ON A PATTERN CLASSIFICATION PROBLEM ON THE BASIS OF A TRAINING SEQUENCE ASSOCIATED WITH DEPENDENT RANDOM VARIABLES

Watanabe, Masafumi
Department of Mathematics, Kyushu University

渡辺, 正文
九州大学数学科

http://hdl.handle.net/2324/13081

出版情報：統計数理研究. 16 (1/2), pp.35-48, 1974-03. 統計科学研究会
バージョン: published
権利関係：
ON A PATTERN CLASSIFICATION PROBLEM ON THE
BASIS OF A TRAINING SEQUENCE ASSOCIATED
WITH DEPENDENT RANDOM VARIABLES

By

Masafumi WATANABE*

(Received July 10, 1973)

§ 1. Introduction and Summary.

In this paper we shall be concerned with the pattern classification problem related
to "learning with a teacher". Many previous authors have studied this problem
under the given situation of a training sequence composed of observed patterns
independently sampled from a common population. From the practical point of view,
however, the situation that observed patterns are independently sampled is rather
restrictive. For this reason, in this paper the author treats the pattern classification
problem on the basis of a dependent sequence of observed patterns. In [8], [9] and
[10], K. Tanaka treated same problem as the present author, but restricted himself
to the parametric case. In this paper, we shall consider the non-parametric case,
and appeal to the method which has been developed in [12]. Consequently, our
various conditions imposed are different from those in [8], [9] and [10].

This paper consists of five sections. In Section 2, we shall give the formulation
of our problem and five assumptions necessary for subsequent arguments. In Section
3, we shall define a recursive algorithm for the pattern classification problem, which
is an application of the dynamic stochastic approximation method [3], and investigate
the convergence of it. The meaning of the convergence is "in the mean". In
Section 4 we shall give two examples.

§ 2. The formulation of the problem and Assumptions.

We consider the two-categories classification problem. Let \( \mathcal{X}^{(n)} \) and \( \Theta \) denote a
pattern space at instant \( n \) and the set of categories, respectively. We assume that
\( \Theta \) consists of two categories \( \theta_1 \) and \( \theta_2 \), i.e. \( \Theta = \{ \theta_1, \theta_2 \} \).

An outcome in pattern classification problem is described by a pair \( (x, \theta) \), where
\( x \) is an observed pattern in pattern space and \( \theta \) specifies the category of the observed
pattern \( x \). Generally, \( \theta \) is unknown to the observer. For a sequence of observed
patterns \( x_1, x_2, \ldots, x_n, \ldots \), we can consider a sequence:

\[
(2.1) \quad (x_1, \theta_1), (x_2, \theta_2), \ldots, (x_n, \theta_n), \ldots
\]

* Department of Mathematics, Kyushu University, Fukuoka.
with $x_n \in \hat{X}^{(n)}$ and $\theta_n \in \Theta$, where $\theta_n = \theta_i$ if $x_n$ is a sample value from a specific category $\theta_i$.

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables, where $X_n$ is a $\hat{X}^{(n)}$-valued random variable and can be from either of two classes $\theta_1$ and $\theta_2$. And let $\Theta_n$ be a $\Theta$-valued random variable for each $n$, where $\Theta_n = \theta_i$ if $X_n$ is from the class $\theta_i$. Thus $(x_1, \theta_1), \ldots, (x_n, \theta_n), \ldots$ are outcomes of $(X_1, \Theta_1), (X_2, \Theta_2) \ldots (X_n, \Theta_n), \ldots$.

In this paper, we shall assume $\hat{X}^{(n)} = R^N$ for all $n$, where $R^N$ is $N$-dimensional Euclidean space. And we shall assume for each $n$, random variable $X_n$ has a probability density function $p(n)(\cdot)$ with respect to $N$-dimensional Lebesgue measure if $X_n$ comes from the class $\theta_i$. Furthermore, for each $n$, suppose that there exists a distribution $(q^{(n)}_1, q^{(n)}_2)$ on $\Theta$, that is, $q^{(n)}_i = \Pr(\Theta_n = \theta_i)$ and $q^{(n)}_j = \Pr(\Theta_n = \theta_j)$. Therefore, for each $n$, a pair of random variables $(X_n, \Theta_n)$ has a probability density function $p(n)(x, \theta), x \in R^N$ and $\theta \in \Theta$, where

$$p(n)(x, \theta_i) = q^{(n)}_i \cdot p(n)(x); \quad x \in R^N \quad \text{and} \quad i = 1, 2, \ldots.$$  

Let us assume tentatively that $p(n)(\cdot, \theta_1)$ and $p(n)(\cdot, \theta_2)$ are all known to us at instant $n$. Let us consider the discriminant function at instant $n$:

$$D(n)(x) = p(n)(x, \theta_1) - p(n)(x, \theta_2), \quad x \in R^N.$$  

And let us consider the following decision rule based on (2.3); decide $X_n$ to be from the class $\theta_1$ if $D(n)(x) \geq 0$ for the outcome $x$ of $X_n$, and decide $X_n$ to be from the class $\theta_2$ if $D(n)(x) < 0$ for the outcome $x$ of $X_n$. It is well known that this decision rule minimizes the probability of misclassification at instant $n$. This decision rule has been called the Bayes decision rule. In this paper, we shall call (2.3) the optimal discriminant function at instant $n$.

In this paper, we shall treat the case when $p(n)(\cdot, \theta_1)$ and $p(n)(\cdot, \theta_2)$ are unknown to us for each $n$, consequently the optimal discriminant function at instant $n$ is unknown to us. In this situation we are supposed to have a training sequence $\{(X_n, \Theta_n)\}_{n=1}^\infty$ with the observed $R^N$-valued random vector $X_n$ and $\Theta$-valued random variable $\Theta_n$. It is assumed that the knowledge of the $\Theta_n$ is obtained from a “teacher” who classifies (without error) the incoming $X_n$ for each $n$.

The pattern classification problem considered here is to find a decision rule to classify the pattern $X_n$ as correctly as possible into either of the two categories for a sufficiently large $n$ on the basis of a training sequence $\{(X_n, \Theta_n)\}_{n=1}^\infty$. It is reasonable, therefore, to consider a method of approximation to the limit of $D(n)(\cdot)$, if it exist, by using a training sequence.

For each $n$, we assume that $(X_n, \Theta_n)$ has a probability density function $p(n)(x, \theta), x \in R^N$ and $\theta \in \Theta$, which is defined in (2.2). Furthermore, for each $n$ and $m$, $(X_{n+m}, \Theta_{n+m})$ has a conditional probability density function given the first $m$ history $(X_1, \Theta_1) = (x_1, \theta_1), (X_2, \Theta_2) = (x_2, \theta_2), \ldots, (X_m, \Theta_m) = (x_m, \theta_m)$. We denote this transition probability density function by

$$p(n+m)(x, \theta | (x_1, \theta_1), \ldots, (x_m, \theta_m)).$$  

Next, we shall give assumptions.
ASSUMPTION 1. There exists a positive constant $P_0$ such that
\begin{equation}
\sup_{(x, \theta, n)} p^{(n)}(x, \theta) \leq P_0.
\end{equation}

ASSUMPTION 2. We put for $n, m = 1, 2, \ldots$,
\begin{equation}
H_{n, m} = \sup_{(x, \theta)} |p^{(n)}(x, \theta) - p^{(m)}(x, \theta)| .
\end{equation}
Then
\begin{equation}
\lim_{n, m \to \infty} H_{n, m} = 0 .
\end{equation}

ASSUMPTION 3. We put for $n = 0, 1, 2, \ldots$ and $m = 1, 2, \ldots$,
\begin{equation}
G_{n, m} = \sup_{(x, \theta, (x_1, \theta_1), \ldots, (x_m, \theta_m))} |p^{(n+m)}(x, \theta) - p^{(n)}(x, \theta)| .
\end{equation}
Then
\begin{equation}
\lim_{n, m \to \infty} G_{n, m} = 0 .
\end{equation}

ASSUMPTION 4. For each $n$, $p^{(n)}(\cdot, \theta_1)$ and $p^{(n)}(\cdot, \theta_2)$ are uniformly continuous functions on $R^N$.

ASSUMPTION 5. For all $n$, $p^{(n)}(\cdot, \theta_1)$ and $p^{(n)}(\cdot, \theta_2)$ satisfy an uniform Lipschitz conditions with constants $C_1$ and $C_2$, respectively : the following inequality holds for all $n$,
\begin{equation}
|p^{(n)}(x, \theta_1) - p^{(n)}(y, \theta_1)| \leq C_i \|x-y\| \quad \text{for all } n ,
\end{equation}
where $\|y\| = (\sum_{i=1}^{N} y_i^2)^{\frac{1}{2}}$ for $y = (y_1, \ldots, y_N) \in R^N$.

REMARK. For each $n$ and $m$, putting
\begin{equation}
H'_{n, m} = \sup_x |D^{(n)}(x) - D^{(m)}(x)| ,
\end{equation}
\begin{equation}
D^{(n+m)}(x| x_1, \theta_1), \ldots, (x_m, \theta_m)) = p^{(n+m)}(x, \theta_1| (x_1, \theta_1), \ldots, (x_m, \theta_m)) - p^{(n+m)}(x, \theta_2| (x_1, \theta_1), \ldots, (x_m, \theta_m)) ,
\end{equation}
and
\begin{equation}
G'_{n, m} = \sup_x |D^{(n+m)}(x| (x_1, \theta_1), \ldots, (x_m, \theta_m)) - D^{(n+m)}(x)| ,
\end{equation}
we have
\begin{equation}
H'_{n, m} \leq 2H_{n, m}
\end{equation}
and
\begin{equation}
G'_{n, m} \leq 2G_{n, m} .
\end{equation}
Then it is easily seen that
\begin{equation}
\sup_{x, n} |D^{(n)}(x)| \leq 2P_0 ,
\end{equation}
\begin{equation}
\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} H_{j, n} = 0 ,
\end{equation}
and
under Assumptions 1, 2 and 3, respectively.

And if Assumption 5 holds, then we have

\[ |D^{(n)}(x) - D^{(n)}(y)| \leq C_0 \|x - y\|, \quad x, y \in \mathbb{R}^N \]

and for all \( n \), where \( C_0 = \max \{C_1, C_2\} \).

§ 3. Main results.

In this paper, all integrals are interpreted as \( N \)-dimensional Lebesgue integral on \( \mathbb{R}^N \). We denote \( \int_{\mathbb{R}^N} g(x) d\mu(x) \) simply by \( \int g(x) dx \), where \( \mu \) is \( N \)-dimensional Lebesgue measure. And Fubini's theorem is invoked without any comments.

Let \( K(\cdot) \) be a real-valued measurable function on \( \mathbb{R}^N \) satisfying the following conditions:

1. \( \forall y \in \mathbb{R}^N, \quad K(y) \geq 0 \)
2. \( \sup_{y \in \mathbb{R}^N} K(y) = K_0 < \infty \)
3. \( \int K(y) dy = 1 \)
4. \( \sup_{y \in \mathbb{R}^N} \|y\|^{1+\alpha} K(y) = K_1 < \infty \) for some \( \alpha > 0 \)
5. \( \int \|y\| \cdot K(y) dy = K_2 < \infty \)

Let \( \{h_n\}_{n=1}^{\infty} \) be a sequence of positive numbers satisfying the following conditions:

1. \( 1 \geq h_1 \geq h_2 \geq \cdots \geq h_n \geq \cdots \)
2. \( \lim_{n \to \infty} h_n = 0 \)
3. \( \lim_{n \to \infty} nh_n = \infty \)

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

\[ K_n(x, y) = h_n^{-N} \cdot K[h_n^{-1}(x-y)] \quad x, y \in \mathbb{R}^N \]

The following lemma is a modification of Theorem 1A in [5].

**Lemma 1.** Let Assumption 1, Assumption 2 and Assumption 4 hold. Let \( K(\cdot) \) be a real-valued measurable function on \( \mathbb{R}^N \) satisfying (K1), (K2), (K3) and (K4), and let \( \{h_n\}_{n=1}^{\infty} \) be a sequence of positive numbers satisfying (S1) and (S2). Define \( g_n(\cdot) \) by

\[ g_n(x) = \int K_n(x, y) D^{(n)}(y) dy \quad \text{for } n = 1, 2, \ldots \]

where \( K_n(x, y) \) is defined by (3.1). Then it holds that

\[ \sup |g_n(x) - D^{(n)}(x)| \to 0 \quad \text{as } n \to \infty. \]
PROOF. Note first that
(3.4) \[ g_n(x) - D^{(n)}(x) = \int \{D^{(n)}(x - y) - D^{(n)}(x)\} \cdot \frac{1}{h_n} \cdot K\left(\frac{y}{h_n}\right) dy. \]

Let choose \( \alpha' \) satisfying
(3.5) \[ \alpha' \cdot (1 + \alpha)^{-1} > \alpha' > 0, \]
and split the region of integration into two regions, \( \|y\| \leq \delta_n \) and \( \|y\| > \delta_n \), where \( \delta_n = h_n^\alpha' \). Then
(3.6) \[ |g_n(x) - D^{(n)}(x)| \leq \sup_{\|y\| \leq \delta_n} |D^{(n)}(x - y) - D^{(n)}(x)| \int K(z) dz + \frac{1}{\delta_n} \sup_{\|y\| \leq \delta_n/h_n} \|z\| \cdot (K(z) \cdot \int |D^{(n)}(y)| dy + |D^{(n)}(x)| \cdot \int K(y) dy). \]

From (K4), we have
(3.7) \[ \sup_{\|z\| \leq \delta_n/h_n} \|z\| \cdot K(z) \leq K_1 \cdot h_n^\alpha \delta_n^\alpha, \]
and from the definition of the optimal discriminant function
(3.8) \[ \int |D^{(n)}(y)| dy \leq q^{(1)}_n + q^{(2)}_n = 1 \quad \text{for all } n. \]

Therefore we have, from (3.6), (3.7), (3.8) and Assumption 1,
(3.9) \[ \sup_x |g_n(x) - D^{(n)}(x)| \leq \sup_{\|y\| \leq \delta_n} |D^{(n)}(x) - D^{(n)}(y)| + K_1 \cdot h_n^\alpha \delta_n^{-(1 + \alpha)} + 2P \int \|y\| \leq \delta_n/h_n K(y) dy. \]

Noting that
(3.10) \[ \sup_{\|y\| \leq \delta_n} |D^{(n)}(x) - D^{(m)}(x)| + \sup_{\|y\| \leq \delta_n} |D^{(m)}(x) - D^{(m)}(y)| + \sup_{\|y\| \leq \delta_n} |D^{(m)}(y) - D^{(n)}(y)| \leq 2H_{n,m} + \sup_{\|y\| \leq \delta_n} |D^{(m)}(x) - D^{(m)}(y)| \]
where \( m \) is an arbitrary integer. From Assumption 2, we can choose \( \varepsilon > 0 \) and let \( N_0 \) be such that \( n, m \geq N_0 \) implies \( 2H_{n,m} < \varepsilon \). Let \( m_o \) be some integer satisfying \( m_o \geq N_0 \), then we have
(3.11) \[ \sup_x |g_n(x) - D^{(n)}(x)| \leq \varepsilon + \sup_{\|y\| \leq \delta_n} |D^{(m_o)}(x) - D^{(m_o)}(y)| + K_1 \cdot h_n^\alpha \delta_n^{-(1 + \alpha)} + 2P \int \|y\| \leq \delta_n/h_n K(y) dy \quad \text{for all } n \geq N_0. \]
By \((3.5)\), it is easily seen that
\[
\begin{align*}
\partial_n &= h_n' \rightarrow 0 \quad \text{as } n \rightarrow \infty , \\
h_n^{\sigma} / \partial_n^{1+\alpha} &= h_n^{\sigma - \alpha'(1+\alpha)} \rightarrow 0 \quad \text{as } n \rightarrow \infty , \\
\end{align*}
\]
and
\[
\begin{align*}
\partial_n / h_n &= h_n^{-(1-\alpha')} \rightarrow \infty \quad \text{as } n \rightarrow \infty . \\
\end{align*}
\]
Therefore the right side in \((3.11)\) tends to 0 by letting \(n \rightarrow \infty\) at first and then \(\varepsilon \rightarrow 0\). Thus the proof of the lemma is completed.

The following lemma was proved in [12] (Lemma 4).

**LEMMA 2.** Let Assumption 5 holds. Let \(K(\cdot)\) be a real-valued measurable function on \(R^N\) satisfying \((K1), (K3)\) and \((K5)\), and let \(\{h_n\}_{n=1}^\infty\) be a sequence of positive numbers. Then it hold that
\[
\sup_{x \in R^N} |g_n(x) - D^{\sigma}(x)| \leq C_0 \cdot K \cdot h_n \quad \text{for } n = 1, 2, \ldots
\]
where \(g_n(\cdot)\) was defined in Lemma 1.

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
\[
\begin{align*}
\rho^n &= \rho^n(\Theta_n) = 1 \quad \text{if } \Theta_n = \theta_1 , \\
&= -1 \quad \text{if } \Theta_n = \theta_2 . \\
\end{align*}
\]
for \(n = 1, 2, \ldots\). From \((3.1)\) and \((3.16)\), we have
\[
\begin{align*}
\mathbb{E}[\rho^n \cdot K_n(x, X_n)] &= \int K_n(x, y) D^{\sigma}(y) dy \quad \text{for } n = 1, 2, \ldots.
\end{align*}
\]

In view of the above arguments, we now construct the following algorithm which is an application of the dynamic stochastic approximation method [3].

**LEARNING ALGORITHM.** Let us define the following recursive algorithm:
\[
\begin{align*}
D_0(x) &= 0 , \quad x \in R^N , \\
D_n(x) &= D_n(x) + (n+1)^{-1} \cdot (\rho_{n+1} \cdot K_{n+1}(x, X_{n+1}) - D_n(x)) , \\
D_{n+1}(x) &= n^{-1} \sum_{j=1}^n \rho^j \cdot K_j(x, X_j) , \quad x \in R^N \text{ and } n = 0, 1, 2, \ldots. \\
\end{align*}
\]

The above algorithm is equivalent to the following one;
\[
\begin{align*}
D_n(x) &= n^{-1} \sum_{j=1}^n \rho^j \cdot K_j(x, X_j) , \quad x \in R^N \text{ and } n = 1, 2, \ldots.
\end{align*}
\]

From the above algorithm we can obtain the following lemmas.

**LEMMA 3.** Let Assumptions 1, 2 and 4 hold. Let \(K(\cdot)\) be a real-valued measurable function on \(R^N\) satisfying \((K1), (K2), (K3)\) and \((K4)\), and let \(\{h_n\}_{n=1}^\infty\) be a positive sequence satisfying \((S1)\) and \((S2)\). Then, it holds that
\[
\begin{align*}
\lim_{n \rightarrow \infty} \mathbb{E}[D_n(x)] - D^{\sigma}(x) dx &= 0 .
\end{align*}
\]

**PROOF.** From \((3.20)\) and \((3.17)\), we have
(3.22) \[ |E[D_n(x)] - D^{(n)}(x)| \leq n^{-1} \sum_{j=1}^{n} |g_j(x) - D^{(j)}(x)| \]
\[ + n^{-1} \sum_{j=1}^{n} |D^{(j)}(x) - D^{(n)}(x)| , \]

where \( g_j(\cdot) \) was defined in (3.2). From Lemma 1, we have

(3.23) \[ \sup_x |g_j(x) - D^{(j)}(x)| \longrightarrow 0 \quad \text{as} \quad j \to \infty . \]

And by Assumptions 1 and 2, we have

(3.24) \[ \sup_n n^{-1} \sum_{j=1}^{n} |D^{(j)}(x) - D^{(n)}(x)| \leq n^{-1} \sum_{j=1}^{n} H_{j,n} \longrightarrow 0 \quad \text{as} \quad n \to \infty . \]

Note that

(3.25) \[ (E[D_n(x)] - D^{(n)}(x))^2 \text{dx} \]
\[ \leq \sup_x |E[D_n(x)] - D^{(n)}(x)| \cdot \int \{|E[D_n(x)]| + |D^{(n)}(x)|\} \text{dx} \]
\[ \leq \sup_x |E[D_n(x)] - D^{(n)}(x)| \quad \text{for all} \quad n . \]

Therefore, by (3.23), (3.24) and (3.25), we have (3.21). Thus the proof of the lemma is completed.

**LEMMA 4.** Let Assumption 3 hold. Let \( K(\cdot) \) be a real-valued measurable function on \( \mathbb{R}^N \) satisfying (K1) and (K3), and let \( \{h_n\}_{n=1}^{\infty} \) be a positive sequence. Then, it holds that

(3.26) \[ \int |E[\rho^m \cdot K_m(x, X_n) - \rho^m \cdot K_{n+m}(x, X_{n+m})] \]
\[ - E[\rho^m \cdot K_m(x, X_n)] \cdot E[\rho^m \cdot K_{n+m}(x, X_{n+m})]| \text{dx} \]
\[ \leq 2G_{n,m} \quad \text{for} \quad n = 0, 1, 2, \ldots \quad \text{and} \quad m = 1, 2, \ldots . \]

**PROOF.** We put \( U_n = \rho^n \cdot K_n(x, X_n) \) for \( n = 1, 2, \ldots \), and \( Y_m = ((X_1, \Theta_1), \ldots, (X_m, \Theta_m)) \) for \( m = 1, 2, \ldots \). Then we have

(3.27) \[ |E[U_m \cdot U_{n+m}] - E[U_m] \cdot E[U_{n+m}]| \]
\[ = |E[U_m \cdot E[U_{n+m} | Y_m] - E[U_{n+m}]| | \]
\[ \leq E[|U_m| \cdot \int K_{n+m}(x, y) \cdot |D^{(n+m)}(y | Y_m) - D^{(n+m)}(y)| dy] . \]

From Assumption 3 and (3.27)

(3.28) \[ \int |E[U_m \cdot U_{n+m}] - E[U_m] \cdot E[U_{n+m}]| \text{dx} \leq 2G_{n,m} \cdot E\left[ \int K_n(x, X_n) \text{dx} \right] = 2G_{n,m} . \]

Thus the proof of the lemma is completed.

From Lemma 4 we have the following lemma.

**LEMMA 5.** Let the conditions in Lemma 4 be satisfied. Moreover, suppose that \( K(\cdot) \) satisfies (K2), and \( \{h_n\}_{n=1}^{\infty} \) satisfies (S1) and (S3). Then it holds that

(3.29) \[ \lim_{n \to \infty} E[D_n(x) - E[D_n(x)]]^2 \text{dx} = 0 . \]
PROOF. At first note that

\begin{equation}
\mathbb{E}[D_n(x) - \mathbb{E}[D_n(x)]]^2
= \mathbb{E}\left[ \sum_{j=1}^{n} (U_j - \mathbb{E}[U_j]) \right]^2
= n^{-2} \sum_{j=1}^{n} \mathbb{E}[U_j - \mathbb{E}[U_j]]^2 + n^{-2} \sum_{j \neq j'} \{\mathbb{E}[U_j \cdot U_{j'}] - \mathbb{E}[U_j] \cdot \mathbb{E}[U_{j'}]\}.
\end{equation}

From Lemma 4 we have

\begin{equation}
n^{-2} \sum_{j \neq j'} \{\mathbb{E}[U_j \cdot U_{j'}] - \mathbb{E}[U_j] \cdot \mathbb{E}[U_{j'}]\} \geq 2 \cdot n^{-2} \sum_{i=1}^{n} \sum_{j=-1}^{n-2} \int |\mathbb{E}[U_{i+j} \cdot U_j] - \mathbb{E}[U_{i+j}] \cdot \mathbb{E}[U_j]| \, dx
\end{equation}

\begin{equation}
\leq 4n^{-2} \sum_{i=1}^{n} \sum_{j=0}^{n} G_{i,j} \quad \text{for all } n.
\end{equation}

Next we have

\begin{equation}
n^{-2} \sum_{j=1}^{n} \mathbb{E}[U_j - \mathbb{E}[U_j]]^2 \geq n^{-2} \sum_{i=1}^{n} \mathbb{E}[K_i(x, X_i)]^2 \geq n^{-2} \sum_{j=1}^{n} \mathbb{E}[K_j(x, X_j)]^2 \quad \text{for all } n,
\end{equation}

where \( p^{(j)}(y) = p^{(j)}(y, \theta_j) + p^{(j)}(y, \theta_{j'}), y \in \mathbb{R}^N \) and \( j = 1, 2, \ldots \). By (K2) and (S1), we have

\begin{equation}
n^{-2} \sum_{j=1}^{n} \mathbb{E}[U_j - \mathbb{E}[U_j]]^2 \leq K_1 (n \cdot h_n)^{-1} \quad \text{for } n = 1, 2, \ldots.
\end{equation}

Therefore by (3.31), (3.33), Assumption 3 and (S3), we have

\begin{equation}
\int \mathbb{E}[D_n(x) - \mathbb{E}D_n(x)]^2 \, dx \longrightarrow 0 \quad \text{as } n \to \infty.
\end{equation}

Thus the proof of the lemma is completed.

From Lemma 3 and Lemma 5, we have the following theorem.

THEOREM 1. Let Assumption 1, 2, 3 and 4 hold. Let \( K(\cdot) \) be a real-valued measurable function on \( \mathbb{R}^N \) satisfying (K1), (K2), (K3) and (K4), and let \( \{h_n\}_{n=1}^{\infty} \) be a sequence of positive numbers satisfying (S1), (S2) and (S3). Then it holds that

\begin{equation}
\lim_{n \to \infty} \mathbb{E}\left[ \left( D_n(x) - D^{(n)}(x) \right)^2 \right] = 0.
\end{equation}

PROOF. Noting that

\begin{equation}
\mathbb{E}\left[ \left( D_n(x) - D^{(n)}(x) \right)^2 \right] = \int \mathbb{E}[D_n(x) - \mathbb{E}D_n(x)]^2 \, dx + \int (\mathbb{E}[D_n(x)] - D^{(n)}(x))^2 \, dx,
\end{equation}

it is easily seen that (3.34) is direct from Lemmas 3 and 5.
Next, we shall give a theorem concerning the order of the convergence. The following lemma was given by the present author in [11].

**Lemma 6.** Let \( \{a_n\}_{n=1}^{\infty} \) be a sequence of non-negative numbers. Suppose that there exist a position integer \( n_0 \), two sequences of positive numbers \( \{a_n\}_{n=1}^{\infty} \) and \( \{b_n\}_{n=1}^{\infty} \), and two positive constants \( C \) and \( \alpha_0 \) such that

\[
A_{n+1} \leq (1-a_{n+1})A_n + C \cdot a_{n+1} \cdot b_{n+1} \quad \text{for all } n \geq n_0,
\]

\[
\sum_{n=1}^{\infty} a_n = \infty \quad \text{and} \quad \lim_{n \to \infty} a_n = 0,
\]

\[
0 < \alpha_0 < 1,
\]

and

\[
(1-a_{n+1}) \cdot b_n / b_{n+1} \leq (1-\alpha_0 a_{n+1}) \quad \text{for all } n \geq n_0.
\]

Then, there exists a positive constant \( C_0 \) such that

\[
A_n \leq C_0 \cdot b_n \quad \text{for all } n.
\]

**Remark.** In Lemma 6, when \( a_n = n^{-1} \) and \( b_n = n^{-\beta} \) \((0 < \beta < 1)\) the conditions (3.37) and (3.39) are satisfied by \( \alpha_0 \) such that \( 0 < \alpha_0 + \beta < 1 \). And when \( A_n = n^{-1} \sum_{j=1}^{n} b_j \), the condition (3.36) is satisfied by \( C = 1 \). Therefore, there exists a position constant \( C_0 \) such that

\[
n^{-1} \sum_{j=1}^{n} b_j \leq C_0 \cdot b_n \quad \text{for all } n.
\]

**Theorem 2.** Let Assumption 1, 2, 3 and 5 hold. Let \( K(\cdot) \), be a real-valued measurable function satisfying (K1), (K2), (K3) and (K5), and let \( \{h_n\}_{n=1}^{\infty} \) be a sequence of positive numbers satisfying (S1), (S2) and (S3). Moreover, let the following conditions be satisfied: there exist positive constant \( H, G \) and \( \alpha_0 \), and positive integer \( n_0 \) such that

\[
n \cdot H_{n-1} \leq H \cdot b_n \quad \text{for } n = 1, 2, \ldots,
\]

\[
\sup_m G_{n,m} \leq G \cdot b_n \quad \text{for } n = 1, 2, \ldots,
\]

\[
0 < \alpha_0 < 1,
\]

and

\[
b_n / b_{n+1} \leq 1 + \alpha_0 n^{-1} \quad \text{for all } n \geq n_0,
\]

where

\[
b_n = \max \{h_n, (n \cdot h_n)^{-1}\} \quad \text{for } n = 1, 2, \ldots.
\]

Then, there exists a positive constant \( C^* \) such that

\[
E[\int (D_n(x) - D^{(\alpha)}(x))^2 dx] \leq C^* \cdot b_n \quad \text{for all } n.
\]

**Proof.** From (3.22) and (3.25) in the proof of Lemma 3, and from Lemma 2 we have
From (3.31) and (3.33) in the proof of Lemma 5 we have

\[ (3.48) \int E(D_n(x)-D^{(n)}(x))^2dx \leq 4n^{-2}\sum_{i=1}^{n} G_{i,j} + K_1(nh_n^n)^{-1} \] for \( n = 1, 2, \ldots \).  

Therefore,

\[ (3.49) E\left[ \int (D_n(x)-D^{(n)}(x))^2dx \right] \leq C_0 \cdot K_2 \cdot \sum_{i=1}^{n} h_i + 2n^{-1} \sum_{j=1}^{n} H_{j,n} + 4n^{-2}\sum_{i=1}^{n} G_{i,j} + K_1(nh_n^n)^{-1} \] for \( n = 1, 2, \ldots \).  

Noting that

\[ (3.50) \sum_{j=1}^{n} b_j \leq 2(H_{j,j+1}+H_{j+1,j+2}+\cdots+H_{n-1,n}) \]

by (3.41) we have

\[ (3.51) n^{-1} \sum_{j=1}^{n} H_{j,n} \leq 2H \cdot n^{-1}\sum_{j=1}^{n} \left( \frac{b_{j+1}}{j+1} + \cdots + \frac{b_{n}}{n} \right) \]

\[ \leq 2H \cdot n^{-1}\sum_{j=1}^{n} b_j \] for \( n = 1, 2, \ldots \).  

By (3.42)

\[ (3.52) n^{-2}\sum_{j=1}^{n} G_{i,j} \leq G \cdot n^{-1}\sum_{j=1}^{n} b_j \] for \( n = 1, 2, \ldots \).  

There by (3.45), (3.49), (3.51) and (3.52), we have

\[ (3.53) E\left[ \int (D_n(x)-D^{(n)}(x))^2dx \right] \leq \tilde{C} \cdot \left( n^{-1}\sum_{j=1}^{n} b_j + b_n \right) \] for \( n = 1, 2, \ldots \),

where \( \tilde{C} = \max \{C_0, K_2, 4G, K_1 \} \).  And by (3.44) we have

\[ (1-(n+1)^{-1})b_{n}/b_{n+1} \leq (1-\alpha_0'(n+1)^{-1}) \] for all \( n \geq n_0 \),

where \( \alpha_0' = 1-\alpha_0 \). Therefore from Lemma 6 and its remark, there exists a positive constant \( C' \) such that

\[ (3.54) n^{-1}\sum_{j=1}^{n} b_j \leq C' \cdot b_n \] for all \( n \).

By (3.53) and (3.54), we have

\[ (3.55) E\left[ \int (D_n(x)-D^{(n)}(x))^2dx \right] \leq C^* \cdot b_n \] for all \( n \),

where \( C^* = \max \{C_0, K_1, C', 4H \cdot C', K_1 \} \). Thus the proof of the theorem is completed.

**REMARK.** When \( h_n = n^{-\beta} \) (\( 0 < \beta < N^{-1} \)), we can remove the conditions (3.43) and (3.44). In this case, we have

\[ E\left[ \int (D_n(x)-D^{(n)}(x))^2dx \right] \leq C^* \cdot n^{-\beta_0} \] where \( \beta_0 = \min \{\beta, 1-N\beta\} \).
Let $P_{\tilde{H}}(\varepsilon)$ be the probability of misclassification by using a discriminant function $d(\cdot)$ at instant $n$, and let $P(n)(\varepsilon)$ be the probability of misclassification by using the optimal discriminant function $D^{(n)}(\cdot)$ at instant $n$.

We put $\hat{P}^{(n)}(x) = \hat{P}^{(n)}(x, \theta_1) + \hat{P}^{(n)}(x, \theta_2)$, $x \in \mathbb{R}^N$. Choose $\varepsilon > 0$. Let $\{S_{\varepsilon}^{(n)}\}_{n=1}^{\infty}$ be a sequence of bounded sets in $\mathbb{R}^N$ such that for each $n$, the following inequality is satisfied

\begin{equation}
\int_{S_{\varepsilon}^{(n)}} \hat{P}^{(n)}(x)dx \geq 1 - \varepsilon .
\end{equation}

**Theorem 3.** Let the hypotheses in Theorem 1 hold. In addition, let $\varepsilon > 0$ be given then there exists a positive constant $M_\varepsilon$ such that

\begin{equation}
\mu(S_{\varepsilon}^{(n)}) \leq M_\varepsilon \quad \text{for all } n ,
\end{equation}

where $\mu$ is $N$-dimensional Lebesgue measure on $\mathbb{R}^N$ and $\{S_{\varepsilon}^{(n)}\}_{n=1}^{\infty}$ is sequence of bounded sets in $\mathbb{R}^N$ such that for each $n$, (3.56) is satisfied. Then it holds that

\begin{equation}
\lim_{n \to \infty} \mathbb{E} | P_{B_n}^{(n)}(\varepsilon) - P^{(n)}(\varepsilon) | = 0 .
\end{equation}

**Proof.** For each $n$, define sets

\begin{equation}
H^{(n)} = \{ x \in \mathbb{R}^N ; D^{(n)}(x) \geq 0 \}
\end{equation}

and

\begin{equation}
H_n = \{ x \in \mathbb{R}^N ; D_n(x) \geq 0 \} .
\end{equation}

And for a set $A$ in $\mathbb{R}^N$ denote the complement of $A$ by $\tilde{A}$, and by $I_A$ the indicator of $A$.

Now, we have

\begin{equation}
P^{(n)}(\varepsilon) = \int_{B(n)} P^{(n)}(x, \theta_1)dx + \int_{H_n} P^{(n)}(x, \theta_2)dx
\end{equation}

and

\begin{equation}
P_{B_n}^{(n)}(\varepsilon) = q_1^{(n)} + \int \left[ - D_n(x) \right] I_{H_n}(x)dx .
\end{equation}

Therefore

\begin{equation}
P_{B_n}^{(n)}(\varepsilon) - P^{(n)}(\varepsilon) = \int D^{(n)}(x)[I_{H^{(n)}(x)} - I_{H_n}(x)] \cdot I_{S^{(n)}_\varepsilon}(x)dx
\end{equation}

\begin{equation}
+ \int D^{(n)}(x)[I_{H^{(n)}(x)} - I_{H_n}(x)] \cdot I_{S^{(n)}_\varepsilon}(x)dx .
\end{equation}

Noting that

\begin{equation}
\int \left[ - D_n(x) \right] \cdot [I_{H^{(n)}(x)} - I_{H_n}(x)] \cdot I_{S^{(n)}_\varepsilon}(x)dx \geq 0
\end{equation}

for each $n$, and $0 \leq P_{B_n}^{(n)}(\varepsilon) - P^{(n)}(\varepsilon)$ for each $n$, we have

\begin{equation}
0 \leq P_{B_n}^{(n)}(\varepsilon) - P^{(n)}(\varepsilon)
\end{equation}

\begin{equation}
\leq \int \left[ D^{(n)}(x) - D_n(x) \right] \cdot [I_{H^{(n)}(x)} - I_{H_n}(x)] \cdot I_{S^{(n)}_\varepsilon}(x)dx
\end{equation}
By (3.64)
\[\begin{align*}
\int D'(x) - D_n(x) & | x \in S_n(x) d x + \int S_n(x) d x \\
& \leq E \left[ \int (D'(x) - D_n(x))^2 d x \cdot \mu(S_n(x)) \right]^{1/2} + \frac{\varepsilon}{2}.
\end{align*}\]

Thus, the theorem is proved.

The following theorem can be obtained by modified arguments of the proof in Theorem 3.

**Theorem 4.** Let the hypotheses of Theorem 2 hold. In addition, let, for each \( n \), \( p(n)(\cdot, \theta_1) \) and \( p(n)(\cdot, \theta_2) \) have bounded supports. We put
\[ S_i^{(n)} = \{ x \in \mathbb{R}^N ; p(n)(x, \theta_i) \neq 0 \} \quad i = 1, 2, \text{ and for } n = 1, 2, \ldots. \]

If there exist a positive constant \( M \) such that
\[ \mu(S_i^{(n)}) \leq M \quad \text{for all } i \text{ and } n, \]
then it holds that
\[ E \left| P_{n_i}^{(n)}(e) - P^{(n)}(e) \right| \leq (C \cdot b_n)^{1/2} = M \frac{1}{2} \quad \text{for all } n, \]
where \( b_n \) was defined in Theorem 2 and \( C^* \) is a positive constant which satisfies (3.46).

§ 4. Examples.

**Example 1.** Let \( \{(X_n, \Theta_n)\}_{n=1}^{\infty} \) be a strictly stationary Markov process satisfying the condition \( D_n \) (See \( [2] \), p. 221), and let \( p(x, \theta) \) and \( p^{(n)}(x, \theta | (x', \theta')) \) be a initial probability density function and a \( n \)-step transition probability density function, respectively. From \( [2] \).
\[ \sup \left| p^{(n)}(x, \theta | (x', \theta')) - p(x, \theta) \right| \leq 2 \nu \cdot \lambda^n \quad \text{for all } n = 1, 2, \ldots, \]
where \( 0 < \nu \) and \( 0 < \lambda < 1. \)

If \( p(x, \theta_1) \) and \( p(x, \theta_2) \) are uniformly continuous functions on \( \mathbb{R}^N \), then Assumptions 1, 2, 3 and 4 hold. And it is easily seen that the hypotheses of Theorem 3 are satisfied. Thus, the results of Theorem 1 and Theorem 3 hold.

If \( p(x, \theta_1) \) and \( p(x, \theta_2) \) satisfy the uniform Lipschitz conditions:
\[ |p(x, \theta_i) - p(y, \theta_i)| \leq C_i \cdot ||x - y||, \quad i = 1, 2, \ldots \]
and let \( h_n = n^{-1/N+1} \), then Assumptions 1, 2, 3 and 5 hold, and there exists positive constant \( G \) such that
\[ \lambda^n \leq G \cdot n^{-1/N+1} \quad \text{for all } n = 1, 2, \ldots, \]
and \( b_n = \max \{ h_n, n^{-1} h_n^{-N} \} = h_n = n^{-1/N+1}. \) Therefore, Theorem 2 hold. In addition, if \( p(\cdot, \theta_1) \) and \( p(\cdot, \theta_2) \) have bounded supports in \( \mathbb{R}^N \), then Theorem 4 also hold.
On a Pattern Classification Problem on the Basis

EXAMPLE 2. Let the following conditions be satisfied:

(4.4) \[ p^{(n+m)}(x, \theta_i | (x_1, \theta_1), \ldots, (x_m, \theta_m)) = q^{(n+m)}(\theta_i | \theta_1, \ldots, \theta_m) \cdot p_i(x) \quad i = 1, 2 \text{ and } n, m = 1, 2, \ldots, \]

(4.5) \[ \sum_{i=1}^{2} q^{(n+m)}(\theta_i | \theta_1, \ldots, \theta_m) = 1 \quad \text{for } n, m = 1, 2, \ldots, \]

that is, \( q^{(n+m)}(\theta_i | \theta_1, \ldots, \theta_m) \) is a conditional probability of a category \( \theta_i \) at instant \( n+m \) for given categories \( \theta_1, \theta_2, \ldots, \theta_m \),

(4.6) \[ \sup_{\theta_1, \ldots, \theta_m} | q^{(n+m)}(\theta_i | \theta_1, \ldots, \theta_m) - q^{(n+m)}_{\theta_i} | \to 0 \quad \text{as } n, m \to \infty. \]

And

(4.7) \[ | q^{(n)}_{\theta_i} - q^{(m)}_{\theta_i} | \to 0 \quad \text{as } n, m \to \infty \]

for \( i = 1, 2 \).

If \( p_1(\cdot) \) and \( p_2(\cdot) \) are uniformly continuous functions on \( \mathbb{R}^N \), then the hypothesis of Theorem 1 and Theorem 3 hold. Therefore, the results of Theorem 1 and Theorem 3 hold.

If \( p_1(\cdot) \) and \( p_2(\cdot) \) satisfy uniform Lipschitz conditions,

(4.8) \[ \sup_{\theta_1, \ldots, \theta_m} | q^{(n+m)}(\theta_i | \theta_1, \ldots, \theta_m) - q^{(n+m)}_{\theta_i} | \leq G' \cdot h_m \quad \text{for all } m, \]

where \( G' \) is some positive constant and

(4.9) \[ | q^{(n)}_{\theta_i} - q^{(n+1)}_{\theta_i} | \leq n^{-1} \cdot h_n \quad \text{for } n = 1, 2, \ldots, \]

then Theorem 2 hold. In addition, if \( p_1(\cdot) \) and \( p_2(\cdot) \) have bounded supports in \( \mathbb{R}^N \), then Theorem 4 also hold.

§ 5. Acknowledgement.

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

References