ON CONVERGENCES OF ASYMPTOTICALLY OPTIMAL DISCRIMINANT FUNCTIONS FOR PATTERN CLASSIFICATION PROBLEMS

Watanabe, Masafumi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13080

出版情報：統計数理研究. 16 (1/2), pp.23-34, 1974-03. Research Association of Statistical Sciences

ON CONVERGENCES OF ASYMPTOTICALLY OPTIMAL DISCRIMINANT FUNCTIONS FOR PATTERN CLASSIFICATION PROBLEMS

By

Masafumi WATANABE*

(Received July 10, 1973)

§ 1. Introduction and Summary.

This is a continuation of the paper [11] and is concerned with the pattern classification problem related to "learning with a teacher". In [11], in the case when the optimal discriminant function is assumed to belong to the L^2 space and the case when that is assumed to be uniformly continuous, we gave algorithms, which were applications of the stochastic approximation method, for constructing the asymptotically optimal estimates, and investigated the convergence (mean convergence and almost sure convergence) of the algorithms. But we did not consider the rate of almost sure convergence. In this paper we shall discuss the convergence of the algorithm in the case when the "optimal discriminant function" (o. d. f.) is continuous and the rate of the almost sure convergence.

This paper consists of five sections. In Section 2, we shall give definition of the o. d. f. and of asymptotically optimal estimates to the o. d. f., and we shall prepare several lemmas to be used throughout subsequent sections. In Section 3, we shall treat the case when the o. d. f. is continuous, and give an algorithm which is more general than the form in [10]. And we shall discuss the almost sure convergence and the mean convergence of asymptotically optimal estimates. In Section 4, we shall give some inequalities concerning the rates of convergences.

§ 2. The formulation of the problem and Preliminaries.

We consider the two-categories classification problem. In this paper, we assume that a pattern space is equal to the N-dimensional Euclidian space R^N, and the set of categories is equal to $\{A, B\}$.

Let (q_A, q_B) denote a given priori distribution on $\{A, B\}$, where $q_A, q_B > 0$ and $q_A + q_B = 1$, and let $f(\cdot|A)(f(\cdot B))$ denote the probability density function of the observed R^N-valued random vector X if X is drown from the class $A(B)$.

Let's assume tentatively q_A, q_B, $f(\cdot|A)$ and $f(\cdot|B)$ are all known to us. Let's consider the discriminant function;

* Department of Mathematics, Kyushu University, Fukuoka.

$$D(x)=q_A\cdot f(x|A)-q_B\cdot f(x|B), \qquad x\in R^N. \tag{2.1}$$

And let's consider the following decision rule based on (2.1); decide X to be from the class A if $D(x)\geqq 0$ for outcome x of X, and decide X from the class B if $D(x)<0$ for the outcome x of X. It is well known that this decision rule minimizes the probability of misclassification, and this rule has been called the Bayes decision rule. In this paper, we shall call (2.1) the optimal discriminant function (hereafter, abbreviated as o. d. f.).

In this paper we shall treat the case when none of q_A, q_B $f(\cdot|A)$ and $f(\cdot|B)$ is known to us, consequently the o. d. f. is unknown to us. In this situation we are supposed to have a training sequence $\{(X_n, \Theta_n)\}_{n=1}^{\infty}$ with the observed pattern $X_n\in R^N$ and the category $\Theta_n\in\{A, B\}$ from which X_n is actually drawn. It is assumed that the category from which an each observed pattern has been drawn is correctly indicated by a teacher at each instant n.

We assume the training sequence $\{(X_n, \Theta_n)\}_{n=1}^{\infty}$ is independently and identically distributed, each X_n has probability density function $f(\cdot|A)(f(\cdot|B))$ if $\Theta_n=A(\Theta_n=B)$, and each Θ_n is distributed as $q_A=\mathrm{P}[\Theta_n=A]$ and $q_B=\mathrm{P}[\Theta_n=B]$. Throughout this paper these assumptions remain valid.

Let $D_m(x)$ be an estimate of the o. d. f. $D(x)$ based on $\{(X_k, \Theta_k)\}_{k=1}^{m}$. Let $P_g(e)$ be the probability of misclassification using a discriminant function $g(\cdot)$ which is a realvalued function on R^N. Here to classify by using $g(\cdot)$ means that we make use of the decision rule; decide x come from the class A if $g(x)\geqq 0$, and x from B if $g(x)<0$.

DEFINITION 1. $\{D_n(\cdot)\}_{n=1}^{\infty}$ is said to be the asymptotically optimal sequence of type I (hereafter, abbreviated as AO(I)), if

$$\lim_{n\to\infty}\mathrm{E}|P_{D_n}(e)-P_D(e)|=0, \tag{2.2}$$

and $\{D_n(\cdot)\}_{n=1}^{\infty}$ is said to be the asymptotically optimal sequence of type II (hereafter, abbreviated as AO(II)), if

$$\lim_{n\to\infty}|P_{D_n}(e)-P_D(e)|=0 \qquad \text{with pr. } 1. \tag{2.3}$$

DEFINITION 2. Let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence of positive numbers such that $\lim_{n\to\infty}\alpha_n=0$. Then, $\{D_n(\cdot)\}_{n=1}^{\infty}$ is said to be the asymptotically optimal sequence of order $\{\alpha_n\}$ of Type I (AO$-$I($\{\alpha_n\}$)), if there exists a constant $C>0$ such that

$$\mathrm{E}|P_{D_n}(e)-P_D(e)|\leqq C\cdot\alpha_n \qquad \text{for all } n\geqq 1, \tag{2.4}$$

and $\{D_n(\cdot)\}_{n=1}^{\infty}$ is said to be assymptotically optimal sequence of order $\{\alpha_n\}$ of type II (AO$-$II($\{\alpha_n\}$)), if for any $\delta>0$ there exists a constant $C(\delta)>0$ such that

$$\mathrm{P}[|P_{D_n}(e)-P_D(e)|\leqq C(\delta)\cdot\alpha_n \quad \text{for all } n\geqq 1]>1-\delta. \tag{2.5}$$

The following lemma was essentially proved in [11].

LEMMA 2.1. *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of non-negative numbers.*

Suppose that there exist three sequences of non-negative numbers $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ and $\{K_n\}_{n=1}^{\infty}$ such that

(2.6) $$A_{n+1} \leqq (1-a_{n+1})A_n + a_{n+1} \cdot b_{n+1} + K_{n+1} \quad \text{for all } n \geqq 1,$$

(2.7) $$\sum_{n=1}^{\infty} a_n = \infty \quad \text{and} \quad \lim_{n\to\infty} a_n = 0,$$

(2.8) $$\lim_{n\to\infty} b_n = 0,$$

(2.9) $$\sum_{n=1}^{\infty} K_n < \infty.$$

Then, it holds that $\lim_{n\to\infty} A_n = 0$.

And if $K_n \equiv 0$ *for all* $n \geqq 1$ *in* (2.6) *and there exist a constant* $\alpha_0 > 0$ *and some positive integer* n_0 *such that*

(2.10) $$(1-a_{n+1}) \cdot \frac{b_n}{b_{n+1}} \leqq (1-\alpha_0 a_{n+1}) \quad \text{for all } n \geqq n_0,$$

where $\{b_n\}_{n=1}^{\infty}$ *need not satisfy the condition* (2.8), *then there exists a constant* $C > 0$ *such that*

(2.11) $$A_n \leqq C \cdot b_n \quad \text{for all } n \geqq 1.$$

The following lemma is a direct application of Theorem III in [2] and Lemma 2.1.

LEMMA 2.2. *Let* $\{U_n\}_{n=1}^{\infty}$ *and* $\{V_n\}_{n=1}^{\infty}$ *be two sequences of random variables on a probability space* $(\Omega, \mathcal{F}, P)$. *Let* $\{\mathcal{F}_n\}_{n=1}^{\infty}$ *be a sequence of σ-fields,* $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$, *where* U_n *and* V_n *are measurable with respect to* $\mathcal{F}_n$ *for each n. And let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Suppose that the following conditions be satisfied*

(i) $0 \leqq U_n$ *a.s.* *for all* $n \geqq 1$

(ii) $E[U_1] < \infty$

(iii) $E[U_{n+1} | \mathcal{F}_n] \leqq (1-a_{n+1})U_n + V_n$ *a.s.* *for all* $n \geqq 1$.

(iv) $\sum_{n=1}^{\infty} E|V_n| < \infty$

(v) $a_n \geqq 0$ $(n=1, 2, \cdots)$, $\lim_{n\to\infty} a_n = 0$ *and* $\sum_{n=1}^{\infty} a_n = \infty$.

Then, $\lim_{n\to\infty} U_n = 0$ *a.s. and* $\lim_{n\to\infty} E[U_n] = 0$.

LEMMA 2.3. *Let* $\{U_n\}_{n=1}^{\infty}$ *be a sequence of random variables on a probability space* $(\Omega, \mathcal{F}, P)$. *Let* $\{\mathcal{F}_n\}_{n=1}^{\infty}$ *be a sequence of σ-fields,* $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$, *where* U_n *is measurable with respect to* $\mathcal{F}_n$ *for each n. And let* $\{a_n\}_{n=1}^{\infty}$ *and* $\{v_n\}_{n=1}^{\infty}$ *be two sequences of positive numbers. Suppose that there exist a positive integer* n_0 *and two positive numbers* $0 < M < \infty$ *and* $0 < \lambda < 1$ *satisfying*

(2.12) $$U_n \geqq 0 \quad \text{a.s. for all } n \geqq n_0,$$

(2.13) $$M \geqq U_{n_0} \quad \text{a.s.}$$

(2.14) $$E[U_{n+1} | \mathcal{F}_n] \leqq (1-a_{n+1})U_n + v_{n+1} \quad \text{a.s. for all } n \geqq n_0.$$

(2.15) $$(1-a_{n+1})\left(\frac{v_n}{v_{n+1}}\right)^{\lambda} \leqq 1 \quad \text{for all } n \geqq n_0,$$

and

$$\sum_{n=1}^{\infty} v_n^{1-\lambda} < \infty . \tag{2.16}$$

Then for any $\delta > 0$ there exists a constant $C(\delta)$ such that

$$P[U_n \leqq C(\delta) \cdot v_n^{\lambda} \quad \textit{for all } n \geqq n_0] > 1-\delta .$$

PROOF. Let A_n be the event such that

$$A_n = [U_k \leqq C \cdot v_n^{\lambda} \quad \text{for } k = n_0,\ n_0+1, \cdots, n] \tag{2.17}$$

for $n \geqq n_0$, where C is a positive constant. It is easily seen that $A_n \in \mathcal{F}_n$. And for $n = n_0+1,\ n_0+2, \cdots$, we put

$$P_n = P[U_k \leqq C \cdot v_k^{\lambda} \quad \text{for } k = n_0, \cdots, n-1 \text{ and } U_n > C \cdot v_n^{\lambda}] . \tag{2.18}$$

Since

$$P_n = P[A_{n-1}] - P[A_n] \qquad \text{for } n \geqq n_0+1 ,$$

we have

$$P[A_n] = P[A_{n_0}] - \sum_{k=n_0+1}^{n} P_n \qquad n \geqq n_0+1 . \tag{2.19}$$

We put

$$E_1[U_{n+1}] = \int_{A_n} E[U_{n+1} \mid \mathcal{F}_n] dP \qquad \text{for } n \geqq n_0 \tag{2.20}$$

$$E_2[U_n] = \int_{A_n} U_n dP \qquad \text{for } n \geqq n_0 . \tag{2.21}$$

By the definition of the conditional expectation, we have

$$E_1[U_{n+1}] = \int_{A_n} U_{n+1} dP \qquad \text{for } n \geqq n_0 . \tag{2.22}$$

From (2.14) we have

$$E_1[U_{n+1}] \leqq (1-a_{n+1}) E_2[U_n] + v_{n+1} \qquad \text{for } n \geqq n_0 . \tag{2.23}$$

And from (2.21) and (2.22), we have

$$\begin{aligned} E_1[U_{n+1}] &= \int_{A_{n+1}} U_{n+1} dP + \int_{A_n - A_{n+1}} U_n dP \\ &\geqq E_2[U_{n+1}] + C \cdot v_{n+1}^{\lambda} \cdot P_{n+1} \qquad \text{for } n \geqq n_0 . \end{aligned} \tag{2.24}$$

Therefore from (2.23) and (2.24), we have

$$E_2[U_{n+1}] + C \cdot v_{n+1}^{\lambda} \cdot P_{n+1} \leqq E_2[U_n] \cdot (1-a_{n+1}) + v_{n+1} \qquad \text{for } n \geqq n_0 . \tag{2.25}$$

In (2.25) we put $z_n = E_2[U_n]/v_n^{\lambda}$ for each $n \geqq n_0$. Then we have

$$z_{n+1} + C \cdot P_{n+1} \leqq \left(\frac{v_n}{v_{n+1}} \right)^{\lambda} (1-a_{n+1}) z_n + v_{n+1}^{1-\lambda} \qquad \text{for } n \geqq n_0 . \tag{2.26}$$

Noting (2.15), we have

$$z_{n+1} + C \cdot P_{n+1} \leqq z_n + v_{n+1}^{1-\lambda} \qquad \text{for } n \geqq n_0 . \tag{2.27}$$

Summing the both side of (2.27) from $n = n_0$ to $n = \infty$ and note that (2.16), we have

(2.28) $$C\cdot\sum_{n=n_0+1}^{\infty}P_n\leqq z_{n_0}+\sum_{n=n_0+1}^{\infty}v_n^{1-\lambda}.$$

Note that $z_{n_0}\leqq \mathrm{E}[U_{n_0}]\cdot v_{n_0}^{-\lambda}\leqq M\cdot v_{n_0}^{-\lambda}$ and $\mathrm{P}[A_{n_0}]=1$ if $C\geqq M\cdot v_{n_0}^{-\lambda}$, choosing C according $C>\max\{M\cdot v_{n_0}^{-\lambda}, (M\cdot v_{n_0}^{-\lambda}+\sum_{n=n_0+1}^{\infty}v_n^{1-\lambda})/\delta\}$ we obtain

(2.29) $$\lim_{n\to\infty}\mathrm{P}[A_n]=1-\sum_{n=n_0+1}^{\infty}P_n>1-\delta.$$

Thus the proof of the lemma is completed.

The following lemma can be found in [7].

LEMMA 2.4. *Let* $\{f_n(\cdot)\}_{n=1}^{\infty}$ *and* $f(\cdot)$ *be a sequence of density functions on* R^N *and a density function on* R^N, *respectively. If* $\lim_{n\to\infty}f_n(x)=f(x)$ *for almost all* $x\in R^N$, *then* $\lim_{n\to\infty}\int_S|f_n(y)-f(y)|dy=0$ *uniformly for all Borel sets* S *in* R^N.

§ 3. Convergences.

Let $K(\cdot)$ be a real-valued Borel function on R^N satisfying

(K1) $K(y)\geqq 0$ for all $y\in R^N$,

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

(K3) $\int_{R^N}K(y)dy=1$,

(K4) $\lim_{\|y\|\to\infty}\|y\|\cdot K(y)=0$ where $\|y\|=(\sum_{i=1}^{N}y_i^2)^{1/2}$ for $y=(y_1,\cdots,y_N)\in R^N$,

and let $\{h_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying

(H1) $1\geqq h_n>0$ for all $n\geqq 1$,

(H2) $\lim_{n\to\infty}h_n=0$.

Then we can define the sequence $\{K_n(x,y)\}_{n=1}^{\infty}$ for $x,y\in R^N$,

(3.1) $$K_n(x,y)=h_n^{-N}\cdot K[h_n^{-1}(x-y)]\qquad \text{for } n=1,2,\cdots.$$

The following lemma can be found in [5].

LEMMA 3.1. *Let* $g(\cdot)$ *be a continuous and bounded function on* R^N *and* $\int_{R^N}|g(x)|dx<\infty$. *Then, it holds that* $\lim_{n\to\infty}\left|\int_{R^N}K_n(x,y)g(y)dy-g(x)\right|=0$ *at all* $x\in R^N$, *where* $K_n(x,y)$ *is defined by* (3.1).

If $f(\cdot|A)$ and $f(\cdot|B)$ are continuous density function on R^N then o. d. f. $D(\cdot)$ is also continuous and bounded function on R^N.

And for $x\in R^N$, we have

(3.2) $$\begin{aligned}&\mathrm{E}[\rho(\Theta_n)\cdot K_n(x,X_n)|(X_1,\Theta_1)\cdots(X_{n-1},\Theta_{n-1})]\\&=\mathrm{E}[\rho(\Theta_n)\cdot K_n(x,X_n)]=\int_{R^N}K_n(x,y)D(y)dy\qquad\text{for } n=1,2,\cdots.\end{aligned}$$

where

$$(3.3)\qquad \rho(\Theta_n)=1 \quad \text{if } \Theta_n=A$$
$$\qquad\qquad =-1 \quad \text{if } \Theta_n=B\,.$$

In this paper, we put

$$(3.4)\qquad D_n^*(x)=\mathrm{E}[\rho(\Theta_n)\cdot K_n(x, X_n)] \qquad \text{for } n=1, 2, \cdots.$$

In view of the above arguments, we shall construct the following algorithm which is an application of stochastic approximation methods.

LEARNING ALGORITHM. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying

(A1) $a_n>0$ $(n=1, 2, \cdots)$ and $\sum_{n=1}^{\infty} a_n=\infty$,

(A2) $\lim_{n\to\infty} a_n=0$.

Then $D_n(x)$ is given by the recursive relation as follows

$$(3.5)\qquad D_0(x)\equiv 0\,, \qquad x\in R^N$$
$$D_{n+1}(x)=D_n(x)+a_{n+1}(\rho(\Theta_{n+1})K_{n+1}(x, X_{n+1})-D_n(x))$$

for $n\geqq 0$ and $x\in R^N$.

The above algorithm is transformed to the following one

$$(3.6)\qquad D_n(x)=\sum_{i=1}^{n}\prod_{j=i+1}^{n}(1-a_j)a_i\cdot\rho(\Theta_i)\cdot K_i(x, X_i)\,, \qquad \text{where } \prod_{j=n+1}^{n}(1-a_j)=1\,.$$

THEOREM 3.1. *Let $f(\cdot|A)$ and $f(\cdot|B)$ be continuous on R^N.*

(i) *If $\lim_{n\to\infty} a_n h_n^{-N}=0$, then it holds that*

$$(3.7)\qquad \lim_{n\to\infty}\mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))^2dx\Big]=0\,.$$

(ii) *If $\sum_{n=1}^{\infty} a_n^2 h_n^{-N}<\infty$, then it holds that*

$$(3.8)\qquad \lim_{n\to\infty}\int_{R^N}(D_n(x)-D(x))^2dx=0 \qquad \textit{with pr. 1 and (3.7)}.$$

PROOF. For each $n\geqq 1$, we put

$$(3.9)\qquad A_n=\int_{R^N}(\mathrm{E}[D_n(x)]-D(x))^2dx\,,$$

$$(3.10)\qquad U_n=\int_{R^N}(D_n(x)-\mathrm{E}[D_n(x)])^2dx$$

and

$$M=\sup_x|D(x)|\,.$$

First, we shall prove that $\lim_{n\to\infty} A_n=0$. From the construction of $D_n(x)$, (3.2) and (3.4), we have

$$(3.11)\qquad \sup_x|\mathrm{E}[D_n(x)]-D(x)|\leqq(1-a_{n+1})\cdot\sup_x|\mathrm{E}[D_n(x)]-D(x)|$$
$$+a_{n+1}\cdot|D_{n+1}^*(x)-D(x)|$$
$$\leqq(1-a_{n+1})\sup|\mathrm{E}[D_n(x)]-D(x)|+2M\cdot a_{n+1}$$

for all $n \geqq n_0$, where n_0 is a positive integer such that $0 < a_n \leqq 1$ for all $n \geqq n_0$. Therefore, from Lemma 2.1, we have

$$\sup_x |\mathrm{E}[D_n(x)] - D(x)| \leqq M_1 \qquad \text{for all } n \geqq 1, \tag{3.12}$$

where M_1 is some positive constant. By the same arguments similar to (3.11), we have

$$A'_{n+1} \leqq (1-a_{n+1})A'_n + a_{n+1}\int_{R^N} |D^*_{n+1}(x) - D(x)|\, dx \tag{3.13}$$

for all $n \geqq n_0$, where

$$A'_n = \int_{R^N} |\mathrm{E}[D_n(x)] - D(x)|\, dx \qquad \text{for } n = 1, 2, \cdots. \tag{3.14}$$

In (3.13), noting that

$$\begin{aligned} |D^*_n(x) - D(x)| \leqq\ & q_A \cdot \left| \int_{R^N} K_n(x, y) f(y|A) dy - f(x|A) \right| \\ & + q_B \cdot \left| \int_{R^N} K_n(x, y) f(y|B) dy - f(x|B) \right| \end{aligned} \tag{3.15}$$

for all $n \geqq 1$ and from Lemma 2.4 and Lemma 3.1, we have

$$\lim_{n\to\infty} \int_{R^N} |D^*_n(x) - D(x)|\, dx = 0. \tag{3.16}$$

Therefore, from Lemma 2.1, we have

$$\lim_{n\to\infty} A'_n = 0. \tag{3.17}$$

And noting that

$$\begin{aligned} A_n &\leqq \sup_x |\mathrm{E}[D_n(x)] - D(x)| \cdot A'_n \\ &\leqq M_1 \cdot A'_n \qquad \text{for all } n \geqq 1. \end{aligned} \tag{3.18}$$

Thus, we have $\lim_{n\to\infty} A_n = 0$.

Noting that $\{(X_n, \Theta_n)\}_{n=1}^{\infty}$ is a sequence of independent random vectors, we have

$$\begin{aligned} & \mathrm{E}[U_{n+1} | (X_1, \Theta_1), \cdots, (X_n, \Theta_n)] \\ & \leqq (1-a_{n+1})U_n + a_{n+1}^2 \cdot \mathrm{E}\left[\int_{R^N} (\rho(\Theta_{n+1}) K_{n+1}(x, X_{n+1}) - D^*_{n+1}(x))^2 dx\right] \end{aligned} \tag{3.19}$$

with $pr.$ 1 for all $n \geqq n_0$. From (K2) and $\sup_x |D(x)| = M$, we have

$$\mathrm{E}\left[\int_{R^N} (\rho(\Theta_n) K_n(x, X_n) - D^*_n(x))^2 dx\right] \leqq 2h_n^{-N} \cdot K + 2M \qquad \text{for all } n \geqq 1. \tag{3.20}$$

Therefore, from Lemma 2.2, if $\sum_{n=1}^{\infty} a_n^2 h_n^{-N} < \infty$ then $\lim_{n\to\infty} \mathrm{E}[U_n] = 0$, and $\lim_{n\to\infty} U_n = 0$ with $pr.$ 1. And taking the expectation on the both side (3.19) and by (3.20), we have

$$\mathrm{E}[U_{n+1}] \leqq (1-a_{n+1})\mathrm{E}[U_n] + 2K \cdot a_{n+1}^2 h_{n+1}^{-N} + 2M a_{n+1}^2 \qquad \text{for all} \tag{3.21}$$

$n \geqq n_0$. Therefore, from Lemma 2.1, if $\lim_{n\to\infty} a_n h_n^{-N} = 0$ then $\lim_{n\to\infty} \mathrm{E}[U_n] = 0$.

Thus the proof of the theorem is completed.

It is easily seen that for any $\varepsilon>0$ there exists a bounded Borel set S_ε in R^N such that

$$|P_{D_n}(e)-P_D(e)| \leqq \int_{R^N} |D_n(x)-D(x)| \cdot I_{S_\varepsilon}(x)dx+\varepsilon \tag{3.22}$$

for all $n \geqq 1$, where $I_S(\cdot)$ is the indicator function of S (see [10]).

By the above argument and Theorem 3.1, we have the following corollary.

COROLLARY 3.1.1. *Let $f(\cdot|A)$ and $f(\cdot|B)$ be continuous densities on R^N.*

(i) *If $\lim_{n\to\infty} a_n h_n^{-N}=0$, then $\{D_n(\cdot)\}_{n=1}^{\infty}$ is* AO(I).

(ii) *If $\sum_{n=1}^{\infty} a_n^2 h_n^{-N}$, then $\{D_n(\cdot)\}_{n=1}^{\infty}$ is* AO(I) *and* AO(II).

NOTE ON DENSITY ESTIMATION. Consider a sequence $X_1, X_2, \cdots$ of independent identically distributed N-dimensional random vectors having a probability density function $f(\cdot)$.

H. Yamato [9] proposed the recursive estimator of the form

$$\hat{f}_n(x)=n^{-1}\sum_{j=1}^{n} K_j(x, X_j) \qquad \text{for all } n \geqq 1 \text{ and } x \in R^N, \tag{3.23}$$

where $K_j(x, y)$ is defined by (3.1), and he showed $\lim_{n\to\infty} \mathrm{E}|\hat{f}_n(x)-f(x)|^2=0$ at all points $x \in R^N$ if $f(\cdot)$ is continuous and $\lim_{n\to\infty} nh_n^N=\infty$.

In this problem, we consider a following modified algorithm

$$f_0(x)=0 \qquad \text{for all } x \in R^N \tag{3.24}$$

$$f_{n+1}(x)=f_n(x)+a_{n+1}(K_{n+1}(x, X_{n+1})) \qquad \text{for } n \geqq 0,\ x \in R^N.$$

If we put $a_n=\dfrac{1}{n}$ $(n=1, 2, \cdots)$ then (3.24) is same as (3.23).

And in this case, by the same arguments similar to Theorem 3.1, it is easily seen that $\lim_{n\to\infty} \mathrm{E}(f_n(x)-f(x))^2=0$ at all $x \in R^N$ if $\lim_{n\to\infty} a_n h_n^{-N}=0$ and $\lim_{n\to\infty} (f_n(x)-f(x))^2 =0$ with *pr.* 1 at all $x \in R^N$ if $\sum_{n=1}^{\infty} a_n^2 h_n^{-N}<\infty$.

§ 4. Rates of Convergences.

Let $K(\cdot)$ be a real-valued Borel function on R^N satisfying (K1), (K2), (K3), (K4) and

$$\int_{R^N} \|y\| K(y)dy=K_0<\infty. \tag{K5}$$

And let $\{a_n\}_{n=1}^{\infty}$ and $\{h_n\}_{n=1}^{\infty}$ be two sequences of real numbers satisfying the conditions in § 3 ((A1), (A2) and (H1), (H2), respectively). In this section, we assume that the above conditions are always satisfied.

ASSUMPTION 1. $f(\cdot|A)$ and $f(\cdot|B)$ are continuous functions on R^N. And there exists a constant $0<C_0<\infty$ such that

$$\max_{\theta \in \{A, B\}} \int_{R^N} |f(x+y|\theta)-f(y|\theta)| dy \leqq C_0 \cdot \|x\| \tag{4.1}$$

for all $x \in R^N$.

ASSUMPTION 2. $f(\cdot|A)$ and $f(\cdot|B)$ satisfy uniform Lipschitz conditions with positive constants C_A and C_B, respectively. And we put $C_f = \max\{C_A, C_B\}$.

THEOREM 4.1. *Let Assumption* 1 *or Assumption* 2 *be satisfied. And if there exist two positive constant* $0 < \alpha_1, \alpha_2 \leqq 1$ *and some positive integer* n_0 *such that*

$$(4.2) \qquad (1-a_{n+1})\frac{h_n}{h_{n+1}} \leqq (1-\alpha_1 a_{n+1}) \qquad \textit{for all } n \geqq n_0,$$

$$(4.3) \qquad (1-a_{n+1})\frac{a_n h_n^{-N}}{a_{n+1}h_{n+1}^{-N}} \leqq (1-\alpha_2 a_{n+1}) \qquad \textit{for all } n \geqq n_0,$$

then there exists a constant $0 < C < \infty$ *such that*

$$(4.4) \qquad \mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))^2 dx\Big] \leqq C \cdot b_n \qquad \textit{for all } n \geqq 1$$

where

$$(4.5) \qquad b_n = \max\{h_n, a_n h_n^{-N}\} \qquad \textit{for each } n \geqq 1.$$

PROOF. Let A_n, A_n' and U_n be defined by (3.9), (3.14) and (3.10), respectively. First, if Assumption 1 holds then

$$(4.6) \qquad \int_{R^N}|D_n^*(x)-D(x)|dx = \int_{R^N}\Big|\int_{R^N}K(z)\cdot D(x-h_n z)dz - D(x)\Big|dx$$
$$\leqq \int_{R^N}\int_{R^N}K(z)|f(x-h_n z|A)-f(x|A)|dzdx$$
$$+\int_{R^N}\int_{R^N}K(z)|f(x-h_n z|B)-f(x|B)|dzdx$$
$$\leqq 2K_0\cdot C_0\cdot h_n \qquad \text{for all } n \geqq 1.$$

We apply Lemma 2.1 to (3.13). Then we have

$$(4.7) \qquad A_n' \leqq C_1 \cdot h_n \qquad \text{for all } n \geqq 1$$

and for some positive constant C_1. Therefore, by (3.18) we have

$$(4.8) \qquad A_n \leqq M_1 \cdot C_1 \cdot h_n \qquad \text{far all } n \geqq 1.$$

In (3.21), by the some argument similar to (4.7) we have

$$(4.9) \qquad \mathrm{E}[U^n] \leqq C_2 \cdot a_n h_n^{-N} \qquad \text{for all } n \geqq 1$$

and for some positive constant C_2. Therefore from (4.8) and (4.9) we have (4.5).

Next, if Assumption 2 holds then we have

$$(4.10) \qquad \sup_x |D_n^*(x)-D(x)| \leqq C_f \cdot K_0 \cdot h_n \qquad \text{for all } n \geqq 1.$$

And note that

$$(4.11) \qquad A_n \leqq \sup_x |\mathrm{E}[D_n(x)]-D(x)| \cdot A_n'.$$

and

$$(4.12)\qquad \sup_x|\mathrm{E}[D_{n+1}(x)]-D(x)|\leqq(1-a_{n+1})\sup_x|\mathrm{E}[D_n(x)]-D(x)| +a_{n+1}\cdot\sup_x|D^*_{n+1}(x)-D(x)| \leqq(1-a_{n+1})\sup_x|\mathrm{E}[D_n(x)]-D(x)|+C_f\cdot K_0\cdot h_{n+1}a_{n+1}.$$

for all $n\geqq 1$.

We apply Lemma 2.1 to (4.12). Then we have

$$(4.13)\qquad \sup_x|\mathrm{E}[D_n(x)]-D(x)|\leqq C_3\cdot h_n \quad \text{for all } n\geqq 1$$

and for some positive constant C_3. From (4.11) and $\lim_{n\to\infty}A'_n=0$, we have

$$(4.14)\qquad A_n\leqq C_4\cdot h_n \quad \text{for all } n\geqq 1$$

and for some positive constant C_4. And (4.9) also holds. Therefore from (4.14) and (4.9), we have (4.5). Thus the proof of the theorem is completed.

THEOREM 4.2. *Let Assumption 1 or Assumption 2 be satisfied. And if there exist two positive constant $0<\alpha_1\leqq 1$, $0<\lambda<1$ and same positive integer n_0 such that* (4.2) *holds and*

$$(4.15)\qquad (1-a_{n+1})\left(\frac{a_n^2\,h_n^{-N}}{a_{n+1}^2h_{n+1}^{-N}}\right)^{\lambda}\leqq 1 \quad \textit{for all } n\geqq n_0$$

and

$$(4.16)\qquad \sum_{n=1}^{\infty}(a_{n+1}^2h_{n+1}^{-N})^{1-\lambda}<\infty,$$

then for any $\delta>0$ there exists a constadt $C(\delta)>0$ such that

$$(4.17)\qquad \mathrm{P}\left[\int_{R^N}(D_n(x)-D(x))^2dx\leqq C(\delta)\cdot b'_n \quad \textit{for all } n\geqq 1\right]>1-\delta,$$

where

$$(4.18)\qquad b'_n=\max\{h_n,(a_n^2\,h_n^{-N})^{\lambda}\} \quad \textit{for all } n\geqq 1.$$

PROOF. From (3.19), (3.20) and (H1), we have

$$(4.19)\qquad \mathrm{E}[U_{n+1}|(X_1,\Theta_1),\cdots,(X_n,\Theta_n)] \leqq(1-a_{n+1})U_n+2(M+K)a_{n+1}^2h_{n+1}^{-N} \quad \text{with } pr.\ 1$$

for all $n\geqq n_1$, where n_1 is some positive integer. Noting (3.6), (K2) and (K3) and applying Lemma 2.3 to (4.19), it holds that for any $\delta>0$ there exists a constant $C'(\delta)$ such that

$$(4.20)\qquad \mathrm{P}[U_n\leqq C'(\delta)(a_n^2\,h_n^{-N})^{\lambda} \quad \text{for all } n\geqq 1]>1-\delta.$$

And from the proof of Theorem 4.1 we have

$$(4.21)\qquad A_n\leqq C''\cdot h_n \quad \text{for all } n\leqq 1,$$

where C'' is some positive constant. Therefore we put $C(\delta)=\max\{C'',C'(\delta)\}$, then (4.17) holds. Thus the proof of the theorem is completed.

If $f(\cdot|A)$ and $f(\cdot|B)$ have bounded supports, then we have

(4.22) $$|P_{D_n}(e)-P_D(e)| \leqq \int_{R^N} |D_n(x)-D(x)| I_S(x)dx \quad \text{for all } n \geqq 1$$

where S is the bounded support of $f(\cdot|A)+f(\cdot|B)$. From (4.22), we have

(4.23) $$|P_{D_n}(e)-P_D(e)| \leqq \int_{R^N} |\mathrm{E}[D_n(x)]-D(x)| dx + \Big[\int_{R^N} (\mathrm{E}[D_n(x)]-D_n(x))^2 dx \cdot \int_{R^N} I_S(x)dx\Big]^{\frac{1}{2}}$$

for all $n \geqq 1$. And noting the proof of Theorem 4.1 and Theorem 4.2 we have the following theorem.

THEOREM 4.3. *Let Assumption* 1 *or Assumption* 2 *be satisfied. And let* $f(\cdot|A)$ *and* $f(\cdot|B)$ *have bounded supports and* $\lim_{n\to\infty} a_n h_n^{-N}=0$.

(i) *If the conditions in Theorem* 4.1 *hold, then* $\{D_n(\cdot)\}_{n+1}^{\infty}$ *is* AO$-$I$\{\beta_n\}$) *where*

(4.24) $$\beta_n = \max\{h_n, (a_n h_n^{-N})^{\frac{1}{2}}\} \quad \textit{for each } n \geqq 1.$$

(ii) *If the conditions in Theorem* 4.2 *hold, then* $\{D_n(\cdot)\}_{n+1}^{\infty}$ *is* AO$-$II($\{\beta'_n\}$) *where*

(4.25) $$\beta'_n = \max\{h_n, (a_n^2 h_n^{-N})^{\lambda}\} \quad \textit{for each } n \geqq 1.$$

EXAMPLES. We put $a_n = n^{-\alpha}$ and $h_n = n^{-t}$ for each $n \geqq 1$. And suppose that

(4.26) $$\frac{1}{2} < \alpha \leqq 1 \quad \text{and} \quad 0 < t < \frac{2\alpha-1}{N}.$$

In this case, the conditions (4.2) and (4.3) in Theorem 4.1 are satisfied and the conditions (4.15) and (4.16) in Theorem 4.2 are also satisfied by λ such that

(4.27) $$\frac{2\alpha-Nt-1}{2\alpha-Nt} > \lambda > 0,$$

moreover (A1), (A2), (H1) and (H2) be satisfied.

And if we put $\alpha=1$ and $t=\frac{1}{N+1}$ then we have

(4.28) $$\mathrm{E}\Big[\int_{R^N} (D_n(x)-D(x))^2 dx\Big] \leqq C \cdot n^{\frac{-1}{N+1}} \quad \text{for all } n \geqq 1$$

and for some constant $C>0$, and for any $\delta>0$ there exists a constant $C(\delta)$ such that

(4.29) $$\mathrm{P}\Big[\int_{R^N} (D_n(x)-D(x)^2 dx \leqq C(\delta) \cdot n^{-\left(1+\frac{1}{N+1}\right)\lambda} \quad \text{for all } n \geqq 1\Big] > 1-\delta$$

for any λ such that $0 < \lambda < \frac{1}{N+2}$.

Moreover, if $f(\cdot|A)$ and $f(\cdot|B)$ have bounded supports then $\{D_n(\cdot)\}_{n+1}^{8}$ is AO$-$I($\{n^{\frac{-1}{2(N+1)}}\}$) and AO$-$II($\{n^{-\frac{\lambda(N+2)}{2(N+1)}}\}$) for any $0 < \lambda < \frac{1}{N+2}$.

If we put $\alpha=1$ and $t=\frac{1}{N+1}$, then $\{D_n(\cdot)|_{n+1}^{\infty}$ is AO$-$I($\{n^{\frac{-1}{N+2}}\}$) and AO$-$II($\{n^{\frac{-\lambda(N+4)}{N+2}}\}$) for any $0 < \lambda < \frac{1}{N+1}$.

§ 5. Acknowledgement.

The author is deeply indebted to Professor T. Kitagawa of Kyushu University, Professor N. Furukawa of Kyushu University and Professor K. Tanaka of Niigata University for their helpful advices and encouragements.

References

[1] E. M. BRAVERMAN and E. S. PYATNITSKII, *Estimation of the rate of convergence of algorithms based on the potential function method.* Automation and Remote Control, **27**, No. 1, (1966), 80-100.

[2] E. M. BRAVERMAN and L. I. ROZONOER, *Convergence of random process in the theory of learning machine, I.* Autom. and Remote Control, **30**, No. 1, (1969), 44-64.

[3] E. M. BRAVERMAN and L. I. ROZONOER, *Convergence of random process in the theory of learning machine, II.* Autom. and Remote Control, **30**, No. 3, (1969), 386-402.

[4] N. V. LOGINOV, *The methods of stochastic approximation.* Autom. and Remote Control, **27**, No. 4, (1966), 706-728.

[5] E. PARZEN, *On the estimation of probability density function and mode.* Ann. Math. Stat., **33** (1962), 1065-1076.

[6] J. V. RYZIN, *Bayes risk consistency of classification procedure using density estimation.* Sankyā, Series A, **28**, (1966), 261-270.

[7] H. SCHEFFE, *A useful convergence theorem for probability distribution.* Ann. Math. Stat., **18** (1947), 434-438.

[8] K. TANAKA, *On the pattern classification by leaning.* Bull. Math. Stat., **14**, No. 3-4, (1971), 15-25.

[9] H. YAMATO, *Sequential estimation of a continuous probability density function and mode.* Bull. Math. Stat., **14**, No. 3-4, (1971), 1-12.

[10] C. T. WOLVERTON and T. J. WAGNER, *Asymptotically optimal discriminant function for pattern classification.* IEEE Trans. on Information Theory, **IT-15**, No. 2, (1969), 258-265.

[11] M. WATANABE, *On asymptotically optimal algorithms for Pattern classification problems.* Bull. Math. Stat., **15**, 3-4, (1973), 31-48.