

学習によるパターン分類について

九大理新大理 田 中 謙 輔

§1. まえがき

最近パターン認識の問題は学習による方法が多く用いられるようになった。これらの方面の研究は、良く知られている様に Ya. Z. Tsypkin, M. A. Aizerman, E. M. Braverman, J. M. Schumpert, S. S. Yau, J. D. Patterson, K. S. Fu, etc などの人々によっていろいろの type でなされた。

ここではいろいろの type の中で、特に H. Robbins, S. Monro, によって導入されている, J. R. Blum によって多次元へ拡張された確率近似法 (Stochastic Approximation) の応用として学習のアルゴリズムが組み立てられている type に注目し、次のような2つの問題を考えた。

(I) 次の (1) (2) (3) をみたすパターン認識の問題

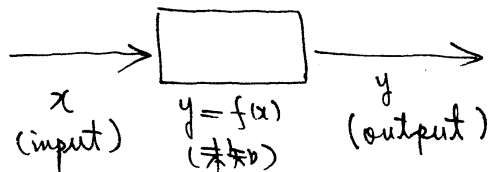
- (1)  $\theta_1, \theta_2$  (category と呼ばれる) なる  $R^m$  にあたるパターンの2つの class が存在するとする。
- (2)  $\theta_1, \theta_2$  の a priori distribution はそれぞれ  $f_1, f_2$  とする
- (3)  $\theta_1, \theta_2$  が真のときパターンの probability density function

は  $z$  に対して  $f_1(x), f_2(x)$  とする。

ここで  $f_1, f_2, f_1(x), f_2(x)$  は未知とする。もしこの問題で  $f_1, f_2, f_1(x), f_2(x)$  が知られていれば Bayes の意味で optimal な判定関数を用いることができる。我々の問題では  $f_1, f_2, f_1(x), f_2(x)$  が未知であるが training sequence  $(x^1, 0^1), \dots$  を用いることができる。これを  $z$  を用いて Bayes の意味で optimal な判定関数  $D(x)$  を近似するアルゴリズムを求め、ただし

$$D(x) = \begin{cases} \frac{f_1 f_1(x) - f_2 f_2(x)}{p(x)} & , p(x) = f_1 f_1(x) + f_2 f_2(x) > 0 \\ 0 & , p(x) = 0 \end{cases}$$

(II) 未知関数を構成する問題



上の system においては、関数  $f(x)$  は未知であるが、input  $x$  と output  $y$  は観測できる。

このとき互に独立な同じ分布にしたがう input の列と、これに対応する output の列から未知関数  $f(x)$  を近似するアルゴリズムを求め、

## §2. 準備

この章では以下の主要結果を証明するために 3 つの補助定理を述べる。  $\{y_n\}_{n=1}^{\infty}$  は実数値 stochastic process とし、  $\{U_n\}_{n=1}^{\infty}$ ,  $\{V_n\}_{n=1}^{\infty}$ ,  $\{Z_n\}_{n=1}^{\infty}$  は非負なる実数値可測関数の 3 つの列とする

$\mathbb{R}^N$  上、 $\delta \geq 0$  の  $U_n, V_n, \zeta_n$  は  $\mathbb{R}^N$  の上に定義されている。  
 このとき  $\{U_n(y^1, y^2, \dots, y^n)\}_{n=1}^{\infty}, \{V_n(y^1, y^2, \dots, y^n)\}_{n=1}^{\infty}, \{\zeta_n(y^1, y^2, \dots, y^n)\}_{n=1}^{\infty}$   
 は再び stochastic process となる。以下では記号を簡単にする  
 ために、 $U_n = U_n(y^1, y^2, \dots, y^n), V_n = V_n(y^1, y^2, \dots, y^n), \zeta_n = \zeta_n(y^1, y^2, \dots, y^n)$   
 と書くことにする。更に  $\{\delta_n\}_{n=1}^{\infty}, \{\mu_n\}_{n=1}^{\infty}$  は実数の  $\geq 0$  の列とする。

今  $\delta > 0$  の stochastic process  $\{U_n\}_{n=1}^{\infty}, \{V_n\}_{n=1}^{\infty}, \{\zeta_n\}_{n=1}^{\infty}$  に対して  
 次の標本基本条件を導入する：

(A1)  $E[U_1]$  と  $E[V_1]$  は存在する。

(A2) 全ての  $n \geq 1$  として

$$E[U_{n+1} | y^1, y^2, \dots, y^n] \leq (1 + \mu_n) U_n - \delta_n V_n + \zeta_n \quad \text{が成立する}$$

(A3)  $\sum_{n=1}^{\infty} \delta_n = \infty$  .

(A4)  $\sum_{n=1}^{\infty} |\mu_n| < \infty$  .

(A5) 次の標本条件を満たす正数の列  $\{M_n\}_{n=1}^{\infty}$  が存在する

$$\text{全ての } n \geq 1 \text{ として } P[\zeta_n \leq M_n] = 1, \quad ,$$

$$\sum_{n=1}^{\infty} M_n < \infty .$$

補助定理 1.  $\delta > 0$  の stochastic process  $\{U_n\}_{n=1}^{\infty}, \{V_n\}_{n=1}^{\infty},$

$\{\zeta_n\}_{n=1}^{\infty}$  が次の標本条件を満たしているとする：(i) 条件 (A1)

~ (A5) が成立する, (ii)  $\lim_{n \rightarrow \infty} \delta_n = 0$ , (iii)  $\exists \{P[\lim_{k \rightarrow \infty} V_{n_k} = 0] = 1$

なる部分列  $\{n_k\}$  が存在すれば,  $P[\lim_{k \rightarrow \infty} U_{n_k} = 0] = 1$  となる。

このとき次の標本事柄が成立する

$$P[\lim_{n \rightarrow \infty} U_n = 0] = 1, \quad \lim_{n \rightarrow \infty} E[U_n^\beta] = 0 \text{ for all } 0 < \beta < 1.$$

補助定理 2. 次の標本条件を満足している非負の実数列  $\{a_n\}_{n=1}^{\infty}$

が存在するとする:

$$(2.1) \quad a_{n+1} \leq (1 - \delta_{n+1})a_n + A_{n+1}, \quad \forall n \geq n_0$$

$$(2.2) \quad \sum_{n=1}^{\infty} \delta_n = \infty,$$

$$(2.3) \quad \lim_{n \rightarrow \infty} \delta_n = 0,$$

$$(2.4) \quad \sum_{n=1}^{\infty} A_n < \infty.$$

ただし  $\{\delta_n\}_{n=1}^{\infty}$ ,  $\{A_n\}_{n=1}^{\infty}$  は 2 の非負の実数列で,  $n_0$  はある正整数とする.

このとき次の標本事柄が成立する.

$$\lim_{n \rightarrow \infty} a_n = 0.$$

補助定理 3. 次の標本条件を満足している非負の実数列  $\{a_n\}_{n=1}^{\infty}$

が存在するとする:

$$(2.5) \quad a_{n+1} \leq (1 - A/n^s) a_n + B/n^t, \quad \forall n \geq n_0,$$

$$(2.6) \quad t \text{ はある実数で } 0 < s < 1 \text{ である.}$$

このとき次の標本事柄が成立する

$$\limsup_{n \rightarrow \infty} n^{t-s} a_n \leq B/A.$$

§3. Teacher をもつ学習によるパターン分類

我々が「学習」と言う言葉に「感じる」「知る」の面を考  
えてみると、これまで論じてきたモデル(I)を次の構に変  
型したモデルに想定するとは自然である構に思われる。こ  
こで「category」の集合  $\mathcal{H} = \{\theta_1, \theta_2\}$  の場合を述べる。

time	1	2	....	n	...
category	$\theta^1$	$\theta^2$	....	$\theta^n$	...
pattern space	$X_1$	$X_2$	....	$X_n$	...
pattern	$x^1$	$x^2$	....	$x^n$	...
transition prob. density fn. (未知)	$p(x^1, \theta^1)$	$p(x^2, \theta^2   x^1, \theta^1)$	....	$p(x^n, \theta^n   \xi^{n-1}, d^{n-1})$ $\xi^{n-1} = (x^1, \dots, x^{n-1})$ $d^{n-1} = (\theta^1, \theta^2, \dots, \theta^{n-1})$	....

理想的な状態 (各時点での transition prob density fn を知って  
いる) では, Bayes の意味で optimal に判定をする。毎各  
時点  $n$  で  $x^n$  を観測したとき

$$D^*(x^n | \xi^{n-1}, d^{n-1}) = \Pi_{x^n}(d^{n-1} \theta_1) - \Pi_{x^n}(d^{n-1} \theta_2) \geq 0 \quad \text{ならば}$$

$x^n$  は category  $\theta_1$  をもつと判定する。

$$D^*(x^n | \xi^{n-1}, d^{n-1}) = \Pi_{x^n}(d^{n-1} \theta_1) - \Pi_{x^n}(d^{n-1} \theta_2) < 0 \quad \text{ならば}$$

$x^n$  は category  $\theta_2$  であると判定する。以下 (

$$\Pi_{x^n}(d^{n-1} \theta_1) = \frac{\Pi(d^{n-1}) p(x^n, \theta_1 | \xi^{n-1}, d^{n-1})}{\sum_{\hat{\theta}^{n-1} \in \mathcal{H}^{n-1}} \Pi(\hat{\alpha}^{n-1}) p(x^n, \hat{\theta}^n | \xi^{n-1}, d^{n-1})}$$

(  $\Pi(d^{n-1})$  は  $\mathcal{H}^{n-1} = \mathcal{H} \times \mathcal{H} \times \dots \times \mathcal{H} \ni d^{n-1}$  の上の確率 )

上の標本判定方法は統計的決定理論より誤りの確率を最小にすることが知られている。

然し我々の問題では各時点での transition prob. density fn は未知であるが、従属関係をもつている training sequence  $(x^1, \theta^1)$ ,  $(x^2, \theta^2)$ , ... が利用できる場合を考える。このとき各時点で見られる orthonormal function の集合  $\{\varphi_i^{(m)}(x)\}_{i=1}^N$  と training sequence を用いて、各時点  $n$  で、

$$I_n = E[(D(x^n | \xi^{n-1}, d^{n-1}) - \sum_{i=1}^N c_i^{(m)}(\xi^{n-1}, d^{n-1}) \varphi_i^{(m)}(x^n))^2 | \xi^{n-1}, d^{n-1}]$$

$$= \int_{X_n} (D(x^n | \xi^{n-1}, d^{n-1}) - \sum_{i=1}^N c_i^{(m)}(\xi^{n-1}, d^{n-1}) \varphi_i^{(m)}(x^n))^2 p(x^n, \theta^n | \xi^{n-1}, d^{n-1}) dx^n$$

$$(D(x^n | \xi^{n-1}, d^{n-1}) = p(x^n, \theta_1 | \xi^{n-1}, d^{n-1}) - p(x^n, \theta_2 | \xi^{n-1}, d^{n-1}))$$

を最小にする  $\{c_i^{(m)}(\xi^{n-1}, d^{n-1})\}_{i=1}^N$  を求めて、 $\sum_{i=1}^N c_i^{(m)}(\xi^{n-1}, d^{n-1}) \varphi_i^{(m)}(x^n)$  を判定関数として、 $x^n$  の category  $\theta^n$  を判定することをした。

$$\text{今 } \frac{\partial I_n}{\partial c_j^{(m)}} = 0 \quad (j=1, 2, \dots, N) \quad \text{より}$$

$$c_{j*}^{(m)}(\xi^{n-1}, d^{n-1}) = E_{\theta_1}[\varphi_j^{(m)}(x^n) | \xi^{n-1}, d^{n-1}] - E_{\theta_2}[\varphi_j^{(m)}(x^n) | \xi^{n-1}, d^{n-1}]$$

から得られる

$$\text{したがって } E_{\theta_1}[\varphi_j^{(m)}(x^n) | \xi^{n-1}, d^{n-1}] = \int_{X_n} \varphi_j^{(m)}(x^n) p(x^n, \theta_1 | \xi^{n-1}, d^{n-1}) dx^n$$

上の議論から T. Kitagawa [6] によって導入され、V. Dupač [4] によって詳細に論じられた "modified stochastic appro-

"ximation" の応用として次の線形アルゴリズムが考えられる。

### Algorithm

最初に 観測される  $x^1$  と teacher = 指示される  $\theta^1$  の対  $(x^1, \theta^1)$  より

各  $j = 1, \dots, N$  について, 次の線形型で  $c_j^{(1)}(x^1, d^1)$  を構成する:

$$c_j^{(1)}(x^1, d^1) = c_j^{(0)} + r_1 [f^{(1)}(\theta^1) \varphi_j^{(1)}(x^1) - (1 - f^{(1)}(\theta^1)) \varphi_j^{(1)}(x^1)]$$

ただし  $c_j^{(0)} = 0$  ( $j = 1, 2, \dots, N$ )

$$f^{(1)}(\theta^1) = \begin{cases} 1 & \theta^1 = \theta_1 \\ 0 & \theta^1 = \theta_2 \end{cases}$$

次に 観測される  $x^2$  と teacher = よって指示される  $\theta^2$  の対  $(x^2, \theta^2)$

より各  $j = 1, \dots, N$  について, 次の線形型で  $c_j^{(2)}(x^2, d^2)$  を構成する:

$$c_j^{(2)}(x^2, d^2) = c_j^{(1)}(x^1, d^1) + r_2 [f^{(2)}(\theta^2) \varphi_j^{(2)}(x^2) - (1 - f^{(2)}(\theta^2)) \varphi_j^{(2)}(x^2)]$$

ただし  $f^{(2)}(\theta^2) = \begin{cases} 1 & \theta^2 = \theta_1 \\ 0 & \theta^2 = \theta_2 \end{cases}$

一般には 観測される  $x^n$  と teacher = よって指示される  $\theta^n$  の対

$(x^n, \theta^n)$  より各  $j = 1, \dots, N$  について 次の線形型で  $c_j^{(n)}(x^n, d^n)$  を構成する:

$$c_j^{(n)}(x^n, d^n) = c_j^{(n-1)}(x^{n-1}, d^{n-1}) + r_n [f^{(n)}(\theta^n) \varphi_j^{(n)}(x^n) - (1 - f^{(n)}(\theta^n)) \varphi_j^{(n)}(x^n)]$$

ただし  $f^{(n)}(\theta^n) = \begin{cases} 1 & \theta^n = \theta_1 \\ 0 & \theta^n = \theta_2 \end{cases}$

ただし上のアルゴリズムで用いられる非負の実数列  $\{r_n\}_{n=1}^{\infty}$  は次の線形性質を満たしている:

$$\sum_{n=1}^{\infty} r_n = \infty, \quad \sum_{n=1}^{\infty} \delta_n^2 < \infty.$$

このとき  $C_j^{(n)}(z^n, d^n)$  と  $C_{j*}^{(n)}(z^n, d^n)$  によって以下の確率定理が成立することが示される。

定理3.1 次の様な条件を満して  $\{M_n\}_{n=1}^{\infty}$  と

3つの非負の実数  $K_1, K_2, K_3$  が存在する:

$$(i) \quad P[(Q_j^{(n)})^2 \leq r_{n+1} M_n] = 1 \quad \forall n, j,$$

$$(ii) \quad \sum_{n=1}^{\infty} M_n < \infty,$$

$$(iii) \quad \text{Var}[Y_j^{(n+1)} | z^n, d^n] \leq K_1 (U_j^{(n)})^2 + K_2 (Q_j^{(n)})^2 + K_3,$$

$$\text{ただし } Q_j^{(n)} = C_{j*}^{(n)}(z^n, d^n) - C_j^{(n+1)}(z^n, d^n), \quad U_j^{(n)} = C_j^{(n)}(z^n, d^n) - C_{j*}^{(n)}(z^n, d^n)$$

$$Y_j^{(n+1)} = f^{(n+1)}(\theta^{n+1}) \varphi_j^{(n+1)}(x^{n+1}) - (1 - f^{(n+1)}(\theta^{n+1})) \varphi_j^{(n+1)}(x^{n+1}).$$

このとき次の様な事柄が成立する

$$P[\lim_{n \rightarrow \infty} U_j^{(n)} = 0] = 1 \quad \forall j, \quad \lim_{n \rightarrow \infty} E[(U_j^{(n)})^{2\beta}] = 0, \quad 0 < \beta \leq 1.$$

証明 各  $j = 1, 2, \dots, N$  に対して,  $C_j^{(n+1)}(z^{n+1}, d^{n+1})$  の構成方法より

$$\begin{aligned} C_j^{(n+1)}(z^{n+1}, d^{n+1}) - C_{j*}^{(n+1)}(z^n, d^n) &= (1 - r_{n+1}) [C_j^{(n)}(z^n, d^n) - C_{j*}^{(n)}(z^n, d^n)] + \\ &+ (1 - r_{n+1}) [C_{j*}^{(n)}(z^n, d^n) - C_{j*}^{(n+1)}(z^n, d^n)] + r_{n+1} [f^{(n+1)}(\theta^{n+1}) \varphi_j^{(n+1)}(x^{n+1}) - \\ &+ (1 - f^{(n+1)}(\theta^{n+1})) \varphi_j^{(n+1)}(x^{n+1}) - C_{j*}^{(n+1)}(z^n, d^n)]. \end{aligned}$$

上式に記号  $U_j^{(n)}, Q_j^{(n)}$  を用いて表わすとよ。て次式を得る

$$\begin{aligned} (U_j^{(n+1)})^2 &\leq (1 - r_{n+1})^2 (U_j^{(n)})^2 + (1 - r_{n+1})^2 (Q_j^{(n)})^2 + r_{n+1}^2 [Y_j^{(n+1)} - C_{j*}^{(n+1)}(z^n, d^n)]^2 \\ &+ 2(1 - r_{n+1})^2 |U_j^{(n)}| |Q_j^{(n)}| + 2(1 - r_{n+1}) r_{n+1} (Y_j^{(n+1)} - C_{j*}^{(n+1)}(z^n, d^n)) (U_j^{(n)}) \\ &+ 2(1 - r_{n+1}) r_{n+1} (Y_j^{(n+1)} - C_{j*}^{(n+1)}(z^n, d^n)) (Q_j^{(n)}). \end{aligned}$$

両辺の条件付き期待値をとり, 定理の条件を用いると



$$E[(U_j^{(n)})^2 | \xi^n, d^n] \leq \{1 - 2\gamma_{n+1} + \gamma_{n+1}^2(1 + K_1)\} (U_j^{(n)})^2 + \{1 + \gamma_{n+1}^2(1 + K_2)\} (O_j^{(n)})^2 + 2|U_j^{(n)}| |O_j^{(n)}|$$

が成立し、更に不等式

$$2|U_j^{(n)}| |O_j^{(n)}| \leq \gamma_{n+1} (U_j^{(n)})^2 + (O_j^{(n)})^2 / \gamma_{n+1} \quad \text{を用いると}$$

によつて

$$(3.1) \quad E[(U_j^{(n)})^2 | \xi^n, d^n] \leq \{1 + \gamma_{n+1}^2(1 + K_1)\} (U_j^{(n)})^2 - \gamma_{n+1} (U_j^{(n)})^2 + \{1 + \gamma_{n+1}^2(1 + K_2)\} \cdot (O_j^{(n)})^2 + (O_j^{(n)})^2 / \gamma_{n+1} + \gamma_{n+1}^2 K_3$$

を得る。よつて補助定理1を用いて次の結果を得る:

$$P[\lim_{n \rightarrow \infty} U_j^{(n)} = 0] = 1, \quad \lim_{n \rightarrow \infty} E[(U_j^{(n)})^{2\beta}] = 0, \quad 0 < \beta < 1.$$

次に(3.1)の両辺の期待値をとるとより次の式が得られる:

$$E[(U_j^{(n)})^2] \leq [1 - \gamma_{n+1}(1 - \gamma_{n+1}(1 + K_1))\{E[(U_j^{(n)})^2] + \{1 + \gamma_{n+1}^2(1 + K_2)\} \cdot E[(O_j^{(n)})^2] + E[(O_j^{(n)})^2] / \gamma_{n+1} + \gamma_{n+1}^2 K_3$$

よつて補助定理2を用いて次の結果を得る:

$$\lim_{n \rightarrow \infty} E[(U_j^{(n)})^2] = 0.$$

更に平均収束の order  $\alpha > 1/2$  については次の補題が成立する。

定理3.2 定理3.1の条件(iii)の他に次の条件が成立する:

$$(iv) \quad \gamma_n = a/n^\alpha, \quad a > 0, \quad 1/2 < \alpha < 1,$$

$$(v) \quad E[(O_j^{(n)})^2] = O(n^{-2\omega}), \quad \omega > \alpha.$$

このとき次の補題が成立する。

$$E[(U_j^{(n)})^2] = \begin{cases} O(n^{-2(w-\alpha)}) & , w < 3/2\alpha \\ O(n^{-\alpha}) & , w \geq 3/2\alpha \end{cases}$$

証明 定理3.1の証明より, 上の定理の条件を用いることによつて次式が成立する:  $\forall n \geq N$

$$E[(U_j^{(n+1)})^2] \leq (1 - C_1/n^\alpha)E[(U_j^{(n)})^2] + C_2/n^{2\alpha} + C_3/n^{2w-\alpha}$$

上式より  $w < (3/2)\alpha$  ならば

$$E[(U_j^{(n+1)})^2] \leq (1 - C_1/n^\alpha)E[(U_j^{(n)})^2] + C_4/n^{2w-\alpha}$$

$w \geq (3/2)\alpha$  ならば

$$E[(U_j^{(n+1)})^2] \leq (1 - C_1/n^\alpha)E[(U_j^{(n)})^2] + C_5/n^{2\alpha}$$

ただし  $C_1, C_2, C_3, C_4, C_5$  は或る定数である。これ故に補助定理3を用いて定理の結果が得られる。

### 例3.1

(i)  $n$  階線形空間  $X_n = X \cdot \forall n$ ,  $(H) = \{0_1, 0_2\}$ ,

(ii)  $p(x^{(n)}, 0_i | \xi^n, \alpha^n) = q^{(n+1)}(0_i | 10^n) f_i(x^{(n+1)}) \quad \forall n$ ,

(iii)  $\sum_{0_i \in (H)} q^{(n+1)}(0_i | 10^n) = 1$ ,

(iv)  $\int_X f_i(x) dx = 1$ ,

(v) 次の標本条件を満たす  $1 \geq f_i \geq 0$  なる  $f_i$  が存在する:

$$P[|q^{(n+1)}(0_i | 10^n) - f_i| \leq C/n^{1+\alpha}] = 1 \quad \forall n$$

ただし  $C, \alpha$  は或る正数である。

上の標本特別なモデルに対して, 更に

$$(vi) \quad \delta_n = 1/n$$

$$(vii) \quad E_{O_i}[\varphi_j^2(x)] = \int_X \varphi_j^2(x) f_i(x) dx < \infty \quad \forall i, j$$

$$(\varphi_j^{(1)}(x) = \varphi_j^{(2)}(x) = \dots = \varphi_j^{(n)}(x) = \dots)$$

ある orthonormal fn の集合  $\{\varphi_i(x)\}_{i=1}^N$  が存在する。

上の様な状態では定理3.1の結果が成立する。

例3.2 例3.1 にあける条件 (i) (ii) (iii) (iv) (vii) の他に

$$(i) \quad \delta_n = a/n^\alpha, \quad a > 0, \quad (1/2) < \alpha < 1,$$

(ii) 次の様な条件を満たす  $1 \geq f_1 \geq 0$  なる  $f_1$  が存在する:

$$E[(\varphi^{(n+1)}(0, 1, 0^n) - f_1)^2] = O((n+1)^{-2\omega}), \quad \omega > \alpha.$$

このような状態では定理3.2の結果が成立する。

#### §4. Self-learning によるパターン分類

ここでは §3 と同じ type のモデルについて考える。ただし

この場合には category の集合  $\mathcal{H} = \{0_1, 0_2, \dots, 0_s\}$  によって

きちんとして知識はなく, training sequence  $(0_i)$  によって

の情報を(含む)も利用できる。この場合には, 各時点で prob.

density fn.  $p(x^n | \xi^n, d^{n-1}) = \sum_{0_i \in \mathcal{H}} p(x^n, 0_i | \xi^n, d^{n-1})$  のモード

に  $\rightarrow$  の category を対応させるようにする。

従ってここでは各時点  $n+1$  で与えられた orthonormal function

の集合  $\{\varphi_i^{(n+1)}(x^{n+1})\}_{i=1}^N$  の一次式で未知な関数  $p(x^{n+1} | \xi^n, d^n)$  を

近似することを考える。この方法は各時点  $n+1$  で

$$J_{n+1} = \int_{X^{n+1}} [p(x^{n+1} | \xi^n, d^n) - \sum_{i=1}^N c_i^{(n+1)}(\xi^n, d^n) \varphi_i^{(n+1)}(x^{n+1})]^2 dx^{n+1}$$

を最小にする  $\{c_{j*}^{(n)}(z^n, d^n)\}_{j=1}^N$  を求める問題となる。

$$\hookrightarrow \frac{\partial J_{n+1}}{\partial c_j^{(n+1)}} = 0 \quad \text{として}$$

$$c_{j*}^{(n+1)}(z^n, d^n) = E[c_j^{(n+1)}(x^{n+1}) | z^n, d^n] = \sum_{i=1}^N \int_{X_{n+1}} c_j^{(n+1)}(x^{n+1}) p(x^{n+1}, \theta_i | z^n, d^n) dx^{n+1}$$

を得る。

上の議論から "modified stochastic approximation" の応用として次の様なアルゴリズムが考えられる。

### Algorithm

最初 = 観測される  $x^1$  より次の様な型で  $c_j^{(1)}(z^1, d^1)$  を構成する:

$$c_j^{(1)}(z^1, d^1) = c_j^{(0)} + \gamma_1 (c_j^{(1)}(x^1) - c_j^{(0)}) \quad j=1, 2, \dots, N$$

$$\text{ただし } c_j^{(0)} = 0 \quad (j=1, 2, \dots, N)$$

次に 観測される  $x^2$  より各  $j=1, \dots, N$  について 次の様な型で  $c_j^{(2)}(z^2, d^2)$  を構成する:

$$c_j^{(2)}(z^2, d^2) = c_j^{(1)}(z^1, d^1) + \gamma_2 (c_j^{(2)}(x^2) - c_j^{(1)}(z^1, d^1))$$

一般には 観測される  $x^n$  より各  $j=1, \dots, N$  について 次の様な型で  $c_j^{(n)}(z^n, d^n)$  を構成する:

$$c_j^{(n)}(z^n, d^n) = c_j^{(n-1)}(z^{n-1}, d^{n-1}) + \gamma_n (c_j^{(n)}(x^n) - c_j^{(n-1)}(z^{n-1}, d^{n-1}))$$

ただし上のアルゴリズムで用いられる非負の実数列  $\{\gamma_n\}_{n=1}^{\infty}$

は次の様な性質を満足している:

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty.$$

このとき  $C_j^{(n)}(\xi^n, d^n)$  と  $C_j^{(n)}(\xi^{n-1}, d^{n-1})$  によって以下の様な定理が成立する。

定理4.1 次の様な条件を満足してゐる非負の実数列  $\{M_n\}_{n=1}^{\infty}$  と  $\gamma > 0$  の非負の実数  $K_1, K_2, K_3$  が存在する:

$$(i) \quad P[(O_j^{(n)})^2 \leq \gamma_{n+1} M_n] = 1 \quad \forall n, j$$

$$(ii) \quad \sum_{n=1}^{\infty} M_n < \infty$$

$$(iii) \quad \text{Var}[C_j^{(n+1)}(x^{n+1}) | \xi^n, d^n] \leq K_1 (U_j^{(n)})^2 + K_2 (O_j^{(n)})^2 + K_3$$

このとき次の様な事柄が成立する

$$P[\lim_{n \rightarrow \infty} U_j^{(n)} = 0] = 1, \quad \forall j, \quad \lim_{n \rightarrow \infty} E[(U_j^{(n)})^2] = 0 \quad 0 < \frac{\gamma}{\gamma_j} \leq 1$$

更に平均収束の order によってはこの様な定理が成立する。

定理4.2 定理4.1 の条件 (iii) の他は

$$(i) \quad \gamma_n = a/n^\alpha, \quad a > 0, \quad (1/2) < \alpha < 1$$

$$(ii) \quad E[(O_j^{(n)})^2] = O(n^{-2\omega}), \quad \omega > \alpha.$$

このとき次の様な事柄が成立する:

$$E[(U_j^{(n)})^2] = \begin{cases} O(n^{-2(\omega-\alpha)}) & , \omega < (3/2)\alpha \\ O(n^{-\alpha}) & , \omega \geq (3/2)\alpha \end{cases}$$

#### 例4.1

(i)  $n$  次元空間  $X_n = X \quad \forall n$ ,  $\mathcal{H} = \{\theta_1, \theta_2, \dots, \theta_s\}$ ,

(ii)  $p(x^{n+1}, \theta_i | \xi^n, d^n) = f^{(n+1)}(\theta_i | \theta^n) f_i(x^{n+1}) \quad \forall i$ ,

$$(iii) \sum_{0_i \in \Omega} q^{(n+1)}(0_i | 10^n) = 1,$$

$$(iv) \int_X f_{0_i}(x) dx = 1 \quad \forall i,$$

(v) 次の様な条件をみたす実数の集合  $\{q_i\}_{i=1}^S$  ( $1 \geq q_i \geq 0, \sum_{i=1}^S q_i = 1$ ) が存在する

$$P[|q^{(n+1)}(0_i | 10^n) - q_i| \leq C / (n+1)^{\alpha}] = 1 \quad \forall i, n$$

ただし  $C, \alpha$  は或る正数である。

上の様な特別なモデルに対して、更に

$$(vi) \quad \sigma_n = 1/n$$

$$(vii) \quad E_{0_i}[q_j^2(x)] = \int_X q_j^2(x) f_{0_i}(x) dx < \infty \quad \forall i, j$$

$$(q_j^{(0)}(x) = q_j^{(1)}(x) = \dots = q_j^{(n)}(x) = \dots)$$

なる orthonormal  $f_n$  の集合  $\{f_i(x)\}_{i=1}^N$  が存在する。

上の様な状態では定理 4.1 の結果が成立する。

例 4.2 例 4.1 の中の条件 (i) (ii) (iii) (iv) (vii) の他に

$$(i) \quad \sigma_n = a/n^\alpha, \quad a > 0, \quad (1/2) < \alpha < 1$$

(ii) 次の様な条件をみたす実数の集合  $\{q_i\}_{i=1}^S$  ( $1 \geq q_i \geq 0, \sum_{i=1}^S q_i = 1$ ) が存在する:

$$E[(q^{(n+1)}(0_i | 10^n) - q_i)^2] = O((n+1)^{-\omega}), \quad \omega > \alpha, \quad \forall i.$$

= a と主定理 4.2 の結果が成立する。

### §.5 未知関数の構成

ある object の特性が関数であるか否かを知る場合には、その input と output を観測する = とより、その関数を構成する = とを第

一章で述べた。これを更に次の様に一般化する:

- (1) 各時点で input 空間  $X$  の中に観測される  $x^n$  は  $x^1, x^2, \dots, x^{n-1}$  の結果に依存する。
- (2) 各時点での未知関数  $f^{(n)}(x)$  が時間と共に少しずつ変化しながら安定する。

従ってモデルは次の様な型になる。

time	1	2	-----	n	---
input space	$X$	$X$	-----	$X$	---
input	$x^1$	$x^2$	-----	$x^n$	---
output	$y^1$	$y^2$	-----	$y^n$	---
	$(=f^{(1)}(x^1))$	$(=f^{(2)}(x^2))$	-----	$(=f^{(n)}(x^n))$	
transition prob. density fn. (未知)	$p(x^1)$	$p(x^2 x^1)$	-----	$p(x^n \xi^{n-1})$	---
				$\xi^{n-1} = (x^1, x^2, \dots, x^2)$	

我々の問題では各時点での prob. density fn. は未知であるが、観測可能な input と output の列  $(x^1, y^1), (x^2, y^2), \dots$  を用いて、各時点  $n+1$  で与えられた linearly independent continuous function の集合  $\{f_i^{(n+1)}(x)\}_{i=1}^N$  の一次式で近似する = とお考える。

この方法では、各時点  $n+1$  で

$$\begin{aligned} I_{n+1} &= E \left[ \left( y^{n+1} - \sum_{i=1}^N c_i^{(n+1)}(\xi^n) f_i^{(n+1)}(x^{n+1}) \right)^2 \mid \xi^n \right] \\ &= E \left[ \left( y^{n+1} - C^{(n+1)*}(\xi^n) f^{(n+1)}(x^{n+1}) \right)^2 \mid \xi^n \right] \end{aligned}$$

を最小にする vector  $C^{(n+1)*}(\xi^n)$  を求める問題になる、ただし

$$C^{(n+1)}(\xi^n) = (C_1^{(n+1)}(\xi^n), C_2^{(n+1)}(\xi^n), \dots, C_N^{(n+1)}(\xi^n))$$

$$\varphi^{(n+1)}(x^{n+1}) = (\varphi_1^{(n+1)}(x^{n+1}), \varphi_2^{(n+1)}(x^{n+1}), \dots, \varphi_N^{(n+1)}(x^{n+1}))'$$

今  $\nabla C^{(n+1)} I_{n+1} = 0$  より

$$E[(y^{n+1} - C_*^{(n+1)}(\xi^n) \varphi^{(n+1)}(x^{n+1})) \varphi^{(n+1)\prime}(x^{n+1}) | \xi^n] = 0 \text{ が得られる.}$$

上の議論から "modified stochastic approximation" の応用として次の標本プログラムが考えられる。

### Algorithm

最初  $n=1$  に 観測された input  $x^1$  と output  $y^1$  の対  $(x^1, y^1)$  を用いて次の標本型で vector  $C^{(1)}(\xi^1)$  を構成する:

$$C^{(1)}(\xi^1) = C^{(0)} + \gamma_1 (y^1 - C^{(0)} \varphi^{(1)}(x^1)) \varphi^{(1)\prime}(x^1)$$

ただし  $C^{(0)} \equiv 0$  .

次に 観測された input  $x^2$  と output  $y^2$  の対  $(x^2, y^2)$  を用いて次の標本型で vector  $C^{(2)}(\xi^2)$  を構成する:

$$C^{(2)}(\xi^2) = C^{(1)}(\xi^1) + \gamma_2 (y^2 - C^{(1)}(\xi^1) \varphi^{(2)}(x^2)) \varphi^{(2)\prime}(x^2)$$

一般には 観測された input  $x^n$  と output  $y^n$  の対  $(x^n, y^n)$  を用いて次の標本型で vector  $C^{(n)}(\xi^n)$  を構成する:

$$C^{(n)}(\xi^n) = C^{(n-1)}(\xi^{n-1}) + \gamma_n (y^n - C^{(n-1)}(\xi^{n-1}) \varphi^{(n)}(x^n)) \varphi^{(n)\prime}(x^n)$$

ただし上のプログラムで用いられる非負の実数列  $\{\gamma_n\}_{n=1}^{\infty}$  は次の標本性質を満足している:

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty .$$

よって  $C^{(n)}(\xi^n)$  と  $C_*^{(n)}(\xi^{n-1})$  に関して次の標本定理が成立する。



定理5.1 全ての  $x^{(n)}, \xi^n$  に対して次の様な条件が成立する.

$$(1) \quad p(x^{(n+1)} | \xi^n) > 0$$

$$(2) \quad \begin{aligned} E[|f^{(n+1)}(x^{(n+1)}) \varphi_i^{(n+1)}(x^{(n+1)})| | \xi^n] &\leq M_1 \quad \forall i, j, n \\ E[|\varphi_i^{(n+1)}(x^{(n+1)}) \varphi_j^{(n+1)}(x^{(n+1)})| | \xi^n] &\leq M_2 \end{aligned}$$

ただし  $M_1, M_2$  は正の数か或る実数である.

(3)  $A^{(n+1)}(\xi^n) = E[\varphi^{(n+1)}(x^{(n+1)}) \varphi^{(n+1)}(x^{(n+1)}) | \xi^n]$  の最小の固有値  $k_0(\xi^n)$  に対して, 次の様な  $k_0$  が存在する

$$0 < k_0 \leq k_0(\xi^n)$$

$$(4) \quad E[\|Y^{(n+1)} - E[Y^{(n+1)} | \xi^n]\|^2 | \xi^n] \leq K_1 \|U^{(n)}\|^2 + K_2 \|\Theta^{(n)}\|^2 + K_3,$$

ただし  $U^{(n)} = C^{(n)}(\xi^n) - C^*(\xi^{n-1})$ ,  $\Theta^{(n)} = C^*(\xi^{n-1}) - C^*(\xi^n)$

$$Y^{(n+1)} = (y^{(n+1)} - C^{(n)}(\xi^n) \varphi^{(n+1)}(x^{(n+1)})) \varphi^{(n+1)}(x^{(n+1)})$$

$$(5) \quad P[\|\Theta^{(n)}\|^2 \leq \delta_{n+1} M_n] = 1, \quad \sum_{n=1}^{\infty} M_n < \infty.$$

このとき次の様な事柄が成立する:

$$P[\lim_{n \rightarrow \infty} U^{(n)} = 0] = 1, \quad \lim_{n \rightarrow \infty} E[\|U^{(n)}\|^{2\beta}] = 0, \quad 0 < \beta < 1.$$

更に平均収束の order について次の様な定理が成立する.

定理5.2 定理5.1の条件 (1) (2) (3) (4) の他に

$$(1)' \quad \delta_n = a/n^\alpha, \quad a > 0, \quad 1/2 < \alpha < 1,$$

$$(2)' \quad E[\|\Theta^{(n)}\|^2] = O(n^{-\omega}), \quad \omega > \alpha.$$

このとき次の様な事柄が成立する:

$$E[\|U^{(n)}\|^2] = \begin{cases} O(n^{-(\omega-\alpha)}) & , \omega < (3/2)\alpha \\ O(n^{-\alpha}) & , \omega \geq (3/2)\alpha . \end{cases}$$

更に input  $i$  に対応して観測される output が input  $i$  に関係して 11  
 の noise をもつ場合にも, noise の平均が 0 で分散が有界ならば  
 は同様のアルゴリズムで同じ結果がとかれる。

### 参考文献

- [1] Браверман, Э.М. и Розоньэр, П.Н.: Сходимость случайных процессов в теории обучения машин I, Автоматика и Телемеханика, №1, (1969), 57-77.
- [2] Браверман, Э.М. и Розоньэр, П.Н.: Сходимость случайных процессов в теории обучения машин II, Автоматика и Телемеханика, №3, (1969), 87-103.
- [3] Chung, K.L.: On a stochastic approximation method, Ann. Math. Stat., vol. 25, (1954), 463-483.
- [4] Дурас, V.: A dynamic stochastic approximation method, Ann. Math. Stat., vol. 36, (1965), 1695-1702.
- [5] Kitagawa, T.: Successive process of statistical controls I, Mem. Fac. Sci. Kyushu Univ., Ser. A., vol. 7, (1952), 13-28.

- [6] Kitagawa, T.: Successive process of statistical controls II, Mem. Fac. Sci. Kyushu Univ., Ser. A., vol. 13, (1959), 1-16.
- [7] Schumpert, J. M. and Yau, S. S.: Design of pattern classifiers with the updating property using stochastic approximation techniques, IEEE Trans. Computers, vol. C-17, No. 9, (1968), 861-872.