AN ALMOST SURE CONVERGENCE THEOREM IN A STOCHASTIC APPROXIMATION METHOD WITH DEPENDENT RANDOM VARIABLES

Watanabe, Masafumi
Department of Applied Mathematics, Faculty of Science, Fukuoka University

https://doi.org/10.5109/13134

出版情報：統計数理研究. 18 (3/4), pp.95-112, 1979-03. Research Association of Statistical Sciences
バージョン：
権利関係：

AN ALMOST SURE CONVERGENCE THEOREM IN A STOCHASTIC APPROXIMATION METHOD WITH DEPENDENT RANDOM VARIABLES

By

Masafumi WATANABE*

(Received October 1, 1978)

1. Introduction and Summary

This paper is a continuation of our papers [9], [10] and [11] and is concerned with a Robbins-Monro type stochastic approximation method when a sequence of dependently distributed random vectors is given.

The method of stochastic approximation has been first proposed by H. Robbins and S. Monro ([5]) and its modifications have been thereafter given by many authors. A typical one of them is as follows. Suppose that an R^N-valued random vector $Y_n(x)$ can be observed at $x \in R^N$ and each instant n, and the expected value of $Y_n(x)$, denoted by $\mathrm{E}[Y_n(x)]=M_n(x)$, is unknown to us. Assuming that the equation $M_n(x)=\mathbf{0}$ has a solution $x=\theta_n$ for each $n=1, 2, \cdots$, it is desire to estimate θ_n for sufficiently large n on the basis of observed values $Y_1(X_0), Y_2(X_1), \cdots, Y_{n+1}(X_n), \cdots$ at the points $X_0, X_1, \cdots, X_n, \cdots$ which are produced by the following recurrence relation,

$$(1.1)\qquad \begin{cases} X_0=\text{an arbitrary constant vector in } R^N, \\ X_{n+1}=X_n-a_{n+1}Y_{n+1}(X_n) \qquad \text{for } n=0, 1, 2, \cdots, \end{cases}$$

where $\{a_n\}_{n=1}^{\infty}$ is a sequence of positive numbers which satisfies $\lim_{n\to\infty} a_n=0$ and $\sum_{n=1}^{\infty} a_n =\infty$, and $Y_{n+1}(X_n)$ is an R^N-valued random vector whose conditional distribution given $X_0=x_0, X_1=x_1, \cdots, X_n=x_n$ coincides with the distribution of $Y_{n+1}(x_n)$, consequently

$$(1.2)\qquad \mathrm{E}[Y_{n+1}(X_n)\,|\,X_0, X_1, \cdots, X_n]=M_{n+1}(X_n) \qquad a.s..$$

In our previous papeper [8], we showed that $\lim_{n\to\infty} \mathrm{E}\|X_n-\theta_n\|^2=0$ and $\lim_{n\to\infty} \|X_n-\theta_n\|=0$ *a. s.* under the condition (1.2) and some additional conditions on $\{M_n(\cdot)\}_{n=1}^{\infty}$, $\{Y_{n+1}(X_n)\}_{n=0}^{\infty}$ and $\{\theta_n\}_{n=1}^{\infty}$.

In applying the above stochastic approximation method, we often encounter the case when $Y_n(x)$ can be expressed in the form of $\Phi_n(x, Y_n)$ where $\{Y_n\}_{n=1}^{\infty}$ is a given sequence of observable random vectors which does not depend on x and $\Phi_n(\cdot, \cdot)$'s are

* Department of Applied Mathematics, Faculty of Science, Fukuoka University, Fukuoka, Japan

measurable transformations. For example, in the learning problem of estimating an unknhwn optimal discriminant function for a pattern classification, $\{Y_n\}_{n=1}^{\infty}$ means a training sequence ([4], [6], [7]), and in the problem of estimating an unknown function on the basis of a sequence of input-output data, $\{Y_n\}_{n=1}^{\infty}$ means a sequence of a pair of inputs and outputs variables ([1], [4]). From these viewpoints, in [9], [10], [11] and this paper, we have considered the case when the random vector $Y_n(x)$ can be expressed in the form of $\Phi_n(x, Y_n)$ where $\{Y_n\}_{n=1}^{\infty}$ is a sequence of R^M-valued random vectors and $\Phi_n(\cdot, \cdot)$'s are Borel measurable mappings from $R^N \times R^M$ into R^N. And we have discussed the problem of finding the solution of the equation

$$\mathrm{E}[\Phi_n(x, Y_n)] = \mathbf{0}$$

for sufficiently large n.

If Y_n's are independently distributed random vectors then the relation (1.2) is automatically satisfied under the procedure (1.1). Hence this is the case of the usual R-M stochastic approximation method. If Y_n's are not independently distributed random vectors then (1.2) may not be satisfied. Therefore we can not apply the usual R-M stochastic approximation method to this case directly. For this reason, in our previous papers ([9], [10], [11]) we discussed the case when the relation (1.2) may not be satisfied. And it was shown that the mean square convergence theorems ([9], [10]) and the $a.s.$ convergence theorem ([11]) hold. In [9] and [11] it was assumed that $\Phi_n(x, y)$ can be expressed in the form of $(x-\theta_n)A_n + \Gamma_n(y)$ where A_n's are $N \times N$-matrices (not random) and Γ_n's are Borel measurable mappings from R^M into R^N. In [10] it was assumed that A_n's are bounded linear operators of a separable Hilbert space H_0 into itself and Γ_n's are measurable mappings from a Hilbert space into H_0.

In this paper we shall try to generalized to the case when the $N \times N$-matrices A_n's depend on $y \in R^M$. And we shall give the $a.s.$ convergence theorem.

This paper consists of four sections. In Section 2, we shall give the formulation of our problem and give some lemmas to be used throughout this paper. In Section 3, we shall give the $a.s.$ convergence theorem which is the main result. In Section 4, we shall give two applications of our result.

2. Formulation and Preliminalies

Throughout this paper R^n denotes the n-dimensional Euclidian space, $\mathbf{0}$ denotes the zero vector and unless otherwise indicated vectors are to be row vectors. And let $(\Omega, \mathcal{A}, P)$ be a probabily space.

Let $\Phi_n^{(i)}(\cdot, \cdot)$ ($i=1, 2, \cdots, N$ and $n=1, 2, \cdots$) be a real valued Borel function defined on $R^N \times R^M$. And we put

$$\Phi_n(x, y) = (\Phi_n^{(1)}(x, y), \cdots, \Phi_n^{(N)}(x, y))$$

for $x \in R^N$, $y \in R^M$ and $n=1, 2, \cdots$. Throughout this paper we suppose that $\{Y_n\}_{n=1}^{\infty}$ is a sequence of R^M-valued random vectors defined on $(\Omega, \mathcal{A}, P)$, i.e. Y_n's are measurable mappings from $(\Omega, \mathcal{A})$ into $(R^M, \mathcal{B}_M)$ where $\mathcal{B}_M$ is the Borel field in R^M.

Now we shall consider the following problem.

PROBLEM. For sufficiently large n, find the solution of the equation;

$$M_n(x)=\mathbf{0} \tag{2.1}$$

where $M_n(x)=\mathrm{E}[\Phi_n(x, y_n)]=(\mathrm{E}[\Phi_n^{(1)}(x, Y_n)], \cdots, \mathrm{E}[\Phi_n^{(N)}(x, Y_n)])$, $n=1, 2, \cdots$, $x\in R^N$.

For the above problem we suppose that $M_n(\cdot)$'s are unknown to us. But we assume that we can make use of the random vector $\Phi_n(x, Y_n)$ at each $x\in R^N$ and $n=1, 2, \cdots$.

Let us give a Robbins-Monro type procedure for estimating the solution of the equation (2.1) for sufficiently large n.

PROCEDURE. Let $X_0=\theta_0$ be an arbitrary constant vector in R^N and define $X_1, X_2, \cdots$ by the following recurrence relation;

$$X_{n+1}=X_n-a_{n+1}\Phi_{n+1}(X_n, Y_{n+1}) \tag{2.2}$$

where $\{a_n\}_{n=1}^{\infty}$ is a sequence of positive numbers which convergees to 0 as $n\to\infty$.

Throughout this paper let $\mathcal{A}_n=\sigma(Y_n)$ $(n=1, 2, \cdots)$ be the σ-field generated by Y_n. And let us define the dependent coefficient of the σ-fields $\mathcal{A}'$ and $\mathcal{A}''$ which are the sub-σ-fields of $\mathcal{A}$ by the relation,

$$\phi(\mathcal{A}', \mathcal{A}'')=\sup_{A\in\mathcal{A}''}(\operatorname*{ess\,sup}_{\omega\in\Omega}|P(A|\mathcal{A}')(\omega)-P(A)|). \tag{2.3}$$

And we put

$$\rho_n=\sup_{1\leq m}\phi(\mathcal{A}_m, \mathcal{A}_{m+n}) \tag{2.4}$$

for $n=1, 2, \cdots$. For the above dependent coefficient we refer to M. Iosifescu and R. Theodorescu ([2]).

With all vectors considered as elements of R^N we adopt the following notations. If $x, y\in R^N$ then $\langle x, y\rangle$ will denote their inner product. The norm of $x\in R^N$ is denoted by $\|x\|$ and, of course, is equal to $\langle x, x\rangle^{1/2}$. If A is a $N\times N$-matrix then we define in the usual way

$$\|A\|_N=\sup_{\|x\|=1}\|xA\|. \tag{2.5}$$

Moreover I denotes the identity matrix.

The following five lemmas will be needed for the proof of the theorem in Section 3.

LEMMA 2.1. *Define* $\|\cdot\|_0$ *by* $\|A\|_0=\max_{1\leq i,j\leq N}|a_{ij}|$ *where* A *is a* $N\times N$*-matrix whose* (i, j)*-th element is* a_{ij} *(real number). Then there exist two positive number* C_1 *and* C_2*, such that*

$$C_1\|A\|_0\leqq\|A\|_N\leqq C_2\|A\|_0 \tag{2.6}$$

for all $N\times N$*-matrix* A*, where* $\|\cdot\|_N$ *is defined by* (2.5).

PROOF. Put $x=(x_1, x_2, \cdots, x_N)$ and $e_i=(0, \cdots, 0, \overset{\overset{i}{\vee}}{1}, \cdots, 0)$ for $i=1, 2, \cdots, N$. Then, using the Schwarz inequality, it follows that

$$(2.7)\qquad \begin{aligned}\|A\|_N &= \sup_{\|x\|=1}\|xA\| \\ &= \sup_{\|x\|=1}\left\{\sum_{i=1}^{N}\left(\sum_{j=1}^{N} a_{ij}x_j\right)^2\right\}^{1/2} \\ &\leqq N^{1/2}\sup_{\|x\|=1}\left\{\max_{1\leqq i\leqq N}\left(\sum_{j=1}^{N} a_{ij}x_j\right)^2\right\}^{1/2} \\ &\leqq N^{1/2}\sup_{\|x\|=1}\left\{\max_{1\leqq i\leqq N}\left(\sum_{j=1}^{N} a_{ij}^2\right)\left(\sum_{j=1}^{N} x_j^2\right)\right\}^{1/2} \\ &= N^{1/2}\max_{1\leqq i\leqq N}\left(\sum_{j=1}^{N} a_{ij}^2\right)^{1/2} \\ &\leqq N\max_{1\leqq i,j\leqq N}|a_{ij}|\,.\end{aligned}$$

And we have

$$(2.8)\qquad \begin{aligned}\|A\|_N &\geqq \|e_iA\| \\ &= \left(\sum_{j=1}^{N} a_{ij}^2\right)^{1/2} \\ &\geqq |a_{ij}| \qquad \text{for all}\quad i,\ j=1,2,\cdots,N\,.\end{aligned}$$

Hence, (2.6) holds.

The following lemma was proved in [6].

LEMMA 2.2. *Let $\{x_n\}_{n=0}^{\infty}$ 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*

(i) $x_{n+1}\leqq(1-a_{n+1})x_n+a_{n+1}b_{n+1}+K_{n+1}$ *for* $n=0, 1, \cdots,$

(ii) $\lim_{n\to\infty} a_n=0,\qquad \sum_{n=1}^{\infty} a_n=\infty\,,$

(iii) $\lim_{n\to\infty} b_n=0,\qquad \sum_{n=1}^{\infty} K_n<\infty\,.$

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

REMARK. Assume $0\leqq a_n<1$ for all $n\geqq 1$. And define

$$(2.9)\qquad x_n=\sum_{k=1}^{n} a_kb_k\prod_{l=k+1}^{n}(1-a_l)+\sum_{k=1}^{n} K_k\prod_{l=k+1}^{n}(1-a_l)$$

for $n=1,2,\cdots$. Then $\{x_n\}_{n=1}^{\infty}$ satisfies (i).

LEMMA 2.3. (*Kronecker's lemma*). *If $\sum x_n$ converges to a finite number and $\{b_n\}_{n=1}^{\infty}$ is a non-increasing sequence of positive numbers which converges to zero as $n\to\infty$, then it follows that* $\lim_{n\to\infty} b_n\sum_{k=1}^{n}\frac{x_k}{b_k}=0$.

PROOF. It is ommited (see [3]).

LEMMA 2.4. *Let $\mathcal{X}$ be a normed linear space. Suppose that there exist four sequences of elements in $\mathcal{X}$, $\{x_n\}_{n=0}^{\infty}$, $\{u_n\}_{n=1}^{\infty}$, $\{v_n\}_{n=1}^{\infty}$ and $\{w_n\}_{n=1}^{\infty}$, such that*

$$(2.10)\qquad x_{n+1}=T_{n+1}x_n+u_{n+1}+v_{n+1}+w_{n+1}\,,\qquad n=0, 1, 2, \cdots,$$

where $\{T_n\}_{n=1}^{\infty}$ *is a sequence of bounded linear operators of* $\mathfrak{X}$ *into itself. And suppose that there exist three sequences of positive numbers* $\{\alpha_n\}_{n=1}^{\infty}$, $\{\beta_n\}_{n=1}^{\infty}$ *and* $\{\gamma_n\}_{n=1}^{\infty}$, *positive numbers* K_i $(i=1,2,3,4,5)$ *and a positive integer* N_0, *such that*

(i) $\lim_{n\to\infty} \alpha_n=0, \qquad \sum_{n=1}^{\infty} \alpha_n=\infty,$

(ii) $\sup_{1\leq n} \beta_n|\alpha_{n+1}^{-1}-\alpha_n^{-1}|\leqq K_1,$

(iii) $\sup_{1\leq n} \beta_n\leqq K_2, \qquad \sup_{1\leq n} \beta_n\beta_{n+1}^{-1}\leqq K_3,$

(iv) $\|T_n\cdot T_{n-1}\cdots T_{m+1}\|\leqq K_4 \prod_{k=m+1}^{n} |1-\alpha_k|$ *for* $n>m\geqq N_0$,

(v) $\|I-T_n\|\leqq\alpha_n\gamma_n$ *for* $n>N_0$, *where* I *is the identity operator,*

(vi) $\sup_{1\leq n} \beta_n\gamma_n\leqq K_5$ *or* $\sum_{n=1}^{\infty} \alpha_n\beta_n\gamma_n<\infty,$

(vii) $\lim_{n\to\infty} \alpha_n^{-1}\|u_n\|=0,$

(viii) $\sum_{n=1}^{\infty} \|v_n\|<\infty,$

(ix) $\lim_{n\to\infty} \alpha_n\beta_n^{-1}\left\|\sum_{k=1}^{n} \alpha_k^{-1}w_k\right\|=0.$

Then it holds that $\lim_{n\to\infty}\|x_n\|=0$.

Proof. Let us put $T_m^{(n)}=T_nT_{n-1}\cdots T_{m+1}$ for $n>m\geqq 0$ and $T_n^{(n)}=I$. Then (2.10) implies

$$x_n=\sum_{k=1}^{n} T_k^{(n)}(u_k+v_k+w_k)+T_0^{(n)}x_0. \tag{2.11}$$

From (i) and (iv) we have

$$\lim_{n\to\infty}\|T_0^{(n)}x_0\|=0. \tag{2.12}$$

By (i), (iv), (vii), (viii) and using Lemma 2.2, we have

$$\lim_{n\to\infty}\sum_{k=1}^{n}\|T_k^{(n)}u_k\|=0 \tag{2.13}$$

and

$$\lim_{n\to\infty}\sum_{k=1}^{n}\|T_k^{(n)}v_k\|=0. \tag{2.14}$$

Next we shall prove that

$$\lim_{n\to\infty}\left\|\sum_{k=1}^{n} T_k^{(n)}w_k\right\|=0. \tag{2.15}$$

Note that $\sum_{k=1}^{n} T_k^{(n)}w_k=\sum_{k=N_0}^{n} T_k^{(n)}w_k+\sum_{k=1}^{N_0-1} T_k^{(n)}w_k$. Firstly, (i) and (iv) imply

$$\lim_{n\to\infty}\sum_{k=1}^{N_0-1}\|T_k^{(n)}w_k\|=0. \tag{2.16}$$

Hence, in order to show (2.15), we have only to prove that

$$\lim_{n\to\infty}\left\|\sum_{k=N_0}^{n}T_k^{(n)}w_k\right\|=0. \tag{2.17}$$

Let us put

$$\begin{cases} s_{N_0-1}=\mathbf{0}, \text{ where } \mathbf{0} \text{ is the zero element in } \mathscr{X}, \\ s_n=\alpha_n\sum_{k=N_0}^{n}\alpha_k^{-1}w_k \quad \text{for} \quad n\geqq N_0 \end{cases} \tag{2.18}$$

and

$$y_n=\sum_{k=N_0}^{n}T_k^{(n)}w_k. \tag{2.19}$$

Then, noting $\alpha_n^{-1}s_n-\alpha_{n-1}^{-1}s_{n-1}=\alpha_n^{-1}w_n$, we obtain

$$\begin{aligned} y_n&=s_n+\sum_{k=N_0}^{n-1}(\alpha_kT_k^{(n)}-\alpha_{k+1}T_{k+1}^{(n)})\alpha_k^{-1}s_k \\ &=s_n+\sum_{k=N_0}^{n-1}T_{k+1}^{(n)}(T_{k+1}-I)s_k+\sum_{k=N_0}^{n-1}\alpha_{k+1}(\alpha_{k+1}^{-1}-\alpha_k^{-1})T_{k+1}^{(n)}s_k. \end{aligned} \tag{2.20}$$

From (i), we can assume that $0<\alpha_n<1$ for all $n\geqq N_0$ without loss of generality. Hence (ii), (iv), (v) and (2.20) imply

$$\begin{aligned} \|y_n\|\leqq\|s_n\|&+K_4\sum_{k=N_0}^{n-1}\alpha_{k+1}\beta_{k+1}\gamma_{k+1}\beta_{k+1}^{-1}\|s_k\|\prod_{l=k+2}^{n}(1-\alpha_l) \\ &+K_1\sum_{k=N_0}^{n-1}\alpha_{k+1}\beta_k^{-1}\|s_k\|\prod_{l=k+2}^{n}(1-\alpha_l). \end{aligned} \tag{2.21}$$

Note that $\beta_{k+1}^{-1}\|s_k\|=(\beta_k\beta_{k+1}^{-1})(\beta_k^{-1}\|s_k\|)$ and $\lim_{n\to\infty}\beta_n^{-1}\|s_n\|=0$. Hence, by (i), (iii), (iv), (vi), (2.21) and using Lemma 2.2 we obtain (2.17).

Thus the proof Lemma 2.4 is completed.

REMARKS. (1) If we put $\alpha_n=n^{-1}$, $\beta_n=n^{-\delta}$ $(0<\delta<1)$ for $n=1,2,\cdots$, then (i), (ii) and (iii) are satisfied.

(2) Let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence of positive numbers which satisfies (i). And put $\beta_n=\alpha_n^{\delta}$ $(0<\delta<1)$ and suppose that (ii) holds. Then it holds that $|\alpha_n\alpha_{n+1}^{-1}-1|\leqq K_1\alpha_n^{1-\delta}$. Hence (iii) holds.

Next we state without proof the lemma which was proved in [2].

LEMMA 2.5. (*A strong law of large numbers for dependent random variables*). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real valued random variables defined on $(\Omega, \mathcal{A}, P)$. Suppose that there exist a positive interger n_0 and a increasing sequence of positive numbers $\{k_n\}_{n=1}^{\infty}$, such that*

(i) $\overline{\lim_{n\to\infty}}\,\phi\left(\bigvee_{i=1}^{n}\mathcal{A}_i, \bigvee_{i=n+n_0}^{\infty}\mathcal{A}_i\right)<1,$

(ii) $\sum_{n=1}^{\infty} \rho_n^{1/2} < \infty$,

(iii) $\lim_{n\to\infty} k_n = \infty$,

(iv) $\sum_{n=1}^{\infty} k_n^{-2} \mathrm{E}(X_n - \mathrm{E}X_n)^2 < \infty$,

where $\mathcal{A}_i$ *is the* σ*-field generated by* X_i *(i.e.* $\mathcal{A}_i = \sigma(X_i)$*),* $\bigvee_{i\in\Lambda} \mathcal{A}_i$ *denotes the* σ*-field generated by the join of the* σ*-field* $\mathcal{A}_i$*,* $i \in \Lambda$*, and* $\rho_n = \sup_{1\leq m} \phi(\mathcal{A}_m, \mathcal{A}_{m+n})$*. Then it holds that*

$$\lim_{n\to\infty} k_n^{-1} \sum_{i=1}^{n} (X_i - \mathrm{E}X_i) = 0 \qquad a.s.$$

EMARK. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random $N\times N$-matrices (vectors) which satisfies the conditions (i) to (iii) and the following condition,

(iv)′ $\sum_{n=1}^{\infty} k_n^{-2} \mathrm{E}\|X_n - \mathrm{E}X_n\|_N^2 < \infty \qquad \left(\sum_{n=1}^{\infty} k_n^{-2} \mathrm{E}\|X_n - \mathrm{E}X_n\|^2 < \infty\right)$.

Then, noting Lemma 2.1 and using Lemma 2.5, we have

$$\lim_{n\to\infty} k_n^{-1} \left\| \sum_{i=1}^{n} (X_i - \mathrm{E}X_i) \right\|_N = 0 \qquad a.s.$$

$$\left(\lim_{n\to\infty} k_n^{-2} \left\| \sum_{i=1}^{n} (X_i - \mathrm{E}X_i) \right\| = 0 \qquad a.s.\right).$$

3. Main result

Let the sequence of positive numbers $\{a_n\}_{n=1}^{\infty}$, the sequence of R^M-valued random vectors $\{Y_n\}_{n=1}^{\infty}$ and the sequence of the σ-fields $\{\mathcal{A}_n\}_{n=1}^{\infty}$ be as defined in Section 2. Moreover, let the sequence of non-negative numbers $\{\rho_n\}_{n=1}^{\infty}$ be as defined by (2.4) in Section 2. And suppose that δ be a number which satisfies $0<\delta\leq 1/7$. The following assumptions about $\{a_n\}_{n=1}^{\infty}$, $\{\Phi_n(\cdot, \cdot)\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ will be used in the proof of the theorem.

ASSUMPTIONS.

(A1) $\{a_n\}_{n=1}^{\infty}$ is a decreasing sequence of positive numbers, such that $\sum_{n=1}^{\infty} a_n = \infty$.

(A2) $\sum_{n=1}^{\infty} a_n^{1+\delta} < \infty$.

(A3) $\sup_{1\leq n} a_n^{\delta}(a_{n+1}^{-1} - a_n^{-1}) < \infty$.

(B1) For each $n=1, 2, \cdots$, $x \in R^N$ and $y \in R^M$, $\Phi_n(x, y)$ is expressed in the form of

$$\Phi_n(x, y) = (x - \theta_n) A_n(y) + \Gamma_n(y),$$

where θ_n is a constant vector in R^N, $A_n(y) = (A_n^{(i,j)}(y))$ is a $N\times N$-matrix whose (i, j)-th element $A_n^{(i,j)}(y)$ is a Borel function defined on R^M and $\Gamma_n(y) = (\Gamma_n^{(1)}(y), \cdots, \Gamma_n^{(N)}(y))$

is a vector in R^N whose i-th element $\Gamma_n^{(i)}(y)$ is a Borel function defined on R^M.

(B2) $\sum_{n=1}^{\infty} a_n^{1+\delta}\mathrm{E}\|A_n(Y_n)\|_N^2<\infty$.

(B3) $\inf_{\|x\|=1}\langle xA_n(Y_n),\ x\rangle\geqq 0$ $a.s.$ for $n=1, 2, \cdots$.

(B4) There exists a positive number α_0 such that

$$\inf_{\|x\|=1}\langle x\mathrm{E}[A_n(Y_n)],\ x\rangle\geqq\alpha_0 \quad \text{for} \quad n=1, 2, \cdots.$$

(B5) $\lim_{n\to\infty}\|\mathrm{E}[\Gamma_n(Y_n)]\|=0$.

(B6) $\sum_{n=1}^{\infty} a_n^{1+\delta}\mathrm{E}\|\Gamma_n(Y_n)-\mathrm{E}[\Gamma_n(Y_n)]\|^2<\infty$.

(B7) $\lim_{n\to\infty} a_{n+1}^{-1}\|\theta_n-\theta_{n+1}\|=0$.

(C1) There exists a positive integer n_0 such that

$$\overline{\lim_{n\to\infty}}\,\phi\Big(\bigvee_{i=1}^{n}\mathcal{A}_i,\ \bigvee_{i=n+n_0}^{\infty}\mathcal{A}_i\Big)<1,$$

where $\bigvee_{i\in\Lambda}\mathcal{A}_i$ denotes the σ-field generated by the join of the σ-fields $\mathcal{A}_i$, $i\in\Lambda$.

(C2) $\sum_{n=1}^{\infty}\rho_n^{1/2}<\infty$.

REMARKS. (1) (B1) and (B5) imply $\lim_{n\to\infty}\|M_n(\theta_n)\|=0$. And suppose that (B1) and the followihg condition,

(B5)′ $\mathrm{E}[\Gamma_n(Y_n)]=\mathbf{0}$ $(n=1, 2, \cdots)$

hold then θ_n $(n=1, 2, \cdots)$ is a solution of the equation (2.1).

(2) Assume that Y_n's are independently distributed random vectors. Then it is easily seen that $\overline{\lim_{n\to\infty}}\,\phi\Big(\bigvee_{i=1}^{n}\mathcal{A}_i,\ \bigvee_{i=n+n_0}^{\infty}\mathcal{A}_i\Big)=0$ for all n_0 and $\rho_n=0$ for all n. Hence, (C1) and (C2) hold.

(3) Put $a_n=n^{-t}$ $((1+\delta)^{-1}<t\leqq 1)$, $n=1, 2, \cdots$. Then (A1) to (A3) are satisfied.

(4) (B2) implies $\mathrm{E}\Big[\sum_{n=1}^{\infty}a_n^{1+\delta}\|A_n(Y_n)\|_N^2\Big]<\infty$, implying that $\sum_{n=1}^{\infty}a_n^{1+\delta}\|A_n(Y_n)\|_N^2<\infty$ $a.s.$. And (A2), (B5) and (B6) imply $\sum_{n=1}^{\infty}a_n^{1+\delta}\|\Gamma_n(Y_n)\|^2<\infty$ $a.s.$.

The following theorem is our main result in this paper.

THEOREM. *Let* $X_0, X_1, \cdots, X_n, \cdots$ *be as defined by* (2.2) *in Section* 2. *Assume* (A1) *to* (A3), (B1) *to* (B7) *and* (C1) *to* (C2). *Then it holds that*

$$\lim_{n\to\infty}\|X_n-\theta_n\|=0 \quad a.s.. \tag{3.1}$$

PROOF. According to the procedure (2.2) and (B1), it follows that

$$\begin{aligned}(3.2)\qquad & X_{n+1}-\theta_{n+1}\\ &=(X_n-\theta_n)(I-a_{n+1}A_{n+1}(Y_{n+1}))+a_{n+1}\{a_{n+1}^{-1}(\theta_n-\theta_{n+1})(I-a_{n+1}A_{n+1}(Y_{n+1}))\\ &\quad-\Gamma_{n+1}(Y_{n+1})\},\qquad n=0, 1, 2, \cdots.\end{aligned}$$

Let us put

$$X_n-\theta_n=x_n,\ \mathrm{E}[A_n(Y_n)]=\hat{A}_n,\ \mathrm{E}[\Gamma_n(Y_n)]=\hat{\Gamma}_n \quad \text{and} \quad I-a_n\mathrm{E}[A_n(Y_n)]=T_n .$$

Then we have

(3.3) $$x_{n+1}=x_nT_{n+1}+a_{n+1}U_{n+1}+a_{n+1}V_{n+1}-a_{n+1}W_{n+1}$$

where

(3.4) $$U_{n+1}=a_{n+1}^{-1}(\theta_n-\theta_{n+1})-\hat{\Gamma}_{n+1},$$

(3.5) $$V_{n+1}=-(\theta_n-\theta_{n+1})\hat{A}_{n+1}$$

and

(3.6) $$W_{n+1}=x_n(\hat{A}_{n+1}-A_{n+1}(Y_{n+1}))+(\theta_n-\theta_{n+1})(\hat{A}_{n+1}-A_{n+1}(Y_{n+1})) -\Gamma_{n+1}(Y_{n+1})+\hat{\Gamma}_{n+1}.$$

By virture of (3.2) to (3.6), in order to show (3.1), it is sufficiently to prove that (3.3) to (3.6) satisfy the conditions in Lemma 2.4 for putting $\alpha_n=\alpha_0a_n$ and $\beta_n=a_n^{\delta}$, $n=1,2,\cdots$. Hence we claim that the following (I) to (V).

(I); There exists a sequence of positive numbers $\{\gamma_n\}_{n=1}^{\infty}$ such that

$$\|I-T_n\|\leqq\alpha_0a_n\gamma_n \quad \text{for} \quad n=1,2,\cdots$$

and

$$\sum_{n=1}^{\infty}a_n^{1+\delta}\gamma_n<\infty .$$

(II); There exist a positive integer N_0 and a positive number K, such that

$$\|T_{m+1}\cdot T_{m+2}\cdots T_n\|_N\leqq K\prod_{k=m+1}^{n}(1-\alpha_0a_k) \quad \text{for} \quad n>m\geqq N_0,$$

where $1-\alpha_0a_k>0$ $(k\geqq N_0)$.

(III); $\lim_{n\to\infty}\|U_n\|=0$.

(IV); $\sum_{n=1}^{\infty}a_n\|V_n\|<\infty$.

(V); $\lim_{n\to\infty}a_n^{1-\delta}\left\|\sum_{k=1}^{n}W_k\right\|=0$ *a.s.*.

Proof of (I). Put $\alpha_0^{-1}\|\hat{A}_n\|_N=\gamma_n$, $n=1,2,\cdots$. Then we have

(3.7) $$\|I-T_n\|_N=a_n\|\hat{A}_n\|_N=\alpha_0a_n\gamma_n, \qquad n=1,2,\cdots.$$

The fact that $\|\hat{A}_n\|_N^2\leqq\mathrm{E}\|A_n(Y_n)\|_N^2$ and the Schwarz inequality imply

(3.8) $$\sum_{n=1}^{\infty}a_n^{1+\delta}\gamma_n=\alpha_0^{-1}\sum_{n=1}^{\infty}a_n^{1+\delta}\|\hat{A}_n\|_N \leqq\alpha_0^{-1}\Big(\sum_{n=1}^{\infty}a_n^{1+\delta}\Big)^{1/2}\Big(\sum_{n=1}^{\infty}a_n^{1+\delta}\mathrm{E}\|A_n(Y_n)\|_N^2\Big)^{1/2}.$$

Hence, according to (A2) and (B2), (I) follows.

PROOF OF (II). According to (A1) it follows that there exist a positive integer N_0 and a positive number k_0, such that

$$1-2\alpha_0 a_n \geqq k_0^{-1} \quad \text{for} \quad n \geqq N_0 . \tag{3.9}$$

(B4) and (3.9) imply

$$\begin{aligned} \|I-a_n\hat{A}_n\|_N^2 &\leqq 1+a_n^2\|\hat{A}\|_N^2-2\alpha_0 a_n \\ &=(1-2\alpha_0 a_n)\{1+(1-2\alpha_0 a_n)^{-1}(a_n^2\|\hat{A}_n\|_N^2)\} \\ &\leqq (1-2\alpha_0 a_n)(1+k_0 a_n^2\|\hat{A}_n\|_N^2) \end{aligned} \tag{3.10}$$

for $n \geqq N_0$. Hence we have

$$\|T_{m+1}\cdots T_n\|_N \leqq \left\{\prod_{k=m+1}^{n}(1+k_0 a_k^2\|\hat{A}_k\|_N^2)\right\}^{1/2} \prod_{k=m+1}^{n}(1-\alpha_0 a_k) \tag{3.11}$$

for $n>m\geqq N_0$. And (B2) implies

$$\sup_{1\leqq m<n} \prod_{k=m+1}^{n}(1+k_0 a_k^2\|\hat{A}_k\|_N^2)<\infty . \tag{3.12}$$

Hence (II) holds.

PROOF OF (III) From (B5) and (B7) it is easily seen that (III) holds.

PROOF OF (IV). Note that $a_n\|V_n\| \leqq a_n^{-1}\|\theta_{n-1}-\theta_n\|\cdot a_n^2\|\hat{A}_n\|_N$ and $1+\delta<2$. Then, (B7) and (3.8) imply (IV).

Next we shall prove (V) which is the main part of the proof of Theorem.

PROOF OF (V). In (3.6), let us put

$$\hat{\Gamma}_{n+1}-\Gamma_{n+1}(Y_{n+1})=W_{1,n+1}, \tag{3.13}$$

$$(\theta_n-\theta_{n+1})(\hat{A}_{n+1}-A_{n+1}(Y_{n+1}))=W_{2,n+1} \tag{3.14}$$

and

$$x_n(\hat{A}_{n+1}-A_{n+1}(Y_{n+1}))=W_{3,n+1} . \tag{3.15}$$

In order to show (V), we have to prove that

$$\lim_{n\to\infty} a_n^{1-\delta}\left\|\sum_{k=1}^{n} W_{i,k}\right\|=0 \qquad a.s. \tag{3.16}$$

for $i=1,2,3$. The fact that $0<\delta\leqq 1/7$, it follows that

$$\frac{1+\delta}{2}<1-\delta . \tag{3.17}$$

Hence, by (A1), (B6), (C1), (C2) and the remark of Lemma 2.5, we have

$$\lim_{n\to\infty} a_n^{1-\delta}\left\|\sum_{k=1}^{n} W_{1,k}\right\|=0 \qquad a.s.. \tag{3.18}$$

Next we have

$$(3.19)\qquad a_n^{1-\delta}\left\|\sum_{k=1}^{n}W_{2,k}\right\|$$
$$\leqq a_n^{1-\delta}\sum_{k=1}^{n}a_k^{-1}\|\theta_{k-1}-\theta_k\|\cdot a_k\|\hat{A}_k\|_N+a_n^{1-\delta}\sum_{k=1}^{n}a_k^{-1}\|\theta_{k-1}-\theta_k\|\cdot a_k\|A_k(Y_k)\|_N.$$

According to the Schwarz inequality, (A2) and the remark (4) of Assumptions, it follows that

$$(3.20)\qquad \sum_{n=1}^{\infty}a_n^{1+\delta}\|A_n(Y_n)\|_N\leqq\left(\sum_{n=1}^{\infty}a_n^{1+\delta}\right)^{1/2}\left(\sum_{n=1}^{\infty}a_n^{1+\delta}\|A_n(Y_n)\|_N^2\right)^{1/2}<\infty\qquad a.s..$$

Hence, by (3.8), (3.20), (B7) and applying Lemma 2.3, we obtain

$$(3.21)\qquad \lim a_n^{1-\delta}\left\|\sum_{k=1}^{n}W_{2,k}\right\|=0\qquad a.s..$$

By virtue of (3.18) and (3.21), in order to show (V) we have only to prove that

$$(3.22)\qquad \lim a_n^{1-\delta}\left\|\sum_{k=1}^{n}W_{3,k}\right\|=0\qquad a.s..$$

From (3.2), we have

$$(3.23)\qquad \|x_{n+1}\|\leqq\|x_n\|\cdot\|I-a_{n+1}A_{n+1}(Y_{n+1})\|_N+a_{n+1}\|R_{n+1}\|,\qquad n=0,1,2,\cdots,$$

where

$$(3.24)\qquad R_{n+1}=a_{n+1}^{-1}(\theta_n-\theta_{n+1})(I-a_{n+1}A_{n+1}(Y_{n+1}))-\Gamma_{n+1}(Y_{n+1}).$$

And (A3) implies

$$(3.25)\qquad \|I-a_nA_n(Y_n)\|_N^2\leqq 1+a_n^2\|A_n(Y_n)\|_N^2\qquad a.s.,\qquad n=1,2,\cdots.$$

Hence we have

$$(3.26)\qquad \|x_{n+1}\|\leqq\|x_n\|(1+a_{n+1}^2\|A_{n+1}(Y_{n+1})\|_N^2)^{1/2}+a_{n+1}\|R_{n+1}\|\qquad a.s..$$

The backward iteration of (3.26) to 0 and noting $x_0=\mathbf{0}$ imply

$$(3.27)\qquad \|x_n\|\leqq\sum_{k=1}^{n}a_k\|R_k\|\prod_{l=k+1}^{n}(1+a_l^2\|A_l(Y_l)\|_N^2)^{1/2}\qquad a.s.,\qquad n=1,2,\cdots.$$

Hence, by the remark (4) of Assumptions, it holds that there exists a positive random variable Z_0 which is $a.s.$ finite, such that

$$(3.28)\qquad \|x_n\|\leqq Z_0\cdot\sum_{k=1}^{n}a_k\|R_k\|\qquad a.s.,\qquad n=1,2,\cdots.$$

And from (3.2), we have

$$(3.29)\qquad \|x_n-x_{n-1}\|\leqq a_n\|x_{n-1}\|\|A_n(Y_n)\|_N+a_n\|R_n\|,\qquad n=1,2,\cdots.$$

By virtue of $0<\delta\leqq 1/7$, it followe that there exists a number $\hat{\delta}$, such that

$$(3.30)\qquad \frac{1-\delta}{2}\geqq\hat{\delta}\geqq 3\delta.$$

For the above $\hat{\delta}$, let us put $a_0=1$ and

$$(3.31)\qquad \begin{cases} S_0=\mathbf{0} \\ S_n=a_n^{1-\hat{\delta}} \sum_{k=1}^{n} (\hat{A}_k-A_k(Y_k)). \end{cases}$$

And (3.30) and (B2) imply

$$(3.32)\qquad \sum_{n=1}^{\infty} a_n^{2-2\hat{\delta}} \mathrm{E}\|\hat{A}_n-A_n(Y_n)\|_N^2$$

$$\leqq \sum_{n=1}^{\infty} a_n^{1+\delta} \mathrm{E}\|A_n(Y_n)\|_N^2<\infty .$$

Hence, by applying Lemma 2.5 and its remark, it follows that

$$(3.33)\qquad \lim_{n\to\infty}\|S_n\|_N=0 \qquad a.s..$$

Now we put

$$(3.34)\qquad y_n=a_n^{1-\delta} \sum_{k=1}^{n} W_{3,k} .$$

Then we have

$$(3.35)\qquad y_n=a_n^{1-\delta} \sum_{k=1}^{n} x_{k-1}(a_k^{\hat{\delta}-1}S_k-a_{k-1}^{\hat{\delta}-1}S_{k-1})$$

$$=a_n^{\hat{\delta}-\delta}x_{n-1}S_n+a_n^{1-\delta} \sum_{k=1}^{n-1} a_k^{\hat{\delta}-1}(x_{k-1}-x_k)S_k .$$

Hence we have

$$(3.36)\qquad \|y_n\|\leqq a_n^{\hat{\delta}-\delta}\|x_{n-1}\|\|S_n\|_N+a_n^{1-\delta} \sum_{k=1}^{n} a_k^{\hat{\delta}-1}\|x_{k-1}-x_k\|\|S_k\|_N .$$

At first we shall prove

$$(3.37)\qquad \lim_{n\to\infty} a_n^{\hat{\delta}-\delta}\|x_{n-1}\|\|S_n\|_N=0 \qquad a.s..$$

By (3.28), we have

$$(3.38)\qquad a_n^{\hat{\delta}-\delta}\|x_{n-1}\|\|S_n\|_N$$

$$\leqq Z_0 a_n^{\hat{\delta}-\delta}\|S_n\|_N \sum_{k=1}^{n} a_k\|R_k\| \qquad a.s..$$

And according to (A1), (A2), (B7), (3.20) and the remark (4), it follows that

$$(3.39)\qquad \sum_{n=1}^{\infty} a_n^{1+\delta}\|R_n\|$$

$$\leqq \sum_{n=1}^{\infty} a_n^{1+\delta}(a_n^{-1}\|\theta_{n-1}-\theta_n\|)+ \sum_{n=1}^{\infty} a_n^{2+\delta}\|A_n(Y_n)\|_N(a_n^{-1}\|\theta_{n-1}-\theta_n\|)$$

$$+\Big(\sum_{n=1}^{\infty} a_n^{1+\delta}\Big)^{1/2}\Big(\sum_{n=1}^{\infty} a_n^{1+\delta}\|\Gamma_n(Y_n)\|^2\Big)^{1/2}$$

$$<\infty \qquad a.s..$$

Hence we obtain (3.37) by applying Lemma 2.3.

Next we shall prove

$$(3.40)\qquad \lim_{n\to\infty} a_n^{1-\delta}\sum_{k=1}^{n} a_k^{\hat{\delta}-1}\|x_{k-1}-x_k\|\|S_k\|_N=0 \qquad a.s..$$

According to (3.28) and (3.29), it follows that

$$(3.41)\qquad a_n^{1-\delta}\sum_{k=1}^{n} a_k^{\hat{\delta}-1}\|x_{k-1}-x_k\|\|S_k\|_N$$

$$\leqq a_n^{1-\delta}Z_0\sum_{k=1}^{\infty} a_k^{\hat{\delta}}\|S_k\|_N\|A_k(Y_k)\|_N\sum_{i=1}^{k} a_i\|R_i\|$$

$$+a_n^{1-\delta}\sum_{k=1}^{n} a_k^{\hat{\delta}}\|R_k\|\|S_k\|_N \qquad a.s..$$

By (3.33), it follows that there exists a positive random variable Z_1 which is $a.s.$ finite, such that

$$(3.42)\qquad \|S_n\|_N\leqq Z_1 \qquad a.s. \qquad \text{for} \quad n\geqq 1.$$

Hence, (3.30), (3.39), (3.42) and Lemma 2.3 imply

$$(3.43)\qquad \lim_{n\to\infty} a_n^{1-\delta}\sum_{k=1}^{n} a_k^{\hat{\delta}}\|R_k\|\|S_k\|_N=0 \qquad a.s..$$

And (3.39) and Lemma 2.3 imply

$$(3.44)\qquad \lim_{n\to\infty} a_n^{\delta}\sum_{k=1}^{n} a_k\|R_k\|=0 \qquad a.s..$$

Hence, (3.20), (3.30) and (3.44) imply

$$(3.45)\qquad \sum_{n=1}^{\infty}\Big(a_n^{1-\delta}a_n^{\hat{\delta}}\|A_n(Y_n)\|_N\sum_{k=1}^{n} a_k\|R_k\|\Big)$$

$$\leqq \sum_{n=1}^{\infty}\Big(a_n^{1+\delta}\|A_n(Y_n)\|_N\cdot a_n^{\delta}\sum_{k=1}^{n} a_k\|R_k\|\Big) \qquad a.s..$$

Therefore, by applying Lemma 2.3, we obtain

$$(3.46)\qquad \lim_{n\to\infty} a_n^{1-\delta}Z_0\sum_{k=1}^{n} a_k^{\hat{\delta}}\|S_k\|_N\|A_k(Y_n)\|_N\sum_{i=1}^{k} a_i\|R_i\|=0 \qquad a.s..$$

Hence, (3.40) holds. And (V) is proved.

Thus the proof of Theotem is completed.

REMARKS. (1) Suppose that $\sup_{1\leqq n}\|E[A_n(Y_n)]\|_N<\infty$ and the assumptions in Theotem are satisfied. Then it holds that

$$(3.47)\qquad \lim_{n\to\infty} M_n(X_n)=\mathbf{0} \qquad a.s..$$

(2) In this paper, it is assumed that $A_n(\cdot)$ and $\Gamma_n(\cdot)$ are not depend on $x\in R^N$. But it will be required to extend the present work to the case when $\Phi_n(x, y)$ can be expressed in the form of $\Phi_n(x, y)=(x-\theta_n)A_n(x, y)+\Gamma_n(x, y)$. In a fothcoming paper we shall discuss the abeve case.

4. Applibations

I. *The learing problem for a pattern classification.* Let ue consider the problem of estimating an unknown optimal discriminant function ([4], [6], [7]),

$$D^{(n)}(z)=p_1^{(n)}f_1^{(n)}(z)-p_2^{(n)}f_2^{(n)}(z) \tag{4.1}$$

for sufficiently large n, where $(p_1^{(n)}, p_2^{(n)})$ is a priori distribution on categories set $\{1,2\}$ at instant n and $f_1^{(n)}(\cdot)$, $f_2^{(n)}(\cdot)$ are the probability density function with respect to Lebesgue measure on R^{M-1} under the categories 1 and 2 respectively. Let $\{(Z_n, \Delta_n)\}_{n=1}^{\infty}$ be a training sequence, where Z_n is an S-valued random vector and has probability density function $f_1^{(n)}(\cdot)$ $(f_2^{(n)}(\cdot))$ if Z_n is from the class 1(2), and Δ_n is a $\{1,2\}$-valued random variable whose distribution is equal to $(p_1^{(n)}, p_2^{(n)})$.

By many auteors ([4], [6], [7] etc.), it has been discussed the case when the training sequence is independently distributed random variables. In this paper, we shall discuss the case when (Z_n, Δ_n)'s are dependently distributed. And the learning problem considered here is to estimate the optimal discriminant function (4.1) on tne bases of the dependently distributed training sequence $\{(Z_n, \Delta_n)\}_{n=1}^{\infty}$.

Firstly, we assume that there exists a system of orthonormal functions $\{\varphi_i(\cdot)\}_{i=1}^{N}$ defined on the pattern space $S\subset R^{M-1}$, such that

$$\int_S \varphi_i(z)\varphi_j(z)dz=\begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i\neq j\,. \end{cases} \tag{4.2}$$

Secondly, we approximate $D^{(n)}(\cdot)$ at instant n by a finite series

$$\sum_{i=1}^{N} c_i^{(n)}\varphi_i(z) \qquad (=\langle \boldsymbol{c}^{(n)}, \boldsymbol{\varphi}(z)\rangle) \tag{4.3}$$

which minimizes a quantity $J_n(\boldsymbol{c})$ defined by

$$J_n(\boldsymbol{c})=\int_S (D^{(n)}(z)-\langle \boldsymbol{c}, \boldsymbol{\varphi}(z)\rangle)^2 dz \tag{4.4}$$

where $\boldsymbol{c}^{(n)}=(c_1^{(n)}, c_2^{(n)}, \cdots, c^{(n)})$, $\boldsymbol{c}=(c_1, c_2, \cdots, c_N)$ and $\boldsymbol{\varphi}(z)=(\varphi_1(z), \varphi_2(z), \cdots, \varphi_N(z))$. By differentiating (4.4) with respect c_i $(i=1, 2, \cdots, N)$, equating the derivative to zero we have

$$\mathrm{E}[\boldsymbol{c}^{(n)}-d(\Delta_n)\boldsymbol{\varphi}(Z_n)]=\mathbf{0}\,, \qquad n=1, 2, \cdots, \tag{4.5}$$

where

$$d(\delta)=\begin{cases} 1 & \text{if } \delta=1 \\ -1 & \text{if } \delta=2\,. \end{cases}$$

Hence our problem becomes to find the solution of the equation

$$\mathrm{E}[\boldsymbol{c}-d(\Delta_n)\varphi(Z_n)]=\mathbf{0} \tag{4.6}$$

for sufficiently large n. In Section 3, for $n=1, 2, \cdots$, let us put

$$x=\boldsymbol{c}\in R^N\,, \quad y=(z, \delta)\in S\times\{1, 2\}\subset R^M\,, \quad \theta_n=\boldsymbol{c}^{(n)}\,, \tag{4.7}$$

(4.8) $$\Phi_n(x,\ y)=\Phi_n(\boldsymbol{c},\ (z,\ \delta))=(\boldsymbol{c}-\boldsymbol{c}^{(n)})+(\boldsymbol{c}^{(n)}-d(\delta)\boldsymbol{\varphi}(z))\,,$$

(4.9) $$A_n(y)\equiv I$$

and

(4.10) $$\Gamma_n(y)=\Gamma_n((z,\ \delta))=\boldsymbol{c}^{(n)}-d(\delta)\boldsymbol{\varphi}(z)\,.$$

And we construct the following procedure,

(4.11) $$\boldsymbol{C}_0=\text{an arbitrary constant vector in } R^N$$
$$\boldsymbol{C}_{n+1}=\boldsymbol{C}_n-(n+1)^{-1}\{\boldsymbol{C}_n-d(\Delta_{n+1})\boldsymbol{\varphi}(Z_{n+1})\}\,,$$
$n=0,1,2,\cdots$.

As for the above procedure (4.11) we have the following theorem which is the direct application of Theorem in Section 3.

THEOREM 4.1. *Let the following conditions be satisfied.*

(i) $\sup_{\substack{1\leq i\leq N\\ 1\leq n}} \mathrm{E}[\varphi_i^2(Z_n)]<\infty$.

(ii) $\lim_{n\to\infty}(n+1)\left\{\int_S(D^{(n+1)}(z)-D^{(n)}(z))^2dz\right\}^{1/2}=0$.

(iii) $\overline{\lim_{n\to\infty}}\ \phi\left(\bigvee_{i=1}^{n}\mathcal{A}_i,\ \bigvee_{i=n+n_0}^{\infty}\mathcal{A}_i\right)<1$ *for some* n_0, *where* $\mathcal{A}_i=\sigma(Z_i,\ \Delta_i)$ $(i=1,2,\cdots)$.

(iv) $\sum_{n=1}^{\infty}\rho_n^{1/2}<\infty$.

Then it follows that

(4.12) $$\lim_{n\to\infty}\|\boldsymbol{C}_n-\boldsymbol{c}^{(n)}\|=0 \qquad a.s..$$

PROOF. By the remark (3) of Assumptions in Section 3, (A1) to (A3) follows. And (4.6) implies (B5). (i) implies (B6). Moreover (4.9) implies (B2), (B3) and (B4). Next, by (4.2), (4.5) and using the Schwarz inequality, we have

(4.13) $$\begin{aligned}&\|\boldsymbol{c}^{(n)}-\boldsymbol{c}^{(n-1)}\|\\ &=\left[\sum_{i=1}^{N}\left\{\int_S(D^{(n)}(z)-D^{(n+1)}(z))\varphi_i(z)dz\right\}^2\right]^{1/2}\\ &\leqq\left[\sum_{i=1}^{N}\left\{\int_S\varphi_i^2(z)dz\right\}\left\{\int_S(D^{(nn)}(z)-D^{(n+1)}(z))^2dz\right\}\right]^{1/2}\\ &=N^{1/2}\left\{\int_S(D^{(n)}(z)-D^{(n+1)}(z))^2dz\right\}^{1/2}.\end{aligned}$$

Hence, (ii) implies (B7).

Thus the all assumptions in Section 3 are satisfied.

II. *The construction of an unknown function.* Let us consider the problem of estimating an unknown function on the basis of inputs and outputs data ([1], [4]). An outcome in this problem is denoted by a pair $(\alpha,\ \beta)$. The element α is an observable input and β is an observable output corresponding to α. Let $S_0\subset R^M$, $S_1\subset R$

denote respectively an input space and an output space. For the input S_0-valued random vectors $\alpha^1, \alpha^2, \cdots$, we consider a sequence, (α^1, β^1) $(\alpha^2, \beta^2), \cdots$. Here we assume that the output β^n observed corresponding to the input α^n is expressed by

(4.14) $$\beta^n = f(\alpha^n), \qquad n=1, 2, \cdots,$$

where $f(\cdot)$ is a Borel function defined on S_0 and the true type of f is unknown to us.

By many authors ([1], [4], etc.), it has been discussed the case when the inputs variables are independently distributed. Here we shall discuss the case when α^n's are denpendently distributed.

Let us take a system of Borel functions $\{\varphi_i(\cdot)\}_{i=1}^N$ defined on $S_0 \subset R^M$. At each n, we would like to approximate the unknown function f by a finite series

(4.15) $$\sum_{i=1}^N c_i^{(n)} \varphi_i(\alpha) \qquad (= \langle \boldsymbol{c}^{(n)}, \boldsymbol{\varphi}(\alpha) \rangle)$$

which minimizes the quantity $J_n(\boldsymbol{c})$ defined by

(4.16) $$J_n(\boldsymbol{c}) = \mathrm{E}[\beta^n - \langle \boldsymbol{c}, \boldsymbol{\varphi}(\alpha^n) \rangle]^2$$

where $\boldsymbol{c}^{(n)} = (c_1^{(n)}, \cdots, c_N^{(n)})$, $\boldsymbol{c} = (c_1, \cdots, c_N)$ and $\boldsymbol{\varphi}(\alpha) = (\varphi_1(\alpha), \cdots, \varphi_N(\alpha))$. By differentiating (4.16) with respect c_i $(i=1, \cdots, N)$, equating the derivative to zero, we obtain

(4.17) $$\mathrm{E}[\boldsymbol{c}\boldsymbol{\varphi}(\alpha^n)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha^n) - \beta^n \boldsymbol{\varphi}(\alpha^n)] = \boldsymbol{0}, \qquad n=1, 2, \cdots,$$

where A^{T} denotes the transpose of a matrix A. Hence the above problem to find the solution of the equation (4.17) for sufficiently large n.

Let ue suppose that $\mathrm{E}[\boldsymbol{\varphi}(\alpha^n)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha^n)]$'s are positive definite matrices. Then the equation (4.17) has the unique solution $\boldsymbol{c}^{(n)}$ which minimizes the quantity $J_n(\boldsymbol{c})$ for $n=1, 2, \cdots$. And in Section 3, let us put

(4.18) $$x = \boldsymbol{c} \in R^N, \quad y = \alpha \in S_0 \subset R^M, \quad \theta_n = \boldsymbol{c}^{(n)},$$

(4.19) $$A_n(\alpha) = \boldsymbol{\varphi}(\alpha)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha)$$

and

(4.20) $$\Gamma_n(\alpha) = \boldsymbol{c}^{(n)} \boldsymbol{\varphi}(\alpha)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha) - f(\alpha) \boldsymbol{\varphi}(\alpha).$$

And we construct the following procedure,

(4.21) $$\begin{aligned} &\boldsymbol{C}_0 = \text{an arbitrary constant vector in } R^N, \\ &\boldsymbol{C}_{n+1} = \boldsymbol{C}_n - (n+1)^{-1} \{\boldsymbol{C}_n \boldsymbol{\varphi}(\alpha^n)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha^n) - \beta^n \boldsymbol{\varphi}(\alpha^n)\}, \end{aligned}$$

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

As for the above procedure (4.21) we have the following theorem which is the direct application of Theorem in Section 3.

THEOLEM 4.2. *Suppose that the distributions of α^n's do not depend on n. And suppose that $\mathrm{E}[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}} \boldsymbol{\varphi}(\alpha^1)]$ is a positive definite matrix and the following conditions are satisfied.*

(i) $\mathrm{E}[f(\alpha^1) \varphi_i(\alpha^1)]^2 < \infty$ for $i=1, 2, \cdots N$.

(ii) $E[\varphi_i(\alpha^1)\varphi_j(\alpha^1)]^2<\infty$ for $i, j=1, 2, \cdots, N$.

(iii) $\overline{\lim_{n\to\infty}}\, \phi\Big(\bigvee_{i=1}^{n} \mathcal{A}_i, \bigvee_{i=n+n_0}^{\infty} \mathcal{A}_i\Big)<1$ *for some* n_0, *where* $\mathcal{A}_i=\sigma(\alpha^i)$ $(i=1, 2, \cdots)$.

(iv) $\sum_{n=1}^{\infty} \rho_n^{1/2}<\infty$.

Then the equation (4.17) *has the unique solution* $\boldsymbol{c}^*$ *and it holds that*

$$\lim_{n\to\infty}\|\boldsymbol{C}_n-\boldsymbol{c}^*\|=0 \qquad a.s.. \tag{4.22}$$

PROOF. By the remark (3) of Assumptions, (A1) to (A3) follows. Noting $a_n=n^{-1}$, (ii) implies (B2). Note that

$$\begin{aligned}\langle \boldsymbol{c}\boldsymbol{\varphi}(\alpha^{n\mathrm{T}}\boldsymbol{\varphi}(\alpha^n), \boldsymbol{c}\rangle &= \sum_{i=1}^{N}\Big(\sum_{j=1}^{N} \varphi_i(\alpha^n)\varphi_j(\alpha^n)c_j\Big)c_i \\ &=\Big(\sum_{i=1}^{N} c_i\varphi_i(\alpha^n)\Big)^2 \qquad \text{for all} \quad \boldsymbol{c}=(c_1, \cdots c_N)\in R^N.\end{aligned} \tag{4.23}$$

Hence, (4.23) implies (B3). And the fact that $E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)]$ is a positive definite matrix implies

$$\langle \boldsymbol{c}E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)], \boldsymbol{c}\rangle \geqq \alpha_0\|\boldsymbol{c}\|^2 \tag{4.24}$$

for all $\boldsymbol{c}\in R^N$, where α_0 is the minimum eigenvalue of $E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)]$. Hence, (B4) holds. Since $E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)]$ is a positive definite matrix, there exists the inverse matrix. Hence, the euqation (4.17) has the unique solution $\boldsymbol{c}^*$, that is

$$\boldsymbol{c}^*=E[f(\alpha^1)\boldsymbol{\varphi}(\alpha^1)](E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)])^{-1}. \tag{4.25}$$

Therefore, (i) and (ii) imply (B6). And noting that the distributions of α^n's do not depend on n, (4.25) implies (B7).

Thus the all assumptions in Section 3 are satisfied.

REMARK. If α^1 has a positive probability density function with respect to Lebesgue measure on R^N and $\{\varphi_i(\cdot)\}_{i=1}^{N}$ is a system of linearly independent continuous functions then it is easily seen that $E[\boldsymbol{\varphi}(\alpha^1)^{\mathrm{T}}\boldsymbol{\varphi}(\alpha^1)]$ is a positive definite matrix.

References

[1] AIZERMAN, M.A., BRAVERMAN, E.M. and ROZONOER, L.I.: *The method of potential function for the problem of restoring the characteristic of a function converter from randomly observed points*, Autom. and Remote Control, **25** (1964), 1546-1556.

[2] IOSIFESCU, M. and THEODORESCU, R.: Rendom processes and learning, Springer, 1969.

[3] LOÉVE, M.: Probability theory, Van Nostrand, New York, 1960.

[4] MENDEL, J.M. and FU, K.S.: Addaptive, learning and pattern recognition system, Theory and applications, Academic Press, 1970.

[5] ROBBINS, H. and MONRO, S.: *A stochastics approximation method*, Ann. Math. Stat., **22** (1951), 400-407.

[6] WATANABE, M.: *On a asymptotically optimal algorithms for a pattern classification problems*, Bull. Math. Stat., **15**, No. 3-4 (1973), 31-48.

[7] WATANABE, M.: *On convergences of asymptotically optimal discriminant functions for pattern classification problems*, Bull. Math. Stat., **16**, No. 1-2 (1974), 23-34.

[8] WATANABE, M.: *On Robbins-Monro stochastic approximation method with time varying observations*, Bull. Math. Stat., **16**, No. 3-4 (1975), 73-91.

[9] WATANABE, M.: *On Robbins-Monro stochastic approximation method with a sequence of dependent random variables*, to appear in Studia Sci. Math. Hungar..

[10] WATANABE, M.: *On Robbins-Monro procedure with a sequence dependent random elements in Hilbert space*, to appear in Proceedings of the 41st Session of the International Statistical Institute in New Delhi 5-15 December 1977.

[11] WATANABE, M.: *On Robbins-Monro stochastic approximation method with a sequence of dependent random variables* II, Fukuoka Univ. Sci. Reports, Vol. 8, No. 2(1978), 57-64.