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 \( E[Y_n(x)] = M_n(x) \), is unknown to us. Assuming that the equation \( M_n(x) = 0 \) has a solution \( x = \theta_n \) for each \( n = 1, 2, \ldots \), it is desire to estimate \( \theta_n \) for sufficiently large \( n \) on the basis of observed values \( Y_1(X_0), Y_2(X_1), \ldots, Y_{n+1}(X_{n}), \ldots \) at the points \( X_0, X_1, \ldots, X_n, \ldots \) which are produced by the following recurrence relation,

\[
\begin{align*}
X_0 &= \text{an arbitrary constant vector in } R^n, \\
X_{n+1} &= X_n - a_{n+1} Y_{n+1}(X_n) \quad \text{for } n = 0, 1, 2, \ldots,
\end{align*}
\]

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 \( X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n \) coincides with the distribution of \( Y_{n+1}(x_n) \), consequently

\[
E[Y_{n+1}(X_n)|X_0, X_1, \ldots, X_n] = M_{n+1}(X_n) \quad \text{a. s.}
\]

In our previous paper [8], we showed that \( \lim_{n \to \infty} E\|X_n - \theta_n\|^2 = 0 \) 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 unknown 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 \(\Phi_n(x, \cdot)\)'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

\[
E[\Phi_n(x, Y_n)] = 0
\]

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 \(\Gamma_n\)'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 \(\Gamma_n\)'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 Preliminaries

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

\[
(2.1) \quad M_n(x) = 0
\]

where \( M_n(x) = \mathbb{E}[\Phi_n(x, y_n)] = (\mathbb{E}[\Phi_n^{(1)}(x, y_n)], \ldots, \mathbb{E}[\Phi_n^{(N)}(x, y_n)]) \), \( n = 1, 2, \ldots, x \in \mathbb{R}^N \).

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 \( x \in \mathbb{R}^N \) and \( n = 1, 2, \ldots \).

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 \( \mathbb{R}^N \) and define \( X_1, X_2, \ldots \) by the following recurrence relation:

\[
(2.2) \quad X_{n+1} = X_n - a_n \Phi_{n+1}(X_n, y_{n+1})
\]

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

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

\[
(2.3) \quad \phi(\mathcal{A}', \mathcal{A}^n) = \sup_{A \in \mathcal{A}'} (\operatorname{ess} \sup_{\omega \in \Omega} |P(A | A') - P(A)|).
\]

And we put

\[
(2.4) \quad \rho_n = \sup_{m \leq n} \phi(\mathcal{A}_m, \mathcal{A}_{m+n})
\]

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

With all vectors considered as elements of \( \mathbb{R}^N \) we adopt the following notations. If \( x, y \in \mathbb{R}^N \) then \( \langle x, y \rangle \) will denote their inner product. The norm of \( x \in \mathbb{R}^N \) 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

\[
(2.5) \quad \|A\| = \sup_{i \in [1, N]} \|x_i A\|.
\]

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 \( \|A\|_\infty \) by \( \|A\|_\infty = \max_{1 \leq i, j \leq N} |a_{ij}| \) 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

\[
(2.6) \quad C_1 \|A\|_\infty \leq \|A\|_N \leq C_2 \|A\|_\infty
\]

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

**PROOF.** Put \( x = (x_1, x_2, \ldots, x_N) \) and \( e_i = (0, \ldots, 0, 1, \ldots, 0) \) for \( i = 1, 2, \ldots, N \). Then, using the Schwarz inequality, it follows that
\( \|A\|_y = \sup_{|z|=1} \|zA\| \)

\[ = \sup_{|z|=1} \left\{ \sum_{j=1}^{N} \left( \sum_{i=1}^{N} a_{ij}x_j \right)^2 \right\}^{1/2} \]

\[ \leq N^{1/2} \sup_{|z|=1} \left\{ \max_{1 \leq i < N} \left( \sum_{j=1}^{N} a_{ij}x_j \right)^2 \right\}^{1/2} \]

\[ \leq N^{1/2} \sup_{|z|=1} \left\{ \max_{1 \leq i < N} \left( \sum_{j=1}^{N} a_{ij} \right)^2 \right\}^{1/2} \]

\[ = N^{1/2} \max_{1 \leq i, j \leq N} |a_{ij}|. \]

And we have

\[ \|A\|_y \geq \|e_iA\| \]

\[ = \left( \sum_{j=1}^{N} a_{ij}^2 \right)^{1/2} \]

\[ \geq |a_{ij}| \quad \text{for all } i, j = 1, 2, \ldots, N. \]

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) \( x_{n+1} \leq (1-a_{n+1})x_n + a_{n+1}b_{n+1} + K_{n+1} \) for \( n = 0, 1, \ldots \),

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

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

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

**Remark.** Assume \( 0 \leq a_n < 1 \) for all \( n \geq 1 \). And define

\[ x_n = \sum_{k=1}^n a_kb_k \prod_{l=k+1}^\infty (1-a_l) + \sum_{k=1}^n K_k \prod_{l=k+1}^\infty (1-a_l) \]

for \( n = 1, 2, \ldots \). 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 \( \lim_{n \to \infty} b_n \sum_{k=1}^n x_k b_k = 0 \).

**Proof.** It is omitted (see [3]).

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

\[ x_{n+1} = T_{n+1}x_n + u_{n+1} + v_{n+1} + w_{n+1}, \quad n = 0, 1, 2, \ldots, \]

and

\[ \sum_{n=0}^\infty \|u_n\| < \infty \quad \text{and} \quad \sum_{n=1}^\infty \|v_n\| < \infty. \]
An almost sure convergence theorem

where \( \{T_n\}_{n=1}^\infty \) is a sequence of bounded linear operators of \( X \) 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

1. \( \lim_{n \to \infty} \alpha_n = 0 \), \( \sum_{n=1}^\infty \alpha_n = \infty \),
2. \( \sup_{n \geq 1} \beta_n |\alpha_n^{-1} - \alpha_n^{-1}| \leq K_1 \),
3. \( \sup_{n \geq 1} \beta_n \leq K_2 \), \( \sup_{n \geq 1} \beta_n \beta_n^{-1} \leq K_3 \),
4. \( \|T_n \cdots T_{n-1} T_{m+1}\| \leq K_4 \prod_{k=n+1}^{\infty} |1 - \alpha_k| \) for \( n > m \geq N_0 \),
5. \( \|I - T_n\| \leq \alpha_n \gamma_n \) for \( n > N_0 \), where \( I \) is the identity operator,
6. \( \sup_{n \geq 1} \beta_n \gamma_n \leq K_5 \) or \( \sum_{n=1}^\infty \alpha_n \beta_n \gamma_n < \infty \),
7. \( \lim_{n \to \infty} \alpha_n^{-1} u_n = 0 \),
8. \( \sum_{n=1}^\infty \|u_n\| < \infty \),
9. \( \lim_{n \to \infty} \|\gamma_{n+1}^{-1} \sum_{k=1}^{n} \alpha_k^{-1} w_k\| = 0 \).

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

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

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

From (i) and (iv) we have

\[
(2.12) \quad \lim_{n \to \infty} \|T_0^n x_0\| = 0 .
\]

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

\[
(2.13) \quad \lim_{n \to \infty} \sum_{k=1}^{n} \|T_k^n u_k\| = 0
\]
and

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

Next we shall prove that

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

Note that \( \sum_{k=1}^{n} T_k^n w_k = \sum_{k=n+N_0}^{n} T_k^n w_k + \sum_{k=1}^{n+N_0-1} T_k^n w_k \). Firstly, (i) and (iv) imply
Hence, in order to show (2.15), we have only to prove that

\[
\lim_{n \to \infty} \sum_{k=N_0}^{n} \| T_k^{(n)} w_k \| = 0.
\]

Let us put

\[
\begin{align*}
\{ s_{N_0 - 1} = 0, & \text{ where } 0 \text{ is the zero element in } \mathcal{X}, \\
s_{n} = \alpha_n \sum_{k=N_0}^{n} \alpha_{k-1} w_k & \text{ for } n \geq N_0
\end{align*}
\]

and

\[
\begin{align*}
y_n &= \sum_{k=N_0}^{n} T_k^{(n)} w_k.
\end{align*}
\]

Then, noting \( \alpha_{n-1} s_n = \alpha_{n-1} s_{n-1} = \alpha_{n-1} w_n \), we obtain

\[
y_n = s_n + \sum_{k=N_0}^{n-1} (\alpha_k T_k^{(n)} - \alpha_{k+1} T_{k+1}^{(n)}) \alpha_{k-1} s_k
\]

\[
= s_n + \sum_{k=N_0}^{n-1} T_{k+1}^{(n)} (T_{k+1} - I) s_k + \sum_{k=N_0}^{n-1} \alpha_{k+1} (\alpha_{k+1} - \alpha_{k+2}) T_{k+1}^{(n)} s_k.
\]

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

\[
\begin{align*}
\| y_n \| & \leq \| s_n \| + K_n \sum_{k=N_0}^{n-1} \alpha_{k+1} \beta_k \beta_{k+1} s_k \| s_k \| \prod_{l=k+2}^{n} (1 - \alpha_l)
\end{align*}
\]

\[
+ K_n \sum_{k=N_0}^{n-1} \alpha_{k+1} \beta_{k-1} s_k \| s_k \| \prod_{l=k+2}^{n} (1 - \alpha_l).
\]

Note that \( \beta_{k+1} s_k \| s_k \| = (\beta_k \beta_{k+1} s_k \| s_k \|) \) 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 \( \alpha_n = \alpha^1_n = n^{-6} \) \( (0 < \delta < 1) \) for \( n = 1, 2, \ldots \), 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 \( |\alpha_n \alpha_{n+1}^\delta - 1| \leq K_n \alpha_{n+1}^\delta \). 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 \( (\Omega, \mathcal{A}, \mathcal{P}) \). Suppose that there exist a positive integer \( n_0 \) and a increasing sequence of positive numbers \( \{ k_n \}_{n=1}^{\infty} \), such that

\[
\begin{align*}
(1) \lim_{n \to \infty} \phi\left( \sum_{i=1}^{n} \mathcal{A}_i, \sum_{i=n+n_0}^{\infty} \mathcal{A}_i \right) < 1,
\end{align*}
\]
An almost sure convergence theorem

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

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

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

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

\[
\lim_{n \to \infty} k_n^{-1} \sum_{i=1}^{n} (X_i - EX_i) = 0 \quad a.s.
\]

Remark. 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) \( \sum_{n=1}^{\infty} k_n^2 E \|X_n - EX_n\|^2 < \infty \) \( \left( \sum_{n=1}^{\infty} k_n^2 E \|X_n - EX_n\|^2 < \infty \right) \).

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

\[
\lim_{n \to \infty} k_n^{-1} \sum_{i=1}^{n} (X_i - EX_i) = 0 \quad a.s.
\]

\[
\left( \lim_{n \to \infty} k_n^{-1} \sum_{i=1}^{n} (X_i - EX_i) \right) = 0 \quad a.s.
\]

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 \( \sigma \)-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 \( \delta \) 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) \( \sup_{n} a_n^{\delta} (a_n^{1/2} - a_{n+1}^{1/2}) < \infty \).

B1) For each \( n = 1, 2, \ldots \), \( x \in R^N \) 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 \( \theta_n \) is a constant vector in \( R^N \), \( A_n(y) = (A_n^i(y)) \) is a \( N \times N \)-matrix whose \( (i, j) \)-th element \( A_n^i(y) \) is a Borel function defined on \( R^M \) and \( \Gamma_n(y) = (\Gamma_n^u(y), \ldots, \Gamma_n^v(y)) \)
is a vector in $R^N$ whose $i$-th element $I^{(i)}_n(y)$ is a Borel function defined on $R^M$.

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

(B3) $\inf \langle xA_n(Y_n), x \rangle \geq 0$ \ a.s. \ for \ $n=1, 2, \ldots$.

(B4) There exists a positive number $\alpha_0$ such that
$$\inf \langle xE[A_n(Y_n)], x \rangle \geq \alpha_0 \ \text{for} \ n=1, 2, \ldots.$$ 

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

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

(B7) $\lim_{n \to \infty} a_n^{1+\delta}E[I_n(Y_n)] = 0$.

(C1) There exists a positive integer $n_0$ such that
$$\inf_{\bigvee_{i \in I} A_i} \phi\left(\frac{n}{t}, \bigvee_{i=n+n_0}^{\infty} A_i\right) < 1,$$
where $\bigvee_{i \in I} A_i$ denotes the $\sigma$-field generated by the join of the $\sigma$-fields $A_i$, $i \in I$.

(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 following condition,

(B5') $E[I_n(Y_n)] = 0 \ \ (n=1, 2, \ldots)$

hold then $\theta_n \ (n=1, 2, \ldots)$ 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 $\lim_{n \to \infty} \phi\left(\bigvee_{i \in I} A_i, \bigvee_{i=n+n_0}^{\infty} A_i\right) = 0$ for all $n_0$ and $\rho_n = 0$ for all $n$. Hence, (C1) and (C2) hold.

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

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

The following theorem is our main result in this paper.

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

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

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

(3.2) $X_{n+1} - \theta_{n+1} = (X_n - \theta_n)(I - a_{n+1}A_{n+1}(Y_{n+1}))) + a_{n+1}(a_n^{-1}(\theta_n - \theta_{n+1}))(I - a_{n+1}A_{n+1}(Y_{n+1})))$ 
$- I_{n+1}(Y_{n+1}))$, \ \ \ \ n=0, 1, 2, \ldots$. 

Let us put
\[ X_n - \theta_n = x_n, \quad E[A_n(Y_n)] = \hat{A}_n, \quad E[I_n(Y_n)] = \hat{I}_n \] and \[ I - a_n E[A_n(Y_n)] = T_n. \]
Then we have
\begin{equation}
X_{n+1} = x_n T_n + a_{n+1} U_{n+1} + a_{n+1} V_{n+1} - a_{n+1} W_{n+1}
\end{equation}
where
\begin{align}
U_{n+1} &= a_{n+1} \left( \theta_n - \theta_{n+1} \right) - \hat{I}_{n+1}, \\
V_{n+1} &= - \left( \theta_n - \theta_{n+1} \right) \hat{A}_{n+1}
\end{align}
and
\begin{align}
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})) \\
&\quad - I_{n+1}(Y_{n+1}) + \hat{I}_{n+1}.
\end{align}

By virtue 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 \( \alpha_n = \alpha_0 a_n \) 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 \( \{ y_n \} \) such that
\[ \| I - T_n \| \leq \alpha_0 a_n y_n \quad \text{for } n=1, 2, \cdots \]
and
\[ \sum_{n=1}^{\infty} \alpha_0^{1+\delta} y_n < \infty. \]

(II); There exist a positive integer \( N_0 \) and a positive number \( K \), such that
\[ T_m T_{m+1} \cdots T_n \leq K \prod_{k=m+1}^{n} (1 - \alpha_0 a_k) \quad \text{for } n>m \geq N_0, \]
where \( 1 - \alpha_0 a_k > 0 \) (\( k \geq N_0 \)).

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

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

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

PROOF OF (I). Put \( \hat{\alpha}_n = \hat{\alpha}_0 a_n \| \hat{A}_n \| \) and \( \| \hat{A}_n \| \leq y_n \), \( n=1, 2, \cdots \). Then we have
\begin{equation}
\| I - T_n \| = a_n \| \hat{A}_n \| = \alpha_0 a_n y_n, \quad n=1, 2, \cdots.
\end{equation}
The fact that \( \| \hat{A}_n \| \leq E \| A_n(Y_n) \| \) and the Schwarz inequality imply
\begin{equation}
\sum_{n=1}^{\infty} a_n^{1-\delta} y_n = \alpha_0^{-1} \sum_{n=1}^{\infty} a_n^{1+\delta} \| \hat{A}_n \| \leq a_0^{-\frac{1}{2}} \left( \sum_{n=1}^{\infty} a_n^{1+\delta} E \| A_n(Y_n) \| \right)^{\frac{1}{2}}.
\end{equation}
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

\[
1 - 2\alpha_0 a_n \geq k_0^{-1} \quad \text{for } n \geq N_0.
\]

(B4) and (3.9) imply

\[
1 - 2\alpha_0 a_n \leq \frac{1}{n} + \frac{\lambda}{n-2\alpha_0 a_n} \leq 2\alpha_0 a_n \leq (1-2\alpha_0 a_n) \left( 1 + (1-2\alpha_0 a_n)^{-1}(a_n^2 + \|A_n\|^2) \right) \leq (1-2\alpha_0 a_n)(1+k_0 a_n 2\|A_n\|^2)
\]

for \( n \geq N_0 \). Hence we have

\[
\|I - a_n A_n\| \leq \left\{ \prod_{k=m+1}^{n} \left( 1 + k_0 a_n 2\|A_k\|^2 \right) \right\}^{1/2} \prod_{k=m+1}^{n} (1 - \alpha_0 a_k)
\]

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

\[
\sup_{1 \leq m \leq n} \prod_{k=m+1}^{n} (1 + k_0 a_n 2\|A_k\|^2) < \infty.
\]

Hence (II) holds.

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

PROOF OF (IV). Note that \( a_n \|V_n\| \leq a_n \|\theta_{n-1} - \theta_n\| \cdot a_n \|A_n\| \) 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

\[
\hat{\gamma}_{n+1} - \gamma_{n+1}(Y_{n+1}) = W_1, n+1,
\]

\[
(\theta_n - \theta_{n+1})(\hat{A}_{n+1} - A_{n+1}(Y_{n+1})) = W_2, n+1
\]

and

\[
x_n(\hat{A}_{n+1} - A_{n+1}(Y_{n+1})) = W_3, n+1.
\]

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

\[
\lim_{n \to \infty} a_n^{1/2} \left\| \sum_{k=1}^{n} W_{i,k} \right\| = 0 \quad a.s.
\]

for \( i = 1, 2, 3 \). The fact that \( 0 < \delta \leq 1/7 \), it follows that

\[
\frac{1 + \delta}{2} < 1 - \delta.
\]

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

\[
\lim_{n \to \infty} a_n^{1/2} \left\| \sum_{k=1}^{n} W_{i,k} \right\| = 0 \quad a.s.
\]

Next we have
(3.19) \[ a_n^{-\delta} \left\| \sum_{k=1}^{n} W_{z,k} \right\| \leq a_n^{-\delta} \sum_{k=1}^{n} a_k^{\gamma} \| \theta_{k-1} - \theta_k \| \cdot a_k \| A_k \|_N + a_n^{-\delta} \sum_{k=1}^{n} a_k^{\gamma} \| \theta_{k-1} - \theta_k \| \cdot a_k \| A_k(Y_k) \|_N. \]

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

(3.20) \[ \sum_{n=1}^{\infty} a_n^{-\delta} \| A_n(Y_n) \|_N \leq \left( \sum_{n=1}^{\infty} a_n^{\gamma} \right)^{1/2} \left( \sum_{n=1}^{\infty} a_n^{-\delta} \| A_n(Y_n) \|_N^2 \right)^{1/2} < \infty \quad a.s. \]

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

(3.21) \[ \lim_{n \to \infty} a_n^{-\delta} \left\| \sum_{k=1}^{n} W_{z,k} \right\| = 0 \quad a.s. \]

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

(3.22) \[ \lim_{n \to \infty} a_n^{-\delta} \left\| \sum_{k=1}^{n} W_{n,k} \right\| = 0 \quad a.s. \]

From (3.2), we have

(3.23) \[ \| x_{n+1} \| \leq \| x_n \| \cdot \| I - a_{n+1} A_{n+1}(Y_{n+1}) \|_N + a_{n+1} \| R_{n+1} \|, \quad n = 0, 1, 2, \ldots \]

where

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

And (A3) implies

(3.25) \[ \| I - a_n A_n(Y_n) \|_N^{1/2} \leq 1 + a_n \| A_n(Y_n) \|_N^{1/2} \quad a.s., \quad n = 1, 2, \ldots \]

Hence we have

(3.26) \[ \| x_{n+1} \| \leq \| x_n \| (1 + a_{n+1} \| A_{n+1}(Y_{n+1}) \|_N^{1/2}) + a_{n+1} \| R_{n+1} \| \quad a.s. \]

The backward iteration of (3.26) to 0 and noting \( x_0 = 0 \) imply

(3.27) \[ \| x_n \| \leq \left\| \sum_{k=1}^{n} a_k \| R_k \| \prod_{l=k+1}^{\infty} (1 + a_k \| A_l(Y_l) \|_N^{1/2})^ {1/2} \quad a.s., \quad n = 1, 2, \ldots \]

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) \[ \| x_n \| \leq Z_0 \cdot \left\| \sum_{k=1}^{n} a_k \| R_k \| \right\| \quad a.s., \quad n = 1, 2, \ldots \]

And from (3.2), we have

(3.29) \[ \| x_n - x_{n-1} \| \leq a_n \| x_{n-1} \| \| A_n(Y_n) \|_N + a_n \| R_n \|, \quad n = 1, 2, \ldots \]

By virtue of \( 0 < \delta \leq 1/7 \), it follows that there exists a number \( \delta \), such that

(3.30) \[ \frac{1-\delta}{2} \leq \delta \leq 3\delta. \]

For the above \( \delta \), let us put \( a_0 = 1 \) and
\[
\begin{align*}
S_n &= 0 \\
S_n &= a_n^{1+\delta} \sum_{k=1}^n (\Lambda_k - A_n(Y_k)).
\end{align*}
\]

And (3.30) and (B2) imply
\[
\sum_{n=1}^{\infty} a_n^{1+\delta} E \| \Lambda_n - A_n(Y_n) \|_V < \infty.
\]

Hence, by applying Lemma 2.5 and its remark, it follows that
\[
\lim_{n \to \infty} \| S_n \|_N = 0 \quad a.s.
\]

Now we put
\[
y_n = a_n^{1-\delta} \sum_{k=1}^n W_{n,k}.
\]

Then we have
\[
y_n = a_n^{1-\delta} \sum_{k=1}^n x_n (a_k^{\delta-1} S_k - a_k^{\delta-1} S_{k-1})
\]
\[
= a_n^{\delta-\delta} x_{n-1} S_n + a_n^{1-\delta} \sum_{k=1}^{n-1} a_k^{\delta-1} (x_{k-1} - x_k) S_k.
\]

Hence we have
\[
\| y_n \| \leq a_n^{\delta-\delta} \| x_{n-1} \| \| S_n \|_N + a_n^{1-\delta} \sum_{k=1}^n a_k^{\delta-1} \| x_{k-1} - x_k \| \| S_k \|_N.
\]

At first we shall prove
\[
\lim_{n \to \infty} a_n^{\delta-\delta} \| x_{n-1} \| \| S_n \|_N = 0 \quad a.s.
\]

By (3.28), we have
\[
\| S_n \|_N \leq Z \| S_n \|_N \sum_{k=1}^n \| a_k \| R_k \| a.s.
\]

And according to (A1), (A2), (B7), (3.20) and the remark (4), it follows that
\[
\sum_{n=1}^{\infty} a_n^{1+\delta} \| R_n \|
\]
\[
\leq \sum_{n=1}^{\infty} a_n^{1+\delta} (\| a_n^{-1} \| \theta_{n-1} - \theta_n \|) + \sum_{n=1}^{\infty} a_n^{2+\delta} \| A_n(Y_n) \|_N (\| a_n^{-1} \| \theta_{n-1} - \theta_n \|)
\]
\[
+ \left( \sum_{n=1}^{\infty} a_n^{1+\delta} \right)^{1/2} \left( \sum_{n=1}^{\infty} a_n^{1+\delta} \| \Gamma_n(Y_n) \|_V \right)^{1/2}
\]
\[
< \infty \quad a.s.
\]

Hence we obtain (3.37) by applying Lemma 2.3.
Next we shall prove

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

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

\[(3.41) \quad a_n^{-\delta} \sum_{k=1}^{n} a_k^{\delta-1} \|x_{k-1} - x_k\| \|S_k\|_N \leq a_n^{-\delta} Z_0 \sum_{k=1}^{n} a_k^{\delta} \|S_k\|_N \|A_k(Y_k)\|_N \sum_{i=1}^{k} a_i \|R_i\| \]

\[+ a_n^{-\delta} \sum_{k=1}^{n} a_k^{\delta} \|S_k\|_N \quad a.s.\]

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

\[(3.42) \quad \|S_n\|_N \leq Z_1 \quad a.s. \quad \text{for } n \geq 1.\]

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

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

And (3.39) and Lemma 2.3 imply

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

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

\[(3.45) \quad \sum_{n=1}^{\infty} \left( a_n^{-\delta} a_n^{\delta} \|A_n(Y_n)\|_N \sum_{k=1}^{n} a_k \|R_k\| \right) \leq \sum_{n=1}^{\infty} \left( a_n^{\delta} \|A_n(Y_n)\|_N \cdot a_n^{\delta} \sum_{k=1}^{n} a_k \|R_k\| \right) \quad a.s.\]

Therefore, by applying Lemma 2.3, we obtain

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

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

Thus the proof of Theorem is completed.

**Remarks.** (1) Suppose that \(\sup_{1 \leq n \leq N} \|E[A_n(Y_n)]\|_N < \infty\) and the assumptions in Theorem are satisfied. Then it holds that

\[(3.47) \quad \lim_{n \to \infty} M_n(X_n) = 0 \quad a.s.\]

(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 forthcoming paper we shall discuss the above case.
4. Applications

1. The learning problem for a pattern classification. Let us consider the problem of estimating an unknown optimal discriminant function ([4], [6], [7]),

\[ D^{(n)}(z) = p_1^{(n)} f_1^{(n)}(z) - p_2^{(n)} f_2^{(n)}(z) \]

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 \( (Z_n, A_n)_{n=1}^{\infty} \) 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 \( A_n \) is a \{1, 2\} -valued random variable whose distribution is equal to \( (p_1^{(n)}, p_2^{(n)}) \).

By many authors ([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 \( (Z_n, A_n) \)'s are dependently distributed. And the learning problem considered here is to estimate the optimal discriminant function (4.1) on the bases of the dependently distributed training sequence \( (Z_n, A_n)_{n=1}^{\infty} \).

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

\[ \int_S \varphi_i(z)\varphi_j(z)dz = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \]

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

\[ \sum_{i=1}^{N} c_i^{(n)} \varphi_i(z) = \langle c^{(n)}, \varphi(z) \rangle \]

which minimizes a quantity \( J_n(c) \) defined by

\[ J_n(c) = \int_S (D^{(n)}(z) - \langle c, \varphi(z) \rangle)^2 dz \]

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

\[ E[e^{(n)} - d(A_n)\varphi(Z_n)] = 0, \quad n=1, 2, \ldots, \]

where

\[ d(\tilde{\sigma}) = \begin{cases} 1 & \text{if } \tilde{\sigma} = 1 \\ -1 & \text{if } \tilde{\sigma} = 2 \end{cases} \]

Hence our problem becomes to find the solution of the equation

\[ E[e - d(A_n)\varphi(Z_n)] = 0 \]

for sufficiently large \( n \). In Section 3, for \( n=1, 2, \ldots \), let us put

\[ x = e \in R^{N}, \quad y = (z, \tilde{\sigma}) \in S \times \{1, 2\} \subset R^{M}, \quad \theta_n = e^{(n)}, \]
An almost sure convergence theorem

\[ (4.8) \quad \Phi_n(x, y) = \Phi_n(c, (z, \delta)) = (c - c^{(n)}) + (e^{(n)} - d(\delta)\varphi(z)), \]

\[ (4.9) \quad A_n(y) = I \]

and

\[ (4.10) \quad \Gamma_n(y) = \Gamma_n((z, \delta)) = c^{(n)} - d(\delta)\varphi(z). \]

And we construct the following procedure,

\[ (4.11) \quad C_0 = \text{an arbitrary constant vector in } \mathbb{R}^N \]

\[ C_{n+1} = C_n - (n+1)^{-1} \{ C_n - d(\mathcal{J}_n+1)\varphi(Z_{n+1}) \}, \]

\[ n = 0, 1, 2, \ldots. \]

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) \( \sup_{1 \leq i \leq N, \ t \geq 0} \mathbb{E}[\varphi_t(Z_t)] < \infty. \)

(ii) \( \lim_{n \to \infty} (n+1) \left\{ \int_0^1 (D^{(n+1)}(z) - D^{(n)}(z))^2 dz \right\}^{1/2} = 0. \)

(iii) \( \lim_{n \to \infty} \rho_n \left( \bigvee_{i=1}^N A_i \right) < 1 \) for some \( n_0, \) where \( A_i = \sigma(Z_t, \mathcal{A}_t) \) \( (i = 1, 2, \ldots). \)

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

Then it follows that

\[ (4.12) \quad \lim_{n \to \infty} \| C_n - c^{(n)} \| = 0 \quad \text{a.s.}. \]

**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

\[ (4.13) \quad \| c^{(n)} - c^{(n-1)} \| = \left[ \sum_{i=1}^N \left\{ \int_0^1 (D^{(n)}(z) - D^{(n+1)}(z)) \varphi_i(z) dz \right\}^2 \right]^{1/2} \]

\[ \leq \left[ \sum_{i=1}^N \left\{ \int_0^1 \varphi_i(z) dz \right\} \left\{ \int_0^1 (D^{(n+1)}(z))^2 dz \right\} \right]^{1/2} \]

\[ = N^{1/2} \left\{ \int_0^1 (D^{(n)}(z) - D^{(n+1)}(z))^2 dz \right\}^{1/2}. \]

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 ([I], [4]). An outcome in this problem is denoted by a pair \( (\alpha, \beta) \). The element \( \alpha \) is an observable input and \( \beta \) is an observable output corresponding to \( \alpha \). Let \( S_0 \subseteq \mathbb{R}^M, S_1 \subseteq \mathbb{R} \)
denote respectively an input space and an output space. For the input \(S_0\)-valued random vectors \(\alpha^1, \alpha^2, \ldots\), we consider a sequence, \((\alpha^1, \beta^1), (\alpha^2, \beta^2), \ldots\). Here we assume that the output \(\beta^n\) observed corresponding to the input \(\alpha^n\) is expressed by

\[(4.14) \quad \beta^n = f(\alpha^n), \quad n = 1, 2, \ldots, \]

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 \(\alpha^n\)'s are dependently distributed.

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

\[(4.15) \quad \sum_{i=1}^N c_i^{(n)} \phi_i(\alpha) = (c, \phi(\alpha)) \]

which minimizes the quantity \(J_n(c)\) defined by

\[(4.16) \quad J_n(c) = E\left[\beta^n - <c, \phi(\alpha^n)>\right]^2 \]

where \(c^{(n)} = (c_1^{(n)}, \ldots, c_N^{(n)})_n\), \(c = (c_1, \ldots, c_N)\) and \(\phi(\alpha) = (\phi_1(\alpha), \ldots, \phi_N(\alpha))\). By differentiating (4.16) with respect to \(c_i\) \((i = 1, \ldots, N)\), equating the derivative to zero, we obtain

\[(4.17) \quad E[\phi(\alpha^n)\phi(\alpha^n)^T - \beta^n \phi(\alpha^n)] = 0, \quad n = 1, 2, \ldots, \]

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 us suppose that \(E[\phi(\alpha^n)\phi(\alpha^n)]\)'s are positive definite matrices. Then the equation (4.17) has the unique solution \(c^{(n)}\) which minimizes the quantity \(J_n(c)\) for \(n = 1, 2, \ldots\). And in Section 3, let us put

\[(4.18) \quad x = c \in R^N, \quad y = \alpha \in S_0 \subseteq R^m, \quad \theta_n = c^{(n)}, \]

\[(4.19) \quad A_n(\alpha) = \phi(\alpha)^T\phi(\alpha) \]

and

\[(4.20) \quad \Gamma_n(\alpha) = c^{(n)}\phi(\alpha)^T\phi(\alpha) - f(\alpha)\phi(\alpha). \]

And we construct the following procedure,

\[(4.21) \quad C_0 = \text{an arbitrary constant vector in } R^N, \]

\[C_{n+1} = C_n - (n + 1)^{-1} \left[C_n\phi(\alpha^n)^T\phi(\alpha^n) - \beta^n \phi(\alpha^n)\right], \quad n = 0, 1, 2, \ldots. \]

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

**Theorem 4.2.** Suppose that the distributions of \(\alpha^n\)'s do not depend on \(n\). And suppose that \(E[\phi(\alpha^n)^T\phi(\alpha^n)]\) is a positive definite matrix and the following conditions are satisfied.

1. \(E[f(\alpha^n)\phi_i(\alpha^n)]^2 < \infty \quad \text{for } i = 1, 2, \ldots, N.\)
An almost sure convergence theorem

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

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

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

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

\[
\lim_{n \to \infty} \| \mathbf{C}_n - \mathbf{c}^* \| = 0 \quad \text{a.s..}
\]

PROOF. By the remark (3) of Assumptions, (A1) to (A3) follows. Noting \( a_n = n \), (ii) implies (B2). Note that

\[
\langle \mathbf{c} \varphi(\alpha^T \varphi(\alpha^1)), \mathbf{c} \rangle = \sum_{i=1}^N \left( \sum_{j=1}^N \varphi_i(\alpha^1)\varphi_j(\alpha^1)c_j \right)c_i = \left( \sum_{i=1}^N c_i \varphi_i(\alpha^1) \right)^2 \quad \text{for all } \mathbf{c} = (c_1, \ldots, c_N) \in \mathbb{R}^N.
\]

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

\[
\langle \mathbf{c} E[\varphi(\alpha^1)^T \varphi(\alpha^1)], \mathbf{c} \rangle \geq \alpha_0 \| \mathbf{c} \|^2
\]

for all \( \mathbf{c} \in \mathbb{R}^N \), where \( \alpha_0 \) is the minimum eigenvalue of \( E[\varphi(\alpha^1)^T \varphi(\alpha^1)] \). Hence, (B4) holds. Since \( E[\varphi(\alpha^1)^T \varphi(\alpha^1)] \) is a positive definite matrix, there exists the inverse matrix. Hence, the equation (4.17) has the unique solution \( \mathbf{c}^* \), that is

\[
\mathbf{c}^* = E[\varphi(\alpha^1)^T \varphi(\alpha^1)]^{-1} E[\varphi(\alpha^1)^T \varphi(\alpha^1)]^{-1}.
\]

Therefore, (i) and (ii) imply (B6). And noting that the distributions of \( \alpha_n^1 \)'s do not depend on \( n \), (4.25) implies (B7).

Thus all the assumptions in Section 3 are satisfied.

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

References


