数理解析研究所講究録
第 128 巻 1971 年 170-185

学習によるパターン分類

新大 理　　田 中 謙 輔

§1. 序

この論文は前回の研究集会に発表した学習理論への dynamic stochastic approximation の応用に関連しているものである。この dynamic stochastic approximation は最初 T. Kitagawa [7] によって successive process of statistical control (1959) の中で導入され V. Dupač [4] によって詳細に研究されている。

ここでは 2 つの場合が取り扱われている。第一の場合は 2 つのパターンを分類するためのベースの決定関数を teacher をもつ学習方法に基づいて求めることである。このとき各時点でのパターンの出現が過去のパターンの出現状況に影響される場合である。更に前回はある与えられた直交関数族の一次結合で近似することを考えたが、ここでは C. T. Wolverton, T. J. Wagner [13] によって扱われた non-parametric な方法を用いている。第二の場合では,各時点でのパターンの確率密度関数が次の様な型をしている:

$$p^{(n)}(x^n|\theta^1,\theta^2,\cdots,\theta^{n-1})=\sum_{i=1}^{s} g^{(n)}(\theta_i|\theta^1,\theta^2,\cdots,\theta^{n-1})f^{(n)}_{\theta_i}(x^n),$$

ただし,θ^i は i 時点におけるパターンの型を表している。更に各時点で $f^{(n)}_{\theta_i}(x^n)$ は既知で $g^{(n)}(\theta_i|\theta^1,\theta^2,\cdots,\theta^{n-1})$ は未知である。このとき各時点での観測値 x^n から未知な $g^{(n)}(\theta_i|\theta^1,\theta^2,\cdots,\theta^{n-1})$ を求めることである。ここでは K. S. Fu [5] によって扱われた方法を用いている。

§2. 補助定理について

この章では以下の主な結果を証明するための補助定理を述べる。$\{y^n\}_{n=1}^{\infty}$ は実数値 stochastic process とし,$\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$, $\{\zeta_n\}_{n=1}^{\infty}$ は非負なる実数値可測関数の3つの列とする。このとき $\{U_n(y^1,y^2,\cdots,y^n)\}_{n=1}^{\infty}$, $\{V_n(y^1,y^2,\cdots,y^n)\}_{n=1}^{\infty}$, $\{\zeta_n(y^1,y^2,\cdots,y^n)\}_{n=1}^{\infty}$ は再び stochastic process となる。更に記号を簡単にするために,$U_n=U_n(y^1,y^2,\cdots,y^n)$, $V_n=V_n(y^1,y^2,\cdots,y^n)$, $\zeta_n=\zeta_n(y^1,y^2,\cdots,y^n)$ と書くことにする。以下では $\{\gamma_n\}_{n=1}^{\infty}$, $\{\mu_n\}_{n=1}^{\infty}$ は2つの実数列とする。

今3つの stochastic processes $\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$, $\{\zeta_n\}_{n=1}^{\infty}$ に対して次の様な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 について

$$E[U_{n+1}|y^1, y^2, \cdots, y^n] \leqq (1+\mu_n)U_n - \gamma_n V_n + \zeta_n ,$$

(A 3) $\sum_{n=1}^{\infty} \gamma_n = \infty$, $(\gamma_n \geqq 0,\ n=1, 2, \cdots)$,

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

(A5) 次の様な条件を満足する正数の列$\{M_n\}_{n=1}^{\infty}$が存在する： 全てのnについて

$$P[\zeta_n \leqq M_n] = 1 \quad , \quad \sum_{n=1}^{\infty} M_n < \infty \quad .$$

〔補助定理1〕 3つの stochastic processes $\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$, $\{\zeta_n\}_{n=1}^{\infty}$が次の様な条件を満足しているとする：

(i) 条件(A1)～(A5)が成立する, (ii) $\lim_{n\to\infty} \gamma_n = 0$,

(iii) もし$P[\lim_{k\to\infty} V_{n_k} = 0] = 1$なる部分列$\{n_k\}$が存在すれば, $P[\lim_{k\to\infty} U_{n_k} = 0] = 1$が成立する.

このとき次の様な事柄が成立する

$$P[\lim_{n\to\infty} U_n = 0] = 1 \ , \ \lim_{n\to\infty} E[(U_n)^{\beta}] = 0, \text{ 全ての } 0<\beta<1 .$$

〔補助定理2〕 次の様な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) $a_{n+1} \leqq (1-\gamma_{n+1})a_n + A_n$, 全ての $n \geqq n_0$,

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

(2.3) $\lim_{n\to\infty} \gamma_n = 0$,

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

ただし $\{\gamma_n\}_{n=1}^{\infty}$, $\{A_n\}_{n=1}^{\infty}$は2つの非負の実数列で, n_0は或る正整数である.

このとき次の様な事柄が成立する

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

〔補助定理3〕 次の様な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) $a_{n+1} \leqq (1 - A/n^s) a_n + B/n^t \quad , \quad \forall n \geqq n_0 \quad ,$

(2.6) tはある実数で $0 < s < 1$ である,

ただし A, B は2つの正の実数で, n_0は或る正整数とする。

このとき次の様な事柄が成立する

$$\limsup_{n\to\infty} n^{t-s} a_n \leqq B/A \quad .$$

§3. teacherをもつ学習によるパターン分類

ここでは category の集合 $\Theta = \{\theta_1, \theta_2, \cdots, \theta_s\}$ のとき各時点での (x^{n+1}, θ^{n+1})の出現の確率密度関数が

$$p(x^{n+1}, \theta^{n+1} | \alpha^n) = g^{(n+1)}(\theta^{n+1} | \alpha^n) f_{\theta^{n+1}}(x^{n+1})$$

の型をしている場合を考えている。ただし $\alpha^n = (\theta^1, \theta^2, \cdots, \theta^n)$ で $f_{\theta_i}(x)$ は R^m の上で定義されている。

このとき,ベースの公式によって x^{n+1} を観測したとき次の様な $\alpha^n \theta^{n+1} = (\theta^1, \theta^2, \cdots, \theta^n, \theta^{n+1})$ についての事後確率密度を考えることができる:

$$\Pi_{x^{n+1}}(\alpha^n \theta^{n+1}) = \frac{\pi(\alpha^n) g^{(n+1)}(\theta^{n+1} | \alpha^n) f_{\theta^{n+1}}(x^{n+1})}{\sum_{\tilde{\alpha}^n \in \Theta^n} \sum_{\tilde{\theta}^{n+1} \in \Theta} \pi(\tilde{\alpha}^n) g^{(n+1)}(\tilde{\theta}^{n+1} | \tilde{\alpha}^n) f_{\tilde{\theta}^{n+1}}(x^{n+1})} ,$$

ただし $\pi(\alpha^n)$ は $\Theta^n = \Theta \times \Theta \times \cdots \times \Theta$ の上の確率分布である。

このとき統計的決定理論より $\Pi_{x^{n+1}}(\alpha^n\theta_1)$, $\Pi_{x^{n+1}}(\alpha^n\theta_2)$, ……… $\Pi_{x^{n+1}}(\alpha^n\theta_s)$ の最大値で各時点での分類を考えれば誤りの確率を最小にすることが知られている。従って特に $\Theta=\{\theta_1,\theta_2\}$ のときには各時点で x^{n+1} を観測したとき

$D^*(x^{n+1}|\alpha^n)=\Pi_{x^{n+1}}(\alpha^n\theta_1)-\Pi_{x^{n+1}}(\alpha^n\theta_2)\geqq 0$ ならば x^{n+1} は category θ_1 を持っていると判定する。

$D^*(x^{n+1}|\alpha^n)=\Pi_{x^{n+1}}(\alpha^n\theta_1)-\Pi_{x^{n+1}}(\alpha^n\theta_2)<0$ ならば x^{n+1} は category θ_2 を持っていると判定する.

上の様な関数はベースの決定関数と呼ばれている. この様な決定規則は次の様な決定規則と同値である:

各時点で x^{n+1} を観測したとき

$D^{(n+1)}(x^{n+1}|\alpha^n)=g^{(n+1)}(\theta_1|\alpha^n)f_{\theta_1}(x^{n+1})-g^{(n+1)}(\theta_2|\alpha^n)f_{\theta_2}(x^{n+1})\geqq 0$ ならば x^{n+1} は category θ_1 を持っていると判定する.

$D_0^{(n+1)}(x^{n+1}|\alpha^n)=g^{(n+1)}(\theta_1|\alpha^n)f_{\theta_1}(x^{n+1})-g^{(n+1)}(\theta_2|\alpha^n)f_{\theta_2}(x^{n+1})<0$ ならば x^{n+1} は category θ_2 を持っていると判定する.

次にこの章では以下の様な条件を満足する R^m の上に定義されている実数値関数 $K(\cdot)$ を考える:

(K1) $K(y)\geqq 0$, $\forall y\in R^m$,

(K2) $\sup K(y)=K<\infty$,

(K3) $\int K(y)dy=1$,

(K4) $\int \|y\|\,K(y)dy=K^*<\infty$,

ただし $\|y(y_1, y_2, \cdots, y_m)\|^2 = \sum_{i=1}^{m}(y_i)^2$.

更に $\{h_n\}_{n=1}^{\infty}$ は次の様な条件を満す正の実数列とする:

(i) $1 \geqq h_1 \geqq h_2 \geqq \cdots\cdots\cdots$,

(ii) $\lim_{n\to\infty} h_n = 0$.

ここでは上の様な関数 $K(\cdot)$ と実数列 $\{h_n\}_{n=1}^{\infty}$ を用いて次の様なアルゴリズムを構成する:

アルゴリズム

最初に観測される x^1 と teacher によって指示される θ^1 の対 (x^1, θ^1) より次の様な型で $D_1(x|x^1, \theta^1)$ を構成する:

$$D_1(x|x^1, \theta^1) = \rho^{(1)}(\theta^1)K_1(x, x^1) - (1-\rho^{(1)}(\theta^1))K_1(x, x^1),$$

ただし $K_1(x, x^1) = h_1^{-m}K[h_1^{-1}(x - x^1)]$,

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

次に観測される x^2 と teacher によって指示される θ^2 の対 (x^2, θ^2) より、次の様な型で $D_2(x|\xi^2, \alpha^2)$ を構成する:

$$D_2(x|\xi^2, \alpha^2) = \frac{1}{2}D_1(x|\xi^1, \alpha^1) + \frac{1}{2}[\rho^{(2)}(\theta^2)K_2(x, x^2) - (1-\rho^{(2)}(\theta^2))K_2(x, x^2)],$$

ただし、$\xi^2 = (x^1, x^2)$, $K_2(x, x^2) = h_2^{-m}K[h_2^{-1}(x - x^2)]$,

$$\rho^{(2)}(\theta^2) = \begin{cases} 1, & \theta^2 = \theta_1 \\ 0, & \theta^2 = \theta_2 \end{cases} .$$

一般に観測される x^{n+1} と teacher によって指示される θ^{n+1} の対 (x^{n+1}, θ^{n+1}) より、次の様な型で $D_{n+1}(x|\xi^{n+1}, \alpha^{n+1})$ を構成する:

$$D_{n+1}(x\mid\xi^{n+1},\alpha^{n+1})=\frac{n}{n+1}D_n(x\mid\xi^n,\alpha^n)+\frac{1}{n+1}\left[\rho^{(n+1)}(\theta^{n+1})K_{n+1}(x,x^{n+1})\right.$$
$$\left.-(1-\rho^{(n+1)}(\theta^{n+1}))K_{n+1}(x,x^{n+1})\right],$$

ただし，$\xi^{n+1}=(x^1,x^2,\cdots\cdots,x^{n+1})$，$K_{n+1}(x,x^{n+1})=h_{n+1}^{-m}K[h_{n+1}^{-1}(x-x^{n+1})]$，

$$\rho^{(n+1)}(\theta^{n+1})=\begin{cases}1, & \theta^{n+1}=\theta_1\\ 0, & \theta^{n+1}=\theta_2\end{cases}.$$

〔補助定理4〕 $f(\cdot)$ は uniform Lipschitz 条件を満足する R^m の上に定義されている実数値関数とする：

$$|f(x)-f(y)|\leqq C\|x-y\| \qquad \forall x,y\in R^m .$$

更に $K(\cdot)$ は(K1)～(K4)を満足している。

このとき次の様な事柄が成立する

$$|f_n(x)-f(x)|\leqq CK^*h_n ,$$

ただし $f_n(x)=\int K_n(x,y)f(y)dy$.

〔証明〕
$$\begin{aligned}|f_n(x)-f(x)|&=\left|\int K_n(x,y)(f(y)-f(x))dy\right|\\&\leqq\int K(z)|f(x-zh_n)-f(x)|dz\\&\leqq C\int K(z)\|h_nz\|dz\\&=CK^*h_n .\end{aligned}$$

このとき $D_{n+1}(x\mid\xi^{n+1},\alpha^{n+1})$ と $D_0^{(n+1)}(x\mid\alpha^n)$ に関して次の様な定理が成立する。

〔定理1〕 次の様な条件を満足しているとする：

(i) $\int(f_{\theta_i}(x))^2dx<\infty \qquad \forall i$,

(ii) 各 $f_{\theta_i}(x)$ は uniform Lipschitz 条件を満足している，

(iii) $\sum_{n=1}^{\infty} n^{-1} h_n < \infty$, $\sum_{n=1}^{\infty} n^{-2} h_n^{-m} < \infty$,

(iv) 全ての i に対して次の条件を満足する非負の実数 g_i ($0 \leqq g_i \leqq 1$, $\sum_{i=1}^{2} g_i = 1$) と正の実数列 $\{M_n\}_{n=1}^{\infty}$ が存在する,

$$n[g^{(n)}(\theta_i|\alpha^{n-1}) - g_i]^2 \leqq M_n \quad i=1,2, \quad \sum_{n=1}^{\infty} M_n < \infty .$$

このとき次の様な事柄が成立する

$$P[\lim_{n\to\infty} I_n = 0] = 1 \ , \ \lim_{n\to\infty} E[I_n^{\beta}] = 0 \quad 0 < {}^{\forall}\beta < 1 ,$$

ただし $I_n = \int [D_n(x|\xi^n, \alpha^n) - D_0^{(n)}(x|\alpha^{n-1})]^2 dx$.

[証明] $D_{n+1}(x|\xi^{n+1}, \alpha^{n+1})$ の構成から

$$D_{n+1}(x|\xi^{n+1}, \alpha^{n+1}) - D_0^{(n+1)}(x|\alpha^n) = (1-\frac{1}{n+1})(D_n(x|\xi^n, \alpha^n) - D_0^{(n)}(x|\alpha^{n-1}))$$
$$+ \frac{1}{n+1}[\rho^{(n+1)}(\theta^{n+1})K_{n+1}(x, x^{n+1}) - (1-\rho^{(n+1)}(\theta^{n+1}))K_{n+1}(x, x^{n+1}) - D_0^{(n+1)}(x|\alpha^n)]$$
$$+ (1-\frac{1}{n+1})[D_0^{(n)}(x|\alpha^{n-1}) - D_0^{(n+1)}(x|\alpha^n)] .$$

したがって上式から次式が得られる

$$E[I_{n+1}|\xi^n, \alpha^n] = (1-\frac{1}{n+1})^2 I_n + (\frac{n}{n+1})^2 \Theta^{(n)} + (\frac{1}{n+1})^2 E[\int (Y_{n+1}(x, x^{n+1}) -$$
$$- D_0^{(n+1)}(x|\alpha^n))^2 dx|\xi^n, \alpha^n] + 2(\frac{n}{n+1})(\frac{1}{n+1})E[\int (D_n(x|\xi^n, \alpha^n) - D_0^{(n)}(x|\alpha^{n-1}))$$
$$(Y_{n+1}(x, x^{n+1}) - D_0^{(n+1)}(x|\alpha^n))dx|\xi^n, \alpha^n] + 2(\frac{n}{n+1})(\frac{1}{n+1})E[\int (D_0^{(n)}(x|\alpha^{n-1}) -$$
$$- D_0^{(n+1)}(x|\alpha^n))(Y_{n+1}(x, x^{n+1}) - D_0^{(n+1)}(x|\alpha^n))dx|\xi^n, \alpha^n] + 2(\frac{n}{n+1})^2 E[$$
$$\int (D_n(x|\xi^n, \alpha^n) - D_0^{(n)}(x|\alpha^{n-1}))(D_0^{(n)}(x|\alpha^{n-1}) - D_0^{(n+1)}(x|\alpha^n))dx|\xi^n, \alpha^n] ,$$

ただし $Y_{n+1}(x, x^{n+1}) = \rho^{(n+1)}(\theta^{n+1})K_{n+1}(x, x^{n+1}) - (1-\rho^{(n+1)}(\theta^{n+1}))K_{n+1}(x, x^{n+1})$,

$\Theta^{(n)} = \int [D_0^{(n)}(x|\alpha^{n-1}) - D_0^{(n+1)}(x|\alpha^n)]^2 dx$.

更に上式は定理の条件より

$$E[I_{n+1}|\xi^n, \alpha^n] \leqq [1 + (\frac{1}{n+1})^2] I_n - (\frac{1}{n+1}) I_n + (\frac{1}{n+1})^2 (2K h_{n+1}^{-m} + 4M_2)$$

$$+2\left(\frac{1}{n+1}\right)M_1(M_3+M_4)K^* h_{n+1}+2M_5M_{n+1} \quad ,$$

ただし M_1, M_2, M_3, M_4, M_5, K^* は定数である.

ここで[補助定理1]を用いて定理の結果が成立する事が示される.

[定理2] $\{D_n(x|\xi^n, \alpha^n)\}_{n=1}^{\infty}$ は上の議論から得られている決定関数の列とする:

$$P[\lim_{n\to\infty}\int[D_n(x|\xi^n,\alpha^n)-D_0^{(n)}(x|\alpha^{n-1})]^2dx=0]=1 \quad .$$

このとき次の様な事柄が成立する

$$P[\lim_{n\to\infty}(P_{D_n(\cdot|\xi^n,\alpha^n)}(e)-P_{D_0^{(n)}(\cdot|\alpha^{n-1})}(e))=0]=1 \quad ,$$

ただし $P_{D(\cdot)}(e)$ は決定関数 $D(\cdot)$ を用いたときの誤りの確率を表している.

[証明] $\varepsilon>0$ に対して次式が成立する様な自然数 N を得る.

$$P[\int(D_N(x|\xi^N,\alpha^N)-D_0^{(N)}(x|\alpha^{N-1}))^2dx<\varepsilon^2/4\int I_{B^N}(x)dx]=1$$

ただし B^N は $\int_{B^N}|D_0^N(x|\alpha^{N-1})|dx \geqq 1-\varepsilon/2$ が成立する様な R^m の上の有界集合で $I_A(x)$ は A の上の indicator である.

このとき次式が成立する

$$\begin{aligned}&P_{D_N(\cdot|\xi^N,\alpha^N)}(e)-P_{D_0^{(N)}(\cdot|\alpha^{N-1})}(e)\\&=\int D_0^{(N)}(x|\alpha^{N-1})[I_{H_0^N}(x)-I_{H^N}(x)]I_{B^N}(x)dx+\int D_0^{(N)}(x|\alpha^{n-1})[I_{H_0^N}(x)-\\&-I_{H^N}(x)]I_{\overline{B^N}}(x)dx \quad ,\end{aligned}$$

ただし $H_0^N=\{x: D_0^{(N)}(x|\alpha^{N-1})\geqq 0\}$, $H^N=\{x: D_N(x|\xi^N,\alpha^N)\geqq 0\}$, $\overline{B^N}$ は B^N の補集合である.

従って $0 \leqq P_{D_N}(\cdot|\xi^n, \alpha^n)(e) - P_{D_0^{(N)}}(\cdot|\alpha^{n-1})(e) < \varepsilon$ となり定理の結果が成立する.

§4. teacher をもたない学習

ここでは category の集合 $\Theta = \{\theta_1, \theta_2, \cdots, \theta_s\}$ のとき、各時点での (x^{n+1}, θ^{n+1}) の出現する確率密度関数が

$$p(x^{n+1}, \theta^{n+1}|\alpha^n) = g^{(n+1)}(\theta^{n+1}|\alpha^n) f^{(n+1)}_{\theta^{n+1}}(x^{n+1})$$

の型をしている場合を考える。ただし $\alpha^n = (\theta^1, \theta^2, \cdots, \theta^n)$ で $f^{(n+1)}_{\theta_i}(x^{n+1})$ は R^m の上に定義されている。

ここの問題は各時点で $f^{(n+1)}_{\theta_i}(x)$ は既知で $g^{(n+1)}(\theta_i|\alpha^n)$ が未知であるとき観測値 x^{n+1} から未知な $g^{(n+1)}(\theta_i|\alpha^n)$ を求めることである。これを求めることによって前に述べた様に統計的決定理論を用いることができる。ここでは、この問題を解くために、各時点で次の様な量を最小にする $g^{(n)}_{i*}$, $i = 1, 2, \cdots, s$ を求めるアルゴリズムを見つける:

$$I_n = \int \Big[\sum_{i=1}^{s} (\hat{g}_i - g^{(n)}(\theta_i|\alpha^{n-1}) f^{(n)}_{\theta_i}(x^n)\Big]^2 dx^n + 2\lambda\Big[\sum_{i=1}^{s} \hat{g}_i - 1\Big],$$

ただし λ はラグランジ乗数である.

I_n を各 $\hat{g}_i$ で微分して zero とおき次式を得る:

$$W^{(n)} Q^{(n)}_*(\alpha^{n-1}) = E[\mathfrak{f}^{(n)}(x^n)|\alpha^{n-1}] - \lambda \sqcup,$$

ただし $W^{(n)}$ は (i, j) 要素 $w^{(n)}_{ij} = \int f^{(n)}_{\theta_i}(x) f^{(n)}_{\theta_j}(x) dx$ をもつ行列 $E[\mathfrak{f}^{(n)}(x^n)|\alpha^{n-1}]$ は i 要素 $E[f^{(n)}_{\theta_i}(x^n)|\alpha^{n-1}] = \int f^{(n)}_{\theta_i}(x^n) p^{(n)}(x^n|\alpha^{n-1}) dx$ をもっている column vector, $\sqcup$ は全ての要素が 1 である column vector, $Q_*^{(n)}(\alpha^{n-1})$ は i 要素 $g_{i*}^{(n)}$ をもっている column vector.

今, $\det W^{(n)}$ ($W^{(n)}$ の行列式) は各時点 n で zero でないと仮定する。このとき

$$g_{i*}^{(n)}=\sum_{k=1}^{s}\left(E[f_{\theta_k}^{(n)}(x^n)|\alpha^{n-1}]-\frac{\sum_{l=1}^{s}\sum_{j=1}^{s}E[f_{\theta_j}^{(n)}(x^n)|\alpha^{n-1}]W_{jl}^{(n)}-\det W^{(n)}}{\sum_{l=1}^{s}\sum_{j=1}^{s}W_{jl}^{(n)}}\right)\frac{W_{ki}^{(n)}}{\det W^{(n)}},$$

ただし $W_{jl}^{(n)}$ は行列 $W^{(n)}$ における $w_{jl}^{(n)}$ の余因子である.

このとき上の式は次の様に書くことができる:

$$g_{i*}^{(n)}=E[F_i^{(n)}(x^n)|\alpha^{n-1}]$$

ただし $F_i^{(n)}(x^n)=\sum_{k=1}^{s}\left(f_{\theta_k}^{(n)}(x^n)-\frac{\sum_{l=1}^{s}\sum_{j=1}^{s}f_{\theta_j}^{(n)}(x^n)W_{jl}^{(n)}-\det W^{(n)}}{\sum_{l=1}^{s}\sum_{j=1}^{s}W_{jl}^{(n)}}\right)\frac{W_{ki}^{(n)}}{\det W^{(n)}}$.

更に $\{\gamma_n\}_{n=1}^{\infty}$ は次の様な条件を満足する非負の実数列とする:

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

ここでは上の実数列 $\{\gamma_n\}_{n=1}^{\infty}$ を用いて次の様なアルゴリズムを構成する.

$\boxed{\text{アルゴリズム}}$

最初に分類されていない観測値 x^1 を用いて次の様な型で $g_i^{(1)}(\xi^1,\alpha^1)$ を構成する:

$$g_i^{(1)}(\xi^1,\alpha^1)=g_i^{(0)}+\gamma_1[F_i^{(1)}(x^1)-g_i^{(0)}] \qquad \forall i \ ,$$

ただし $g_i^{(0)}\equiv 0 \ , \ \forall i$.

次に分類されていない観測値 x^2 を用いて次の様な型で $g_i^{(2)}(\xi^2,\alpha^2)$

を構成する:

$$g_i^{(2)}(\xi^2,\alpha^2)=g_i^{(1)}(\xi^1,\alpha^1)+\gamma_2[F_i^{(2)}(x^2)-g_i^{(1)}(\xi^1,\alpha^1)] \qquad \forall i \ .$$

一般には分類されていない観測値 x^{n+1} を用いて次の様な型で $g_i^{(n+1)}(\xi^{n+1},\alpha^{n+1})$ を構成する:

$$g_i^{(n+1)}(\xi^{n+1},\alpha^{n+1})=g_i^{(n)}(\xi^n,\alpha^n)+\gamma_{n+1}[F_i^{(n+1)}(x^{n+1})-g_i^{(n)}(\xi^n,\alpha^n)] \ .$$

このとき $g_i^{(n+1)}(\xi^{n+1},\alpha^{n+1})$ と $g_{i*}^{(n+1)}$ に関して次の定理が成立する.

[定理3] 次の様な条件を満足しているとする:

(i) $\det W^{(n)} \neq 0 \ , \quad \forall n \ ,$

(ii) 全ての i に対して次の条件を満足する正の実数 g_i ($0<g_i<1$, $\sum_{i=1}^{s} g_i = 1$) と正の実数列 $\{M_n\}_{n=1}^{\infty}$ が存在する:

$$\gamma_n^{-1}[g^{(n)}(\theta_i|\alpha^{n-1})-g_i]^2 \leqq M_n \quad \forall i, \quad \sum_{n=1}^{\infty} M_n<\infty \ ,$$

(iii) $\int f_{\theta_i}(x)f_{\theta_j}(x)dx<\infty$, $\forall i,j$, $\int f_{\theta_i}(x)dx=1$ $\forall i$, $f_{\theta_i}(x)\geqq$ $\forall i$,

(iv) 次の様な条件を満足する正の実数列 $\{M'_n\}_{n=1}^{\infty}$ が存在する:

$$\gamma_n^{-1}(\int f_{\theta_i}^{(n)}(x)f_{\theta_j}^{(n)}(x)dx-\int f_{\theta_i}(x)f_{\theta_j}(x)dx)^2 \leqq M'_n \qquad \forall i,j,$$
$$\sum_{n=1}^{\infty} M'_n<\infty \ .$$

このとき次の様な事柄が成立する.

$$P[\lim_{n\to\infty} U_i^{(n)}=0]=1 \ , \lim_{n\to\infty} E[(U_i^{(n)})^{2\beta}]=0 \ , \quad 0<\forall\beta\leqq 1 \ , \quad \forall i \ ,$$

ただし. $U_i^{(n)}=g_i^{(n)}(\xi^n,\alpha^n)-g_{i*}^{(n)}$.

[証明] $g_i^{(n+1)}(\xi^{n+1},\alpha^{n+1})$ の構成から

$$U_i^{(n+1)}=(1-\gamma_{n+1})U_i^{(n)}+(1-\gamma_{n+1})\theta_i^{(n)}+\gamma_{n+1}[F_i^{(n+1)}(x^{n+1})-g_{i*}^{(n+1)}] \ ,$$

ただし. $U_i^{(n+1)} = g_i^{(n+1)}(\xi^{n+1}, \alpha^{n+1}) - g_{i*}^{(n+1)}$, $\theta_i^{(n)} = g_{i*}^{(n)} - g_{i*}^{(n+1)}$.

これから次式が得られる.

$$E[(U_i^{(n+1)})^2 | \xi^n, \alpha^n] \leqq (1+\gamma_{n+1}^2)(U_i^{(n)})^2 - \gamma_{n+1}(U_i^{(n)})^2 + MM_n + \gamma_{n+1}^2 \sigma^2 ,$$

ただし M, σ^2 は定数である.

従って〔補助定理1〕と〔補助定理2〕を用いて定理の結果が成立することが示される.

次に平均収束の order について次の定理が成立する.

〔定理4〕 次の様な条件を満足しているとする:

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

(ii) $Var[F_i^{(n)}(x^n) | \alpha^{n-1}] \leqq \sigma^2$, $\forall i$, $\forall n$,

(iii) $E[(\theta_i^{(n)})^2] = O(n^{-2\omega})$, $\omega>\alpha$, $\forall i$,

ただし. $\theta_i^{(n)} = g_{i*}^{(n)} - g_{i*}^{(n+1)}$.

このとき次の様な事柄が成立する.

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

〔証明〕 $n \geqq N$ に対して

$$E[(U_i^{(n+1)})^2] \leqq (1 - c_1/n^\alpha) E[(U_i^{(n)})^2] + c_2/n^{2\alpha} + c_3/n^{2\omega-\alpha} ,$$

ただし c_1, c_2, c_3, は定数である.

上式から次式が得られる.

$$E[(U_i^{(n+1)})^2] \leqq (1 - c_1/n^\alpha) E[(U_i^{(n)})^2] + c_4/n^{2\omega-\alpha} , \quad \omega < (3/2)\alpha ,$$

$$E[(U_i^{(n+1)})^2] \leqq (1 - c_1/n^\alpha) E[(U_i^{(n)})^2] + c_5/n^{2\alpha} , \quad \omega \geqq (3/2)\alpha ,$$

ただし C_4, C_5, は定数である.
従って[補助定理3]を用いて定理の結果が成立することが示される.

参考文献

[1] Браверман, Э.М. и Розоноэр, Л.И.: Сходимость случайных процессов в теории обучения машин. 1, Автоматика и телемеханика, No.1, (1969) 57–77.

[2] Браверман, Э.М. и Розоноэр, Л.И.: Сходимость случайных процессов в теории обучения машин. 2, Автоматика и телемеханика, No.3, (1969) 87–103.

[3] Chung, K.L.: *On a stochastic approximation method*, Ann. Math. Stat., vol.25, (1954) 463–483.

[4] Vupač, V.: *A dynamic stochastic approximation method*, Ann. Math. Stat., vol.36, (1965) 1695–1702.

[5] Fu, K.S.: "*Sequential Method in Pattern Recognition and Machin Learning*" Academic Press. New York, 1968.

[6] Kitagawa, T.: Successive process of statistical controls. 1, Mem. Fac. Sci. Kyushu Univ., Ser. A. vol.7 (1952) 13–28.

[7] Kitagawa, T.: Successive process of statistical controls. 2, Mem. Fac. Sci. Kyushu Univ., Ser. A. vol.13, (1959) 1–16.

[8] Parzen, E.: On estimation of a probability density and mode, Ann. Math. Stat., vol. 33 (1962) 1065–1076.

[9] Saridis, G. N., Nikolic, Z. J. and Fu, K. S.: Stochastic approximation algorithms for system identification, estimation and decomposition of mixture, IEEE. Trans. Systems Science and Cybernetics, vol.5, No.1, January (1969) 8–15.

[10] Tanaka, K.: On the pattern classification problems by learning. (1), Bull. Math. Stat., vol. 10, (1970) 31–49.

[11] Tanaka, K.: On the pattern classification problems by learning. (2), Bull. Math. Stat., vol.10, (1970) 61–73.

[12] Tanaka, K.: On the pattern classification problems by learning.(3), Mem. Fac. Sci. Kyushu Univ., Ser. A., vol. 24, (1970) 249–273.

[13] Wolverton, C.T. and Wagner, T.J.: Asymptotically optimal discriminant functions for pattern classification, IEEE Trans. Information Theory, vol. IT–15, (1969) 258–265.