ON ASYMPTOTICALLY OPTIMAL ALGORITHMS FOR PATTERN CLASSIFICATION PROBLEMS

Watanabe, Masafumi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13070

出版情報：統計数理研究. 15 (3/4), pp.31-48, 1973-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON ASYMPTOTICALLY OPTIMAL ALGORITHMS FOR PATTERN CLASSIFICATION PROBLEMS

By

Masafumi WATANABE*

(Received April 3, 1972)

§ 1. Introduction and Summary.

In recent years considerable interests have been given to the pattern classification problem. This problem includes three main aspects, the engineering aspect, the artificial intelligence aspect and the analytical aspect (c. f. [4]). The analytical one is concerned with the mathematical techniques of decision, estimation and optimization under the uncertainty of information. In this paper, our interest will be concentrated on the analytical aspect, especially on the estimation of a discriminant function which is optimum in the sense of the Bayes rule, minize the probability of misclassification, but which is unknown to us. We shall call such the discriminant function an "optimal discriminant function" (o. d. f.). In previous works (e. g. [4], [7], [8], [9] and [11]), under the given situation of a "training sequence" of observed patterns corectly classified by an external indicator, authors tried to obtain algorithms for finding the o. d. f. on the basis of the training sequence. In general, this approach has been called "*learning with a teacher*" in the pattern classification problem. And the method of approach which we shall appeal to in this paper is this case. The problem which will be treated in this paper is to classify the observed patterns into two categories, and the procedure which will be used is an application of the "stochastic approximation method" introduced by H. Robbins and S. Monro. Our object is to construct, on the basis of the given training sequence, estimates of the o. d. f. which are asymptotically optimal in the sense that the probabilities of misclassification from the estimates converge (with probability one or in the mean) to the probability of misclassification from the o. d. f. if it is known to us.

This paper consists of five sections. In Section 2, we shall give definitions of the optimal discriminant function 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 assumed to belong in the L^2 space, and give an algorithm for constructing the asymptotically optimal estimates. J. V. Ryzin [7] treated this case, but his algorithm was not recursive and his convergence of the asymptotically optimal estimates was the one in the

* Department of Mathematics, Kyushu University, Fukuoka.

mean, whereas in ours the algorithm is recursive and the convergence is the one with probability one as well as in the mean. In Section 4, we shall consider the case when the o. d. f. is assumed to satisfy a uniform Lipschitz condition. Such the case was discussed by T. J. Wagner and C. T. Wolverton [11]. Their algorithm was of the form; $D_{n+1}(x)=D_n(x)+1/(n+1)\cdot[\rho(\theta^{n+1})\cdot K_{n+1}(x, x^{n+1})-D_n(x)]$, where $D_n(x)$ denotes the n-th estimate of the o. d. f.. The coefficient $1/(n+1)$ in their algorithm will be replaced by the more general one in our algorithm. Our convergence will be assured by the conditions weaker than theirs. In Section 4 we also consider the case when the o. d. f. is assumed to be uniformly continuous.

§ 2. The formulation of the problem and Preliminaries.

We consider the two-categories classification problem. Let $\hat{X}$ and Θ denote a pattern space and the set of categories, respectively. In this paper, we assume that $\hat{X}$ is a proper subset of the N-dimensional Euclidian space R^N or equal to R^N, and Θ consists of two categories A and B, i. e. $\Theta=\{A, B\}$.

Let (q_A, q_B) denote a given priori distribution on Θ, where $q_A, q_B>0$ and $q_A+q_B=1$, and let $f_A(x)(f_B(x))$ denote the probability density function of the observed random vector X if X is drawn from the class $A(B)$.

Let assume tentatively $q_A, q_B, f_A(\cdot)$ and $f_B(\cdot)$ are all known to us. Let consider the discriminant function;

$$D^0(x)=\begin{cases}\{q_Af_A(x)-q_Bf_B(x)\}/f(x) & \text{if } f(x)>0\,,\\ 0 & \text{if } f(x)=0\,,\end{cases} \tag{2.1}$$

where $f(x)=q_A\cdot f_A(x)+q_B\cdot f_B(x)$. And let consider the following decision rule based on (2.1); decide X to be from the class A if $D^0(x)\geqq 0$ for the outcome x of X, and decide X from the class B if $D^0(x)<0$ 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.

Let us put

$$D(x)=q_A\cdot f_A(x)-q_B\cdot f_B(x)\,. \tag{2.2}$$

Then it is clear that the Bayes decision rule stated above is equivalent to the following rule; decide X to be from A if $D(x)\geqq 0$, and decide X from B if $D(x)<0$ where x is the outcome of X. In this paper we shall call (2.2) the optimal discriminant function (hereafter, abbreviated as o. d. f.).

In this paper we shall treat the case when $q_A, q_B, f_A(\cdot)$ and $f_B(\cdot)$ are all unknown to us, consequently the o. d. f. is unknown. In this situation we are supposed to have a training sequence $\{(x^n, \theta^n)\}_{n=1}^{\infty}$ with the observed pattern $x^n\in\hat{X}$ and the category $\theta^n\in\Theta$ from which x^n is actually drawn. It is assumed that which category each observed pattern has been drawn from is correctly indicated by a teacher.

We assume $\{(x^n, \theta^n)\}_{n=1}^{\infty}$ is independently and identically distributed, each x^n has a probability density function $f_A(x)(f_B(x))$ if $\theta^n=A$ $(\theta^n=B)$, and each θ^n is distributed as $q_A=P_r\{\theta^n=A\}$ and $q_B=P_r\{\theta^n=B\}$ where $q_A, q_B>0$. Throughout this paper these assumptions remain valid.

Let $D_m(x)$ be an estimate of the o. d. f. $D(x)$ based on the training sequence $\{(x^k, \theta^k)\}_{k=1}^m$. Let $P_g(e)$ be the probability of misclassification using a discriminant function $g(\cdot)$. Here, to classify by using $g(\cdot)$ means that we make use of the decision rule: decide x comes from A if $g(x) \geqq 0$, and x from B if $g(x) < 0$.

DEFINITION 1. Let $\{D_n(x)\}_{n=1}^\infty$ be a sequence of estimates of the o. d. f. $D(x)$. Then, $\{D_n(x)\}_{n=1}^\infty$ is said to be asymptotically optimal of type I (hereafter, abbreviated as AO(I)), if

$$(2.3) \qquad |P_{D_n}(e) - P_D(e)| \longrightarrow 0 \quad \text{with pr. 1 as } n \to \infty,$$

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

$$(2.4) \qquad \mathrm{E}|P_{D_n}(e) - P_D(e)| \longrightarrow 0 \quad \text{as } n \to \infty.$$

DEFINITION 2. Let $\{D_n(x)\}_{n=1}^\infty$ be a sequence of estimates of the o. d. f. $D(x)$, and $\{\alpha_n\}_{n=1}^\infty$ be a sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = 0$. Then, $\{D_n(x)\}_{n=1}^\infty$ is said to be asymptotically optimal of order $\{\alpha_n\}_{n=1}^\infty$ (hereafter, abbreviated as AO($\{\alpha_n\}$)), if

$$(2.5) \qquad \mathrm{E}|P_{D_n}(e) - P_D(e)| \leqq C \cdot \alpha_n \quad \text{for all } n,$$

where C is a positive constant.

REMARK. By the definition of the o. d. f. it holds that $P_g(e) - P_D(e) \geqq 0$ for every discriminant function $g(\cdot)$.

The following lemma can be obtained by a slight modification of Theorem 3 in [11].

LEMMA 2.1. *Let* $\{D_n(x)\}_{n=1}^\infty$ *be a sequence of estimates of the o. d. f.* $D(x)$.

(1) *If* $\int_{\hat{X}} (D_n(x) - D(x))^2 dx \to 0$ *with pr.* 1 *(in the mean) as* $n \to \infty$, *then* $\{D_n(x)\}_{n=1}^\infty$ *is* AO(I) (AO(II)).

(2) *Let* $\{\alpha_n\}_{n=1}^\infty$ *be a positive sequence such that* $\lim_{n\to\infty} \alpha_n = 0$. *And suppose that* $D(\cdot)$ *has a bounded support, and there exists a constant* $C > 0$ *such that* $\mathrm{E}\left[\int_{\hat{X}} (D_n(x) - D(x))^2 dx\right] \leqq C \cdot \alpha_n$ *for all* n. *Then,* $\{D_n(x)\}_{n=1}^\infty$ *is* AO($\{\alpha_n^{\frac{1}{2}}\}$).

Next, we shall give two lemmas to be needed for the convergence of estimations of the o. d. f..

LEMMA 2.2. (*Braverman and Rozonoer* [1]). *Let* $\{y^n\}_{n=1}^\infty$ *be a stochastic process on* R^M, *and let* $\{U_n\}_{n=1}^\infty$, $\{V_n\}_{n=1}^\infty$ *and* $\{\zeta_n\}_{n=1}^\infty$ *be three sequences of non-negative real-valued measurable functions, where* U_n, V_n *and* ζ_n *are defined on* R^{nM} *for each* n. *Let us write* $U^n = U_n(y^1, \cdots, y^n)$, $V^n = V_n(y^1, \cdots, y^n)$ *and* $\zeta^n = \zeta_n(y^1, \cdots, y^n)$. *Let* $\{\gamma_n\}_{n=1}^\infty$ *and* $\{\mu_n\}_{n=1}^\infty$ *be two sequences of real numbers, let* n_0 *be an integer, and suppose the following conditions be satisfied;*

(1) $\mathrm{E}[U^{n_0}] < \infty$ *and* $\mathrm{E}[V^{n_0}] < \infty$,

(2) $\mathrm{E}[U^{n+1} | y^1, \cdots, y^n] \leqq (1+\mu_n) U^n - \gamma_n V^n + \zeta^n$ *for all* $n \geqq n_0$,

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

(4) $\sum_{n=1}^{\infty}|\mu_n|<\infty$ *and* $\sum_{n=1}^{\infty}\mathrm{E}[\zeta^n]<\infty$,

(5) *if there exists a subsequence* $\{n_k\}_{k=1}^{\infty}$ *of* $\{n\}_{n=1}^{\infty}$ *such that* $\lim_{k\to\infty} V^{n_k}=0$ *with pr.* 1, *then* $\lim_{k\to\infty} U^{n_k}=0$ *with pr.* 1.

Then, it holds that

$$\lim_{n\to\infty} U^n=0 \qquad \textit{with pr. } 1.$$

The above lemma is owing to a semimartingal convergence theorem (see, Theorem III of [1] or Lemma 1 of [9]).

LEMMA 2.3. *Let* $\{A_n\}_{n=1}^{\infty}$ *be a non-negative sequence. Suppose that there exist a positive integer* n_0, *three non-negative sequences* $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ *and* $\{K_n\}_{n=1}^{\infty}$, *and a positive constant* K *such that*

$$A_{n+1}\leqq(1-a_{n+1})A_n+K\cdot a_{n+1}\cdot b_{n+1}+K_{n+1} \qquad \textit{for all} \quad n\geqq n_0\,, \tag{2.6}$$

$$\sum_{n=1}^{\infty} a_n=\infty \quad \textit{and} \quad \lim_{n\to\infty} a_n=0\,, \tag{2.7}$$

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

$$\sum_{n=1}^{\infty} K_n<\infty\,. \tag{2.9}$$

Then, it holds that

$$\lim_{n\to\infty} A_n=0\,. \tag{2.10}$$

If $K_n=0$ *for all* n *in the above conditions, and if* $\{b_n\}_{n=1}^{\infty}$ *is a positive sequence such that*

$$(1-a_{n+1})\cdot b_n/b_{n+1}\leqq(1-\alpha_0 a_{n+1}) \qquad \textit{for all} \quad n\geqq n_0 \tag{2.11}$$

where α_0 *is some positive constant (in this case,* $\{b_n\}_{n=1}^{\infty}$ *need not satisfy condition* (2.8)), *then there exists a positive constant* C *such that*

$$A_n\leqq Cb_n \qquad \textit{for all} \quad n\,. \tag{2.12}$$

PROOF. First, we shall prove (2.10). The repeated application (2.6) for integers $m\geqq n_0$ give us

$$\begin{aligned} A_m &\leqq (1-a_m)A_{m-1}+K\cdot a_m\cdot b_m+K_m \\ &\leqq \prod_{k=n_0+1}^{m}(1-a_k)\cdot A_{n_0}+K\cdot\sum_{e=n_0+1}^{m}\prod_{k=e+1}^{m}(1-a_k)\cdot a_e\cdot b_e+\sum_{e=n_0+1}^{m}\prod_{k=e+1}^{m}(1-a_k)\cdot K_e\,. \end{aligned}$$

Hence, we have

$$A_m\leqq F(n_0,\,m)A_{n_0}+K\cdot G(n_0,\,m)+H(n_0,\,m)\,, \tag{2.13}$$

where

$$F(n_0,\,m)=\prod_{k=n_0+1}^{m}(1-a_k),\; G(n_0,\,m)=\sum_{e=n_0+1}^{m}\prod_{k=e+1}^{m}(1-a_k)\cdot a_e\cdot b_e$$

and

$$H(n_0,\,m)=\sum_{k=n_0+1}^{m}\prod_{k=e+1}^{m}(1-a_k)\cdot K_e\,.$$

(2.7) together with $F(n_0, m)=\prod_{e=n_0+1}^{m}(1-a_e)$ implies

(2.14)
$$\lim_{m\to\infty} F(n_0, m)=0\,.$$

Again, (2.7) together with $a_e\cdot F(e, m)=F(e, m)-F(e-1, m)$ implies

$$\lim_{m\to\infty}\sum_{e=n_0+1}^{m} a_e\cdot F(e, m)=1-\lim_{m\to\infty} F(n_0, m)=1\,.$$

By (2.8) therefore we have

(2.15)
$$\lim_{m\to\infty} G(n_0, m)=0\,.$$

Furthermore, noting that for an integer N $(m\geqq N\geqq n_0)$

$$H(n_0, m)=\sum_{e=n_0+1}^{N}\prod_{k=e+1}^{m}(1-a_k)\cdot K_e+\sum_{e=N+1}^{m}\prod_{k=e+1}^{m}(1-a_k)\cdot K_e\,,$$

we have two positive constants C_1 and C_2 for which

$$H(n_0, m)\leqq C_1\cdot\prod_{k=N}^{m}(1-a_k)+C_2\cdot\sum_{e=N+1}^{m}K_e\,.$$

Hence,

(2.16)
$$\lim_{m\to\infty} H(n_0, m)=0\,.$$

By (2.13), (2.14), (2.15) and (2.16), we have

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

Next, we shall prove (2.12). Since $K_n=0$ $(n\geqq n_0)$, puting $Z_n=A_n/b_n$ $(n=1, 2, \cdots)$ yields that

(2.17)
$$Z_m\leqq(1-a_m)\cdot Z_{m-1}\cdot b_{m-1}/b_m+K\cdot a_m \qquad \text{for all} \quad m\geqq n_0\,,$$

and by (2.11) yields that

$$\begin{aligned} Z_m &\leqq (1-a_0\cdot a_m)\cdot Z_{m-1}+K\cdot\alpha_m \\ &\leqq \prod_{k=n_0+1}^{m}(1-\alpha_0\cdot a_k)\cdot Z_{n_0}+K\sum_{e=n_0+1}^{m}\prod_{k=e+1}^{m}(1-\alpha_0\cdot a_k)\cdot\alpha_e \qquad \text{for all} \quad m\geqq n_0\,. \end{aligned}$$

By the same arguments as those in (2.15) and (2.16), we have

(2.18)
$$\lim_{m\to\infty}\prod_{k=n_0+1}^{m}(1-\alpha_0\cdot a_k)=0 \quad \text{and} \quad \lim_{m\to\infty}\sum_{e=n_0+1}^{m}\prod_{k=e+1}^{m}(1-\alpha_0\cdot a_k)\cdot a_e=1/\alpha_0\,.$$

Hence, there exists a positive constant C such that

$$Z_m\leqq C \qquad \text{for all} \quad m\,.$$

Thus the proof of the lemma is completed.

EXAMPLE.

(1) When $K_n=0$, $b_n=1$ for all $n\geqq n_0$ in (2.6), that is, $A_{n+1}\leqq(1-a_{n+1})\cdot A_n+K\cdot a_{n+1}$, the condition (2.11) is satisfied. In this case, from Lemma 2.3, we have $A_n\leqq C$ for all n.

(2) When $a_n = n^{-1}$, $b_n = n^{-\beta}$ $(0<\beta<1)$, the condition (2.11) is satisfied by α_0 such that $0<\alpha_0+\beta<1$.

(3) When $a_n = n^{-\alpha}$, $b_n = n^{-\beta}$ $(0<\alpha, \beta<1)$, the condition (2.11) is satisfied by α_0 such that $0<\alpha_0<1$.

§ 3. L^2-case.

Let $L^2(\hat{X})$ denote the space of all square Lebesgue integrable functions defined on the pattern space $\hat{X}$. Throughout this section, we assume $f_A(\cdot), f_B(\cdot) \in L^2(\hat{X})$. Let $\{\varphi_n(\cdot)\}_{n=1}^{\infty}$ be an orthonormal basis in $L^2(\hat{X})$.

J. V. Ryzin [7] has proved that the sequence $\{D_n(x)\}_{n=1}^{\infty}$ constructed on the basis of the training sequence was asymptotically optimal in our sense, under the assumption that $f_A(\cdot), f_B(\cdot) \in L^2(\hat{X})$. In his work, q_A and q_B were known and his algorithm was not recursive, and the convergence was the one in the mean or in probability. In this section, we shall construct a recursive algorithm in which the convergence is the one not only in the mean but also with probability 1. Moreover, we assume that q_A and q_B are unknown.

Let $\Gamma(\cdot)$ be any non-decreasing non-negative real-valued function on $[0, \infty)$ and $[\Gamma(\cdot)]$ be the integral part of $\Gamma(\cdot)$. Then we define the sequence $\{K_n(x, y)\}_{n=1}^{\infty}$ by

$$K_n(x, y) = \sum_{i=1}^{[\Gamma(n)]} \varphi_i(x)\cdot\varphi_i(y) \qquad \text{for all} \quad x, y \in \hat{X} \tag{3.1}$$

and for all n, where $\{\varphi_i(\cdot)\}_{n=1}^{\infty}$ is an orthonormal basis in $L^2(\hat{X})$.

Let $\{(x^n, \theta^n)\}_{n=1}^{\infty}$ be a training sequence which was defined in § 2. Then we define $\{\rho^n\}_{n=1}^{\infty}$ by

$$\begin{aligned} \rho^n \equiv \rho(\theta^n) &= 1 && \text{if} \quad \theta^n = A \\ &\coloneqq -1 && \text{if} \quad \theta^n = B \quad \text{for all} \quad n. \end{aligned} \tag{3.2}$$

By (3.1) and (3.2), we have

$$\begin{aligned} \mathrm{E}[\rho^n\cdot K_n(x, x^n)] &= \int_{\hat{X}} K_n(x, y)\cdot D(y)\,dy \\ &= \sum_{i=1}^{[\Gamma(n)]} C_i\cdot\varphi_i(x) \qquad \text{for all} \quad x \in \hat{X} \end{aligned} \tag{3.3}$$

and for all n, where $C_i = \int_{\hat{X}} D(x)\cdot\varphi_i(x)\,dx$ $i=1, 2, \cdots$. And we put $D_n^*(x) = \mathrm{E}[\rho^n\cdot K_n(x, x^n)]$ for each n.

In view of the above arguments, we shall construct the following algorithm which is an application of the dynamic stochastic approximation method (V. Dupač [3]).

LEARNING ALGORITHM. Let $\{a_n\}_{n=1}^{\infty}$ be a positive sequence such that

$$\lim_{n\to\infty} a_n = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} a_n = \infty . \tag{3.4}$$

Then $\{D_n(x)\}_{n=1}^{\infty}$ is given by the recursive relation as follows:

$$D_0(x) \equiv 0 , \qquad x \in X$$

(3.5) $$D_{n+1}(x)=D_n(x)+a_{n+1}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_n(x)), \qquad x\in\hat{X}$$

and $n=0, 1, 2, \cdots$.

The above algorithm is transformed to the following one;

$$C_i^0\equiv 0 \qquad i=1, 2, \cdots, [\Gamma(0)]$$

(3.6) $$\begin{aligned} C_i^{n+1}&=C_i^n+a_{n+1}(\rho^{n+1}\cdot\varphi_i(x^{n+1})-C_i^n) && \text{for } i=1, 2, \cdots, [\Gamma(n)], \\ &=a_{n+1}\cdot\rho^{n+1}\cdot\varphi_i(x^{n+1}) && \text{for } i=[\Gamma(n)]+1, \cdots, [\Gamma(n+1)]. \end{aligned}$$

Put

$$\tilde{D}_n(x)=\sum_{i=1}^{[\Gamma(n)]} C_i^n\cdot\varphi_i(x) \qquad n=1, 2, \cdots,$$

then we have $\tilde{D}_n(x)=D_n(x)$ for $n=1, 2, \cdots$. Thus (3.5) is equivalet to (3.6).

REMARK.

(1) Coefficients $\{C_i\}_{i=1}^{[\Gamma(n)]}$ in $D_n^*(x)$ minimize the quantity $\Phi(\alpha_1, \cdots, \alpha_{[\Gamma(n)]})=\int_{\hat{X}}(D(x)-\sum_{i=1}^{[\Gamma(n)]}\alpha_i\cdot\varphi_i(x))^2dx$ for each n. Therefore $D_n^*(x)$ is a kind of approximation to $D(x)$ for each n. And by the orthonormality of $\{\varphi_i(\cdot)\}_{i=1}^{\infty}$, we have

(3.7) $$D_{n+1}^*(x)=D_n^*(x)+\sum_{i=[\Gamma(n)]+1}^{[\Gamma(n+1)]} C_i\cdot\varphi_i(x) \qquad \text{for } x\in\hat{X}$$

and $n=1, 2, \cdots$ where $C_i=\int_{\hat{X}}D(x)\cdot\varphi_i(x)dx$ $i=1, 2, \cdots$.

(2) If $\Gamma(t)\uparrow\infty$ as $t\to\infty$, then $\int_{\hat{X}}(D_n^*(x)-D(x))^2dx=\sum_{i=[\Gamma(n)]+1}^{\infty}C_i^2\downarrow 0$ as $n\to\infty$, which implies $\{D_n^*(x)\}_{n=1}^{\infty}$ is a sequence of approximations of $D(x)$. Thus, from the property of $\{D_n^*(x)\}_{n=1}^{\infty}$ stated in (1), our algorithm is considered as an application of the dynamic stochastic approximation method.

In what follows, all integrals are interpreted Lebesgue integral on $\hat{X}$ ($\hat{X}\subseteqq R^N$). And the Fubini's theorem is invoked without any comment for proofs of lemmas and theorems in this section and next section.

THEOREM 3.1. *Let the following conditions be satisfied; there exists a positive constant K such that*

(3.8) $$\int_{\hat{X}}\varphi_n^2(x)\cdot f(x)dx\leqq K \qquad \textit{for all} \quad n$$

where $f(x)=q_A\cdot f_A(x)+q_B\cdot f_B(x)$. Let $\{a_n\}_{n=1}^{\infty}$ be a positive sequence statisfying (3.4).

(1) *If*

(3.9) $$\lim_{n\to\infty} a_n\cdot\Gamma(n)=0,$$

then it holds that

(3.10) $$\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx\Big]\longrightarrow 0 \qquad \textit{as } n\to\infty.$$

(2) *If*

(3.11) $$\sum_{n=1}^{\infty}a_n^2\cdot\Gamma(n)<\infty \quad \textit{and} \quad \Gamma(t)>0 \quad \textit{for some} \quad t\geqq 0,$$

then it holds that

$$(3.12)\qquad \int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx \longrightarrow 0 \quad \textit{with pr. 1. as } n\to\infty .$$

PROOF. First we shall prove (2). By the construction of $D_n(x)$, we have

$$\begin{aligned}(3.13)\qquad &\int_{\hat{X}}(D_{n+1}(x)-D_{n+1}^*(x))^2dx\\ &=(1-a_{n+1})^2\cdot\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx\\ &+a_{n+1}^2\cdot\int_{\hat{X}}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*)^2dx\\ &+(1-a_{n+1})^2\cdot\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx\\ &+2(1-a_{n+1})\cdot a_{n+1}\cdot\int_{\hat{X}}(D_n(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*(x))dx\\ &+2(1-a_{n+1})^2\cdot\int_{\hat{X}}(D_n(x)-D_n^*(x))\cdot(D_n^*(x)-D_{n+1}^*(x))dx\\ &+2(1-a_{n+1})\cdot a_{n+1}\cdot\int_{\hat{X}}(D_n^*(x)-D_{n+1}^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*(x))dx .\end{aligned}$$

Noting that

$$\begin{aligned}\int_{\hat{X}}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*(x))^2dx &\leqq 2\int_{\hat{X}}K_{n+1}^2(x, x^{n+1})dx+2\int_{\hat{X}}(D_{n+1}^*(x))^2dx\\ &=2\sum_{i=1}^{[\Gamma(n+1)]}\varphi_i^2(x^{n+1})+2\sum_{i=1}^{[\Gamma(n+1)]}C_i^2 ,\end{aligned}$$

by (3.8) we have

$$\begin{aligned}(3.14)\qquad &\mathrm{E}\Big[\int_{\hat{X}}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*(x))^2dx\Big]\\ &\leqq 2[\Gamma(n+1)]\cdot K+2\|D\|^2\\ &\leqq 2K\cdot\Gamma(n+1)+2\|D\|^2 , \qquad \text{where } \|D\|^2=\int_{\hat{X}}D^2(x)dx .\end{aligned}$$

Next, we have

$$(3.15)\qquad \mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*)dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big]=0 ,$$

$$(3.16)\qquad \mathrm{E}\Big[\int_{\hat{X}}(D_n^*(x)-D_{n+1}^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D_{n+1}^*(x))dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big]=0 .$$

And by the construction of $D_n(x)$, we have

$$(3.17)\qquad \int_{\hat{X}}(D_n(x)-D_n^*(x))\cdot(D_{n+1}^*(x)-D_n^*(x))dx=0 .$$

On taking the conditional expectation on both sides of (3.13), by (3.14), (3.15), (3.16), (3.4) and (3.11) we have, there exists n_0 such that

$$\begin{aligned}(3.18)\qquad &\mathrm{E}\Big[\int_{\hat{X}}(D_{n+1}(x)-D_{n+1}^*(\theta))^2dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big]\\ &=(1-a_{n+1})\cdot\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx+2K\cdot a_{n+1}^2\cdot\Gamma(n+1)\end{aligned}$$

$$+2a_{n+1}^2\cdot\|D\|^2+\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx \qquad \text{for all} \quad n\geqq n_0\,.$$

Since

$$\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx=\sum_{i=[\iota(n)]+1}^{[\Gamma(n+1)]}C_i^2\,,$$

we have

$$\sum_{n=1}^{\infty}\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx\leqq\|D\|^2\,.$$

In (3.18) put

$$\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx=U^n\,, \qquad a_n=\gamma_n$$

and

$$2K\cdot a_{n+1}^2\cdot\Gamma(n+1)+2a_{n+1}^2\cdot\|D\|^2+\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx=\zeta^n\,,$$

then these quantities satisfy the conditions Lemma 2.2 where $\mu_n=0$ and $U^n\equiv V^n$. Therefore, (2) is a direct application of Lemma 2.2.

Next, we shall prove (1). Taking the expectation on both sides of (3.18), we have

$$\text{(3.19)}\qquad \begin{aligned} \mathrm{E}\Big[\int_{\hat{X}}(D_{n+1}(x)-D_{n+1}^*(x))^2dx\Big]&=(1-a_{n+1})\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx\Big]\\ &\quad+2K_1\cdot a_{n+1}\cdot(a_{n+1}\cdot\Gamma(n+1)+a_{n+1})\\ &\quad+\int_{\hat{X}}(D_{n+1}^*(x)-D_n^*(x))^2dx \qquad \text{for all} \quad n\geqq n_0\,, \end{aligned}$$

where $K_1=\max\{K, \|D\|^2\}$. In (3.19), we put

$$\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx\Big]=A_n\,, \qquad a_n\cdot\Gamma(n)+a_n=b_n$$

and

$$\int_{\hat{X}}(D_n^*(x)-D_{n+1}^*(x))^2dx=K_n\,,$$

then these quantities satisfy the conditions of Lemma 2.3. Therefore, (1) is a direct application of Lemma 2.3.

Thus the proof of Theorem 3.1 is completed.

COROLLARY 3.1.1. *Let* (3.8) *be satisfied. Let* $\{a_n\}_{n=1}^{\infty}$ *be a positive sequence satisfying* (3.4) *and let* $\Gamma(t)\uparrow\infty$ *as* $t\to\infty$.

(1) *If* (3.9) *holds, then* $\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big]\to 0$ *as* $n\to\infty$.

(2) *If* (3.11) *holds, then* $\int_{\hat{X}}(D_n(x)-D(x))^2dx\to 0$ *with pr.* 1 *as* $n\to\infty$.

PROOF. We can easily prove the corollary by Theorem 3.1 and the inequality

$$\int_{\hat{X}}(D_n(x)-D(x))^2dx\leqq 2\int_{\hat{X}}(D_n(x)-D_n^*(x))^2dx+2\int_{\hat{X}}(D_n^*(x)-D(x))^2dx\,.$$

COROLLARY 3.1.2. *Let the conditions in Corollary* 3.1.1 *be satisfied.*

(1) *If* (3.11) *holds, then* $\{D_n(x)\}_{n=1}^{\infty}$ *is* AO(I).

(2) *If* (3.9) *holds, then* $\{D_n(x)\}_{n=1}^{\infty}$ *is* AO(II).

PROOF. It is easily proved by Lemma 2.1 and Corollary 3.1.1.

Next, we shall discuss the order of the mean convergence. By the assumption that $D(\cdot)\in L^2(\hat{X})$, it is easily seen that $\sum_{i=1}^{\infty} C_i^2 \downarrow 0$ as $n\to\infty$, where $C_i=\int_{\hat{X}} D(x)\cdot\varphi_i(x)dx$ $(i=1, 2, \cdots)$. Therefore we have the following theorem.

THEOREM 3.2. *Let* (3.8) *be satisfied. Let the following conditions be satisfied;*

$$\sum_{i=m}^{\infty} C_i^2 \leqq K'\cdot g(m) \qquad \textit{for all} \quad m\,, \tag{3.20}$$

where $C_i=\int_{\hat{X}} D(x)\cdot\varphi_i(x)dx$ $(i=1,2,\cdots)$, K' *is some positive constant, and* $g(\cdot)$ *is a non-increasing positive function on* $[0, \infty)$. *For a positive sequence* $\{a_n\}_{n=1}^{\infty}$ *satisfying* (3.4), *there exist a positive integer* n_0 *and a positive constant* α_0 *such that*

$$(1-a_{n+1})\cdot b_n/b_{n+1} \leqq (1-\alpha_0\cdot a_{n+1}) \qquad \textit{for all} \quad n\geqq n_0\,, \tag{3.21}$$

where $b_n=\max\,\{a_n\cdot\Gamma(n),\, a_n,\, g([\Gamma(n)])\}$ *for* $n=1,2,\cdots$. *Then, there exists some positive constant* C *satisfying*

$$\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big] \leqq C\cdot b_n \qquad \textit{for all} \quad n\,. \tag{3.22}$$

PROOF. By the construction of $D_n(x)$, we have

$$\begin{aligned}
&\int_{\hat{X}}(D_{n+1}(x)-D(x))^2dx \\
&\quad =(1-a_{n+1})^2\cdot\int_{\hat{X}}(D_n(x)-D(x))^2dx + a_{n+1}^2\cdot\int_{\hat{X}}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))^2dx \\
&\quad\quad +2(1-a_{n+1})\cdot a_{n+1}\cdot\int_{\hat{X}}(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))dx\,.
\end{aligned} \tag{3.23}$$

And we have

$$\begin{aligned}
&\mathrm{E}\Big[\int_{\hat{X}}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))^2dx\Big] \\
&\quad \leqq 2\mathrm{E}\Big[\int_{\hat{X}}K_{n+1}^2(x, x^{n+1})dx\Big]+2\|D\|^2 \\
&\quad \leqq K_1\cdot(\Gamma(n+1)+1)\,, \qquad \text{where} \quad K_1=\max\,\{2K, 2\|D\|^2\}\,.
\end{aligned} \tag{3.24}$$

(3.20) yields that

$$\begin{aligned}
&\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big] \\
&\quad =\int_{\hat{X}}(D_n(x)-D(x))\cdot(D_{n+1}^*(x)-D(x))dx \\
&\quad =\int_{\hat{X}}(D_n(x)-D_{n+1}^*(x))\cdot(D_{n+1}^*(x)-D(x))dx+\int_{\hat{X}}(D_{n+1}^*(x)-D(x))^2dx \\
&\quad =\int_{\hat{X}}(D_{n+1}^*(x)-D(x))^2dx \\
&\quad =\sum_{i=[\Gamma(n+1)]+1}^{\infty} C_i^2 \leqq K'\cdot g([\Gamma(n+1)]) \qquad \text{for all} \quad n\,.
\end{aligned} \tag{3.25}$$

Taking the expectation on both sides of (3.23), from (3.23), (3.24) and (3.25) we have, there exists some integer n_1 such that

$$(3.26)\qquad \mathrm{E}\Big[\int_{\hat{X}}(D_{n+1}(x)-D(x))^2dx\Big]\leqq(1-a_{n+1})\cdot\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big]$$
$$+K_1\cdot a_{n+1}^2\cdot\Gamma(n+1)+K_1\cdot a_{n+1}^2+2K'\cdot a_{n+1}\cdot g([\Gamma(n+1)])$$
$$\text{for all}\quad n\geqq n_1\,,$$

which implies

$$(3.27)\qquad \mathrm{E}\Big[\int_{\hat{X}}(D_{n+1}(x)-D(x))^2dx\Big]\leqq(1-a_{n+1})\cdot\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big]$$
$$+K_2\cdot a_{n+1}\cdot b_{n+1}\qquad\text{for all}\quad n\geqq n_1\,,$$

where $K_2=\max\{K_1, 2K'\}$ and $b_n=\max\{a_n\cdot\Gamma(n), a_n, g([\Gamma(n)])\}$. Putting $\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big]=A_n$ in (3.27), the theorem is easily derived from Lemma 2.3.

COROLLARY 3.2.1. *If the conditions in Theorem* 3.2 *and* (3.9) *are satisfied, if* $\Gamma(t)\uparrow\infty$, $g(t)\downarrow 0$ *as* $t\to\infty$, *and if* $f_A(\cdot)$ *and* $f_B(\cdot)$ *have bounded supports, then* $\{D_n(x)\}_{n=1}^{\infty}$ *is* AO($\{\tilde{b}_n\}$) *where* $\tilde{b}_n=[\max\{a_n\cdot\Gamma(n), a_n, g([\Gamma(n)])\}]^{\frac{1}{2}}$.

PROOF. This corollary is obvious from Theorem 3.2 and Lemma 2.1, (2).

EXAMPLE. When $g(t)=t^{-\alpha}$ $(\alpha>0)$, we choose the sequence $\{a_n\}_{n=1}^{\infty}$ and the function $\Gamma(t)$ by the following way;

$$a_n=1/n\ (n=1, 2, \cdots)\quad\text{and}\quad \Gamma(t)=t^{\frac{1}{1+\alpha}}\ \Big(\text{satisfying, } \Gamma\Big(\frac{t}{g(t)}\Big)=t\Big).$$

Then, we have

$$\mathrm{E}\Big[\int_{\hat{X}}(D_n(x)-D(x))^2dx\Big]\geqq C\cdot n^{-\alpha/(1+\alpha)}\qquad\text{for all}\quad n\,.$$

Therefore $\{D_n(x)\}_{n=1}^{\infty}$ is AO($\{n^{-\alpha/(1+\alpha)}\}$), under the assumption that $f_A(\cdot)$ and $f_B(\cdot)$ have bounded supports.

§ 4. Uniform Continuous case.

In this section, we assume that the density functions $f_A(\cdot)$ and $f_B(\cdot)$ are uniformly continuous on R^N. And let $\hat{X}$ be equal to R^N.

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

$$(4.1)\qquad K(y)\geqq 0\qquad\text{for all}\quad y\in R^N\,,$$

$$(4.2)\qquad \sup_{y\in R^N}K(y)=K<\infty\,,$$

$$(4.3)\qquad \int_{R^N}K(y)dy=1\,,$$

$$(4.4)\qquad \int_{R^N}\|y\|\cdot K(y)dy=K^*\qquad\text{where}\quad \|y\|=\Big(\sum_{i=1}^{N}y_i^2\Big)^{\frac{1}{2}}\quad\text{for } y=(y_1,\cdots,y_N)\in R^N\,.$$

C. T. Wolverton and T. J. Wagner ([11]) and J. V. Ryzin ([7], [8]) used this function for a pattern classification problem. In this section, our method is due to Wolverton and Wagner.

Let $\{h_n\}_{n=1}^{\infty}$ be a positive sequence satisfying

$$1 \geqq h_1 \geqq h_2 \geqq \cdots \geqq h_n \geqq \cdots, \tag{4.5}$$

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

$$K_n(x, y) = h_n^{-N} \cdot K[h_n^{-1} \cdot (x-y)] \qquad \text{for all} \quad x, y \in R^N \text{ and } n = 1, 2, \cdots. \tag{4.6}$$

The following lemma can be obtained by a slight modification of Lemma 1 and Lemma 4 in [11].

LEMMA 4.1. *Let $g(\cdot)$ be a uniformly continuous function on R^N and $\sup_{y \in R^N} |g(y)| = M < \infty$. Let $K(\cdot)$ be measurable function on R^N satisfying* (4.1), (4.2) *and* (4.3), *and let $\{h_n\}_{n=1}^{\infty}$ be a positive sequence such that $\lim_{n\to\infty} h_n = 0$. Define the function $g_n(\cdot)$ by*

$$g_n(x) = \int_{R^N} K_n(x, y) \cdot g(y) dy \qquad \textit{for all} \quad x \in R^N \textit{ and } n = 1, 2, \cdots. \tag{4.7}$$

where $K_n(x, y)$ is defined by (4.6). *Then it holds that*

$$g_n(x) \xrightarrow{n} g(x) \qquad \textit{uniformly in } x.$$

(2) *Let $g(\cdot)$ be a real-valued measurable function on R^N satisfying a uniform Lipschitz condition;*

$$|g(x) - g(y)| \leqq C_g \cdot \|x - y\| \qquad \textit{for all} \quad x, y \in R^N.$$

Let $K(\cdot)$ be a real-valued measurable function on R^N satisfying (4.1), (4.2), (4.3) *and* (4.4), *and let $\{h_n\}_{n=1}^{\infty}$ be a positive sequence satisfying* (4.5). *Define $g_n(x)$ by* (4.7). *Then it holds that*

$$|g_n(x) - g_{n+1}(x)| \leqq C_g \cdot K^* \cdot (h_n - h_{n+1}) \qquad \textit{for all} \quad x \in R^N \textit{ and } n = 1, 2, \cdots,$$

and

$$|g_n(x) - g(x)| \leqq C_g \cdot K^* \cdot h_n \qquad \textit{for all} \quad x \in R^N \textit{ and } n = 1, 2, \cdots.$$

If $f_A(\cdot)$ and $f_B(\cdot)$ are uniformly continuous functions, then $D(x)$ is bounded and uniformly continuous on R^N.

And we have

$$\mathrm{E}[\rho^n \cdot K_n(x, x^n)] = \int_{R^N} K_n(x, y) \cdot D(y) dy \qquad \text{for} \quad n = 1, 2, \cdots, \tag{4.8}$$

where ρ^n is defined by (3.2). And we put $\mathrm{E}[\rho^n \cdot K_n(x, x^n)] = D_n^*(x)$.

In view of above arguments, we now construct the following algorithm which is an application of dynamic stochastic approximations.

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

$$\lim_{n\to\infty} a_n = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} a_n = \infty. \tag{4.9}$$

Then $\{D_n(x)\}_{n=1}^{\infty}$ is given by the recursive relation as follows:

$$D_0(x) \equiv 0, \qquad x \in R^N,$$

$$D_{n+1}(x) = D_n(x) + a_{n+1} \cdot (\rho^{n+1} \cdot K_{n+1}(x, x^{n+1}) - D_n(x)), \tag{4.10}$$

$$x \in R^N \text{ and } n = 0, 1, 2, \cdots.$$

As for the above algorithm we have two main theorems. First, we shall consider the case when $f_A(\cdot)$ and $f_B(\cdot)$ satisfy the uniform Lcpschitz conditions.

THEOREM 4.1. *Let $K(\cdot)$ be a real-valued measurable function on R^N satisfying* (4.1), (4.2), (4.3) *and* (4.4), *and let $\{h_n\}_{n=1}^{\infty}$ be a positive sequence satisfying* (4.5). *Let $f_A(\cdot)$ and $f_B(\cdot)$ satisfy the uniform Lipschitz conditions with constants C_A and C_B, respectively. Let $\{a_n\}_{n=1}^{\infty}$ be a positive sequence satisfying* (4.9).

(1) *If $\lim_{n\to\infty} a_n\cdot h_n^{-N}=0$, then $\mathrm{E}\Big[\int_{R^N}(D_n(x)-D_n^*(x))^2dx\Big]\to 0$ as $n\to\infty$.*

(2) *If $\sum_{n=1}^{\infty} a_n^2\cdot h_n^{-N}<\infty$, then $\int_{R^N}(D_n(x)-D_n^*(x))^2dx\to 0$ with pr. 1 as $n\to\infty$.*

PROOF. First, we shall prove (2). By (4.10), we have

$$
\begin{aligned}
(4.11)\qquad & \int_{R^N}(D_{n+1}(x)-D_{n+1}^*(x))^2dx \\
&= (1-a_{n+1})^2\cdot\int_{R^N}(D_n(x)-D_n^*(x))^2dx \\
&\quad + a_{n+1}^2\cdot\int_{R^N}(\rho^{n+1}\cdot K_{n+1}(x,\, x^{n+1})-D_{n+1}^*(x))^2dx \\
&\quad +(1-a_{n+1})^2\cdot\int_{R^N}(D_{n+1}^*(x)-D_n^*(x))^2dx \\
&\quad +2(1-a_{n+1})\cdot a_{n+1}\cdot\int_{R^N}(D_n(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x,\, x^{n+1})-D_{n+1}^*(x))dx \\
&\quad +2(1-a_{n+1})^2\cdot\int_{R^N}(D_n(x)-D_n^*(x))\cdot(D_{n+1}^*(x)-D_n^*(x))dx \\
&\quad +2(1-a_{n+1})\cdot a_{n+1}\cdot\int_{R^N}(D_{n+1}^*(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x,\, x^{n+1})-D_{n+1}^*(x))dx\,.
\end{aligned}
$$

Noting that $f_A(\cdot)$ and $f_B(\cdot)$ are bounded so that $D_n^*(x)$ is bounded, we have

$$
\begin{aligned}
(4.12)\qquad & \int_{R^N}(\rho^{n+1}\cdot K_{n+1}(x,\, x^{n+1})-D_{n+1}^*(x))^2dx \\
&\leqq 2\int_{R^N}K_{n+1}^2(x,\, x^{n+1})dx+2\int_{R^N}(D_{n+1}^*(x))^2dx \\
&\leqq 2K\cdot h_{n+1}^{-N}+2M \qquad \text{for all} \quad n\,,
\end{aligned}
$$

where $M=\sup_x f(x)$ for $f(x)=q_A\cdot f_A(x)+q_B\cdot f_B(x)$. By Lemma 4.2, we have

$$
(4.13)\qquad \int_{R^N}(D_{n+1}^*(x)-D_n^*(x))^2dx\leqq C_D\cdot K^*\cdot(h_n-h_{n+1}) \qquad \text{for all} \quad n\,,
$$

where $C_D=\max\{C_A,\, C_B\}$, and we have

$$
\begin{aligned}
(4.14)\qquad & \int_{R^N}(D_n(x)-D_n^*(x))\cdot(D_{n+1}^*(x)-D_n^*(x))dx \\
&\leqq \Big(\int_{R^N}|D_n(x)|\,dx+1\Big)\cdot C_D\cdot K^*\cdot(h_n-h_{n+1}) \\
&= C_D\cdot K^*\cdot(h_n-h_{n+1})\cdot\int_{R^N}|D_n(x)|\,dx+C_D\cdot K^*\cdot(h_n-h_{n+1})\,.
\end{aligned}
$$

And by (4.8), we have

$$\text{(4.15)} \qquad \mathrm{E}\Big[\int_{R^N}(D_n(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1}) \\ -D_{n+1}^*(x))dx\,|\,(x^1, \theta), \cdots, (x^n, \theta^n)\Big]=0\,,$$

and

$$\text{(4.16)} \qquad \mathrm{E}\Big[\int_{R^N}(D_{n+1}^*(x)-D_n^*(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1}) \\ -D_{n+1}^*(x))dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big]=0\,.$$

On taking the conditional expectation on both sides of (4.11), by (4.12), (4.13), (4.14), (4.15) and (4.16) we have, there exists n_0 such that

$$\text{(4.17)} \qquad \mathrm{E}\Big[\int_{R^N}(D_{n+1}(x)-D_{n+1}^*(x))^2dx\,|\,(x^1, \theta^1), \cdots, (x^n, \theta^n)\Big] \\ \leqq(1-a_{n+1})\cdot\int_{R^N}(D_n(x)-D_n^*(x))^2dx+K^1\cdot a_{n+1}^2\cdot h_{n+1}^{-N} \\ +2C_D\cdot K^*\cdot(h_n-h_{n+1})+2C_D\cdot K^*\cdot(h_n-h_{n+1})\cdot\int_{R^N}|D_n(x)|\,dx \\ \text{for all} \quad n\geqq n_0\,,$$

where $K^1=\max\{2K, M\}$. Noting that

$$\text{(4.18)} \qquad \mathrm{E}\Big[\int_{R^N}|D_{n+1}(x)dx\Big]\leqq(1-a_{n+1})\cdot\mathrm{E}\Big[\int_{R^N}|D_n(x)|\,dx\Big]+a_{n+1} \qquad \text{for} \quad n=1, 2, \cdots,$$

and by Example (1) of Lemma 2.3 we have, there exists a positive constant D_0 such that

$$\text{(4.19)} \qquad \mathrm{E}\Big[\int_{R^N}|D_n(x)|\,dx\Big]\leqq D_0 \qquad \text{for all} \quad n\,.$$

In (4.18), put

$$U^n=\int_{R^N}(D_n(x)-D_n^*(x))^2dx\,, \qquad U^n=V^n\,, \qquad \gamma_n=a_n$$

and

$$\zeta^{n+1}=K'\cdot a_{m+1}\cdot h_{n+1}^{-N}+2C_D\cdot K^*\cdot(h_n-h_{n+1})+2C_D\cdot K^*\cdot(h_n-h_{n+1})\cdot\int_{R^N}|D_n(x)|\,dx\,,$$

then these quantities satisfy the conditions of Lemma 2.2. Therefore (2) is a direct application of Lemma 2.2.

Next, we shall prove (1). Taking the expectation on both sides of (4.17) and by (4.19), we have

$$\text{(4.20)} \qquad \mathrm{E}\Big[\int_{R^N}(D_{n+1}(x)-D_{n+1}^*(x))^2dx\Big] \\ \leqq(1-a_{n+1})\cdot\mathrm{E}\Big[\int_{R^N}(D_n(x)-D_n^*(x))^2dx\Big] \\ +K'\cdot a_{n+1}^2\cdot h_{n+1}^{-N}+K''\cdot(h_n-h_{n+1}) \qquad \text{for all} \quad n\geqq n_0\,,$$

where $K''=\max\{2C_D\cdot K^*, 2C_D\cdot K^*\cdot D_0\}$. In (4.20), put

$$\mathrm{E}\Big[\int_{R^N}(D_n(x)-D_n^*(x))^2dx\Big]=A_n\,,\qquad b_n=a_n\cdot h_n^{-N}\ \text{ and }\ (h_{n-1}-h_n)=K_n\,,$$

then these quantities satisfy the conditions of Lemma 2.3.

Thus the proof of the theorem is completed.

COROLLARY 4.1.1. *Let the conditions in Theorem* 4.1 *be satisfied and* $\lim_{n\to\infty} h_n=0$.

(1) *If* $\lim_{n\to\infty} a_n\cdot h_n^{-N}=0$, *then* $\mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))^2dx\Big]\to 0$ *as* $n\to\infty$.

(2) *If* $\sum_{n=1}^{\infty} a_n^2\cdot h_n^{-N}<\infty$, *then* $\int_{R^N}(D_n(x)-D(x))^2dx\to 0$ *with pr.* 1 *as* $n\to\infty$.

PROOF. We can easily prove the corollary by Theorem 4.1, (2) of Lemma 4.1 and the inequality,

$$\int_{R^N}(D_n(x)-D(x))^2dx\leqq 2\int_{R^N}(D_n(x)-D_n^*(x))^2dx+2\int_{R^N}(D_n^*(x)-D(x))^2dx\,.$$

COROLLARY 4.1.2. *Let the conditions in Corollary* 4.1.1 *be satisfied.*

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

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

PROOF. It is easily proved by Lemma 2.1 and Corollary 3.1.1.

REMARK. In [11], Wolverton and Wagner constructed the following algorithm,

$$D_{n+1}(x)=D_n(x)+(n+1)^{-1}\cdot(\rho^{n+1}\cdot K_{n+1}(x,\,x^{n+1})-D_n(x))\,. \tag{4.21}$$

In their work, in the case when $D(x)$ satisfies the uniform Lipschitz condition the convergence with probability 1 was assured by

$$\sum_{n=1}^{\infty} h_n\cdot n^{-1}<\infty \qquad\text{and}\qquad \sum_{n=1}^{\infty} h_n^{-2N}\cdot n^{-2}<\infty\,, \tag{4.22}$$

and some additional conditions same as in ours.

Next, we shall consider the case when the o. d. f. is assumed to be uniformly continuous. In this case, we shall prove that $\{D_n(x)\}_{n=1}^{\infty}$ is AO(II).

THEOREM 4.2. *Let* $f_A(\cdot)$ *and* $f_B(\cdot)$ *be uniformly continuous functions on* R^N, *and let* $K(\cdot)$ *be a real-valued measurable function on* R^N *satisfying* (4.1), (4.2) *and* (4.3). *Let* $\{h_n\}_{n=1}^{\infty}$ *be a positive sequence satisfying* (4.5) *and*

$$\lim_{n\to\infty} h_n=0\,,\qquad \lim_{n\to\infty} a_n\cdot h_n^{-N}=0 \tag{4.23}$$

where $\{a_n\}_{n=1}^{\infty}$ *is a positive sequence satisfying* (4.9). *Then it holds that*

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

PROOF. By the construction of $D_n(x)$, we have

$$\begin{aligned}&\int_{R^N}(D_{n+1}(x)-D(x))^2dx\\&\quad=(1-a_{n+1})^2\cdot\int_{R^N}(D_n(x)-D(x))^2dx\\&\qquad+a_{n+1}^2\cdot\int_{R^N}(\rho^{n+1}\cdot K_{n+1}(x,\,x^{n+1})-D(x))^2dx\end{aligned} \tag{4.24}$$

$$+2a_{n+1}\cdot(1-a_{n+1})\cdot\int_{R^N}(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))dx\,.$$

By the arguments similar to (4.12), we have

$$\mathrm{E}\Big[\int_{R^N}(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))^2dx\Big]\leqq 2K\cdot h_{n+1}^{-N}+2M \qquad \text{for all } n\,. \tag{4.25}$$

And

$$\begin{aligned}\mathrm{E}\Big[\int_{R^N}&(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))dx\Big]\\ &=\mathrm{E}\Big[\mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})\\ &\qquad\qquad -D(x))dx\,|\,(x_1^1, \theta^1), \cdots, (x^n, \theta^n)\Big]\Big]\\ &=\mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))\cdot(D_{n+1}^*(x)-D(x))\,dx\Big]\\ &\leqq\int_{R^N}\mathrm{E}[\,|D_n(x)-D(x)|\,]\cdot|D_{n+1}^*(x)-D(x)|\,dx\,.\end{aligned} \tag{4.26}$$

By (4.19) and the definition of o. d. f. $D(x)$, we have

$$\int_{R^N}\mathrm{E}[\,|D_n(x)-D(x)|\,]dx\leqq M'<\infty \qquad \text{for all } n\,, \tag{4.27}$$

where M' is some positive constant. Therefore,

$$\begin{aligned}\mathrm{E}\Big[\int_{R^N}&(D_n(x)-D(x))\cdot(\rho^{n+1}\cdot K_{n+1}(x, x^{n+1})-D(x))dx\Big]\\ &\leqq\sup_{x\in R^N}|D_{n+1}^*(x)-D(x)|\cdot M' \qquad \text{for all } n\,.\end{aligned} \tag{4.28}$$

On taking the expectation on both sides of (4.24), by (4.25) and (4.28), there exists a integer n_0 such that

$$\begin{aligned}\mathrm{E}\Big[\int_{R^N}&(D_{n+1}(x)-D(x))^2dx\Big]\\ &\leqq(1-a_{n+1})\cdot\mathrm{E}\Big[\int_{R^N}(D_n(x)-D(x))^2\Big]+K'\cdot a_{n+1}^2\cdot h_{n+1}^{-N}+K'\cdot a_{n+1}\cdot c_{n+1}\\ &\qquad\qquad\qquad\qquad \text{for all } n\geqq n_0\,,\end{aligned} \tag{4.29}$$

where $K'=\max\{2K, 2M, 2M'\}$ and $c_n=\sup_{x\in R^N}|D_n^*(x)-D(x)|$. By Lemma 4.1, (1), we have $\lim_{n\to\infty}c_n=0$. Therefore the theorem is a direct application of Example (1) of Lemma 2.3. Thus the proof of the theorem is completed.

The following corollary is obvious from Tneorem 4.2.

COROLLARY 4.2.1. *Let the conditions in Theorem* 4.2 *be satisfied. Then* $\{D(x)\}_{n=1}^{\infty}$ *is* AO(II).

Next, we shall discuss the order of the mean convergence when $D(\cdot)$ satisfies a uniform Lipschitz condition.

THEOREM 4.3. *Let the conditions of Theorem* 4.1 *be satisfied. If there exist a positive integer* n_0 *and a positive constant* α_0 *such that*

(4.30) $$(1-a_{n+1})\cdot b_n/b_{n+1} \leqq (1-\alpha_0\cdot a_{n+1}) \quad \textit{for all} \quad n \geqq n_0,$$

where $b_n = \max\{a_n\cdot h_n^{-N}, h_n\}$ *for* $n=1, 2, \cdots$. *Then there exists some positive constant* C *satisfying*

(4.31) $$\mathrm{E}\left[\int_{R^N}(D_n(x)-D(x))^2dx\right] \leqq C\cdot b_n \quad \textit{for all} \quad n.$$

PROOF. By the same argument as in Theorem 3.2, we can easily prove the theorem.

By Theorem 4.3 and Lemma 2.1, we have the following corollary.

COROLLARY 4.3.1. *If the conditions in Theorem* 4.3 *hold, if* $f_A(\cdot)$ *and* $f_B(\cdot)$ *have bounded supports, and if* $\lim_{n\to\infty} b_n = 0$. *Then* $\{D_n(x)\}_{n=1}^{\infty}$ *is* AO($\{b_n^{\frac{1}{2}}\}$).

EXAMPLE. In Theorem 4.3, we put $a_n = n^{-\alpha}$, $h_n = n^{-\beta}$ for $0<\alpha\leqq 1$, $\alpha - N\beta > 0$ and $\beta > 0$. Then we have

$$\begin{aligned} \mathrm{E}\left[\int_{R^N}(D_n(x)-D(x))^2dx\right] &\leqq C\cdot n^{-\beta} && \text{if} \quad \alpha/(N+1) \geqq \beta, \\ &\leqq C\cdot n^{-(\alpha-N\beta)} && \text{if} \quad \alpha/(N+1) < \beta. \end{aligned}$$

If $f_A(\cdot)$ and $f_B(\cdot)$ have bounded supports, from Corollary 4.3.1 then it follows that $\{D_n(x)\}_{n=1}^{\infty}$ is AO($n^{-\beta/2}$) if $\alpha/(N+1) \geqq \beta$, and AO($\{n^{-(\alpha-N\beta)/2}$) if $\alpha/(N+1) < \beta$.

§ 5. Acknowledgment.

The author is deeply indepted to Professor N. Furukawa of Kyushu University for helpful advices and critical reading of the original manuscript. The author also expresses his hearty thanks to Professor T. Kitagawa of Kyushu University and Professor K. Tanaka of Niigata University for their helpful advices and encouragement.

References

[1] BRAVERMAN, E. M. and ROZONOER, L. I.: *Convergence of random processes in the theory of learning machines, Part I.* Automation and Remote Control, **30**, No. 1, (1969), 44–64.

[2] BRAVERMAN, E. M. and ROZONOER, L. I.: *Convergence of random processes in the theory of learning machines, Part II.* Automation and Remote Control, **30**, No. 3, (1969), 386–402.

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

[4] HO, Y. and AGRAWALA, A. K.: *"On pattern classification algorithms introduction and survey".* IEEE Trans. Aut. Cont., **AC-31**, No. 6, (1968), 676–690.

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

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

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

[8] RYZIN, J. V.: *A stochastic a posteriori apdating algorithm for pattern recognition.* J. Math. Anal. Appl., **20**, No. 2, (1967), 359–379.

[9] TANAKA, K.: *On the pattern classification problem by learning (I).* Bull. Math. Stat.,

14, No. 1-2, (1970), 31-49.

[10] TSYPKIN, Y. Z.: *Use of the stochastic approximation method in estimating unknown distribution densities from observation.* Automation and Remote Control, 27, No. 3, (1966), 432-434.

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