J. math. Res. & Exposition, 1986, No. 1.

Sufficient And Necessary Condition For Convergence of Conditional Error Probability in NN-Pattern Discrimination*

Chen Gui-jing (陈佳景)   Kong Fan-chao (孔繁超)

(Anhui University)

Abstract

Let (θ_1, X_1), ⋯, (θ_n, X_n), (θ, X) be iid random vectors, where $\theta \in \{0, 1\}$, $X \in R^d$. Denote by θ'_n the nearest neighbour discriminator of θ based on the training samples (θ_1, X_1), ⋯, (θ_n, X_n) and the observed X; put $R \triangleq 2E[P(\theta=0 | X)P(\theta=1 | X)]$ and $L_n \triangleq P\left(\theta'_n \neq \theta \,\middle|\, \begin{matrix}\theta_1, \cdots, \theta_n \\ X_1, \cdots, X_n\end{matrix}\right)$. This paper gives a sufficient and necessary condition for $L_n \xrightarrow{P} R$ as $n \to \infty$, namely $(P(\theta=0, X=x) - P(\theta=1, X=x))^2 \cdot P(\theta=0, X=x) \cdot P(\theta=1, X=x) = 0$ for every $x \in R^d$. This generalizes a previous result of the authors [5] and improves a result of Wagner, T. J. [2].

§ I Introduction and Result

Let $(\theta_1, X_1), \cdots, (\theta_n, X_n)$, (θ, X) be iid random vectors, where $\theta \in \Theta \triangleq \{0, 1\}$, $X \in \mathscr{X}$, $\mathscr{X}$ is a Borel set in R^d. Let the distribution of (θ, X) be defined as

$$\left.\begin{array}{l} P(\theta=i) \triangleq \eta_i, \quad i=0, 1 \\ P(dx | \theta=i) = f_i(x)\mu(dx), \quad x \in \mathscr{X}, i=0, 1 \end{array}\right\} \tag{1}$$

where μ is a σ-finite measure in $(\mathscr{X}, \mathscr{B}_x)$ with support $\mathscr{X}$, where $\mathscr{B}_x$ is the σ-field of all Borel subsets of $\mathscr{X}$. The marginal distribution of X can be written as

$$Q(dx) = [\eta_0 f_0(x) + \eta_1 f_1(x)]\mu(dx) \triangleq f(x)\mu(dx) \tag{2}$$

Without loss of generality, in the sequel we shall assume $f(x) > 0$ for every $x \in \mathscr{X}$. The conditional distribution of θ given x is

$$\eta_i(x) \triangleq P(\theta=i | X=x) = \frac{1}{f(x)} \eta_i f_i(x), \quad i=0, 1,$$

If the distribution Q possesses atoms, denote the set of these atoms by $\mathscr{X}^{(1)} \triangleq \{a_1, a_2, \cdots\}$. Put $\mathscr{X}^{(2)} \triangleq \mathscr{X} - \mathscr{X}^{(1)}$, and

* Received Feb. 23, 1983.

$$q_i(x) \triangleq P(X = x | \theta = i), \quad i = 0, 1, \quad x \in \mathscr{X}, \tag{3}$$

$$q(x) \triangleq P(X = x) = \eta_0 q_0(x) + \eta_1 q_1(x), \quad x \in \mathscr{X}. \tag{4}$$

Notice that $q(x) = q_0(x) = q_1(x) = 0$ for $x \in \mathscr{X}^{(2)}$.

Suppose that we have at our disposal the training samples (θ_i, X_i), $i = 1, \cdots, n$ with given X, the problem is to determine the value θ associated with X. Let $\|\cdot\|$ be the usual Euclidean norm, or $\|x\| = \max_{1<i<d} |b_i|$, where $x = (b_1, \cdots, b_d)'$. The NN-Pattern discrimination is as follows: Denote

$$k_n \triangleq \min\{j: \|X_j - X\| = \min_{1<i<n} \|X_i - X\|\},$$

and use θ_{k_n} as a discrimination value of θ associated with X. In the sequel we shall write $\theta'_n \triangleq \theta_{k_n}$, $X'_n \triangleq X_{k_n}$. The performance of the discrimination can be measured by the error probabilities in various sense, such as

$$R_n \triangleq P(\theta'_n \neq \theta), \qquad T_n \triangleq P(\theta'_n \neq \theta | X_1, \cdots, X_n), \tag{5}$$

$$L_n \triangleq P\left(\theta'_n \neq \theta \,\middle|\, \begin{matrix} \theta_1, \cdots, \theta_n \\ X_1, \cdots, X_n \end{matrix} \right) \tag{6}$$

A fundamental problem in large sample theory of NN-pattern discrimination is to study the convergence of R_n, T_n, L_n. Devroye [1] proved that

$$\lim_{n \to \infty} R_n = 2E[\eta_0(X)\eta_1(X)] \triangleq R \tag{7}$$

regardless of the distribution of (X, θ). Wagner, T.J. [2], Chen and Kong [5] discussed the convergence of L_n, T_n under the assumptions that Q is absolutly continuous with respect to the Lebesgue measure, or that Q is a purely atomic distribution. This paper is devoted to the study of the convergence of L_n, T_n under general distribution (1). Our main result is the following:

Theorem 1. Under the distribution (1) of (θ, X), a sufficient and necessary condition for $L_n \xrightarrow{P} R$ as $n \to \infty$ is

$$(\eta_0 q_0(x) - \eta_1 q_1(x))^2 \eta_0 q_0(x) \eta_1 q_1(x) = 0, \text{ for every } x \in \mathscr{X} \tag{8}$$

or, equivalently,

$$(\eta_0 q_0(a_k) - \eta_1 q_1(a_k))^2 \eta_0 q_0(a_k) \eta_1 q_1(a_k) = 0, \; k = 1, 2, \cdots. \tag{8'}$$

When $\mathscr{X}^{(1)} = \phi$, then (8) is true. Hence it follows from Theorem 1 that

Corollary 1. If the distribution Q of X is nonatomic, then $L_n \xrightarrow{P} R$ as $n \to \infty$.

This corollary gets rid of all supplementary conditions imposed on the distribution of X, thus improves the result in [2].

When $\mathscr{X}^{(2)} = \phi$, from Lemma 4 we have

Corollary 2. If the distribution of X is purely atomic, we have

$$\{L_n \to R\ (a.s.)\ \text{ as } n\to\infty\} \Leftrightarrow (8').$$

This corollary becomes Theorem 2 in [5] when X is a purely discrete random variable.

§ 2 Proof of the Theorem

First we give the notations to be used in the sequel. For $x_1, \cdots, x_n$, where $x_i \in \mathscr{X}$, $i=1,\cdots,n$, write

$$V_{nj} \triangleq \{x: x\in\mathscr{X},\ j=\min(i: \| x_i - x \| = \min_{l<t<n} \| x_t - x \|)\} \tag{9}$$

$$V_{nj}^{(i)} = V_{nj} \cap \mathscr{X}^{(i)} \tag{10}$$

for $j=1,\cdots,n$, $i=1,2$.

$$L_n^{(i)} \triangleq P\left(\theta'_n \neq \theta,\ X\in\mathscr{X}^{(i)} \middle| \begin{matrix} \theta_1, \cdots, \theta_n \\ X_1, \cdots, X_n \end{matrix} \right) \tag{11}$$

$$T_n^{(i)} \triangleq P(\theta'_n \neq \theta,\ X\in\mathscr{X}^{(i)} \mid X_1, \cdots, X_n) \tag{12}$$

$$R^{(i)} \triangleq 2E[\eta_0(X)\eta_1(X) I_{\mathscr{X}^{(i)}}(X)] \tag{13}$$

for $i=1,2$.

Lemma 1. $R = R^{(1)} + R^{(2)}$, $L_n = L_n^{(1)} + L_n^{(2)}$, $T_n = T_n^{(1)} + T_n^{(2)}$, (14)

$$L_n^{(i)} = \sum_{j=1}^{n} \left[\theta_j \int_{V_{nj}^{(i)}} \eta_0 f_0(x)\mu(dx) + (1-\theta_j)\int_{V_{nj}^{(i)}} \eta_1 f_1(x)\mu(dx) \right], \tag{15}$$

$$T_n^{(i)} = \sum_{j=1}^{n} \left[\eta_1(x_j)\int_{V_{nj}^{(i)}} \eta_0 f_0(x)\mu(dx) + \eta_0(x_j)\int_{V_{nj}^{(i)}} \eta_1 f_1(x)\mu(dx) \right], \tag{16}$$

$$R^{(i)} = 2\int_{\mathscr{X}^{(i)}} \eta_1(x)\eta_0(x) f(x)\mu(dx), \tag{17}$$

for $i=1,2$.

Proof. By the definition of NN-Pattern discrimination, we have

$$L_n = \sum_{j=1}^{n} \{ P[X\in V_{nj},\ \theta_j=1,\ \theta=0] + P[X\in V_{nj},\ \theta_j=0,\ \theta=1] \}$$

$$= \sum_{j=1}^{n} \left[\theta_j \int_{V_{nj}} \eta_0 f_0(x)\mu(dx) + (1-\theta_j)\int_{V_{nj}} \eta_1 f_1(x)\mu(dx) \right],$$

from this expression and (10) ~ (12), (15) and (16) follow; (14) and (17) hdd obviously. Lemma 1 is proved.

For definiteness, we shall assume that $\mathscr{X}^{(1)}$ is a countable infinite set, if $\mathscr{X}^{(1)}$ is a finite or empty set, the argument will be much simpler. For every integer $k>0$, denote

$$B_k=\{(x_1,x_2,\cdots): x_i\neq a_k,\ x_i\in\mathscr{X},\ i=1,2,\cdots\}$$

$$A_k=\left\{\begin{pmatrix}\theta_1,\theta_2,\cdots\\ x_1,x_2,\cdots\end{pmatrix}: \begin{matrix}\theta_i\in\Theta\quad, i=1,2,\cdots\\ (x_1,x_2,\cdots)\in B_k\end{matrix}\right\}$$

and $B^*=\bigcup_{k=1}^{\infty}B_k$, $\quad A^*=\bigcup_{k=1}^{\infty}A_k$.

For $(x_1,x_2,\cdots)\bar{\in}B^*$, let

$$j_k=\min\{j: x_j=a_k\},\quad k=1,2,\cdots \tag{18}$$

Lemma 2. $P((X_1,X_2,\cdots)\in B^*)=0$ $\qquad$ (19)

$$P\left(\begin{pmatrix}\theta_1,\theta_2,\cdots\\ X_1,X_2,\cdots\end{pmatrix}\in A^*\right)=0 \tag{20}$$

and for every $(x_1,x_2,\cdots)\bar{\in}B^*$, we have

1° $k\to\infty\Leftrightarrow j_k\to\infty$; $l\neq k\Leftrightarrow j_l\neq j_k$;

2° If $n\geqslant j_k$, then $a_k\in V_{nj_k}^{(i)}$, and when $j>j_k$, $x_j=a_k$, then $V_{nj}^{(1)}=\phi$;

3° When $n\to\infty$, $V_{njk}^{(1)}\downarrow\{a_k\}$ for every $k=1,2,\cdots$.

Proof. Since a_k $(k=1,2,\cdots)$ are atoms of X, by Fubini′s Theorem, (19) and (20) follow easily. The conclusions 1°, 2° hold obviously. In order to prove 3°, note that $V_{nj_k}^{(1)}\downarrow$ as $n\to\infty$. Write $\lim\limits_{n\to\infty}V_{nj_k}^{(1)}=V_k^{(1)}$. For every $a_l\neq a_k$ we have $a_l\in V_{nj_l}^{(1)}$, $a_k\in V_{nj_k}^{(1)}$, $V_{nj_l}^{(1)}\cap V_{nj_k}^{(1)}=\phi$ as $n\geqslant\max\{j_k,j_l\}$ from 2°, hence $a_l\bar{\in}V_k^{(1)}$, $a_k\in V_k^{(1)}$. Since $V_{nj_k}^{(1)}\cap\mathscr{X}^{(2)}=\phi$, therefore, we have $V_k^{(1)}=\{a_k\}$. Lemma 2 is proved.

Lemma 3. For every $\begin{pmatrix}\theta_1,\theta_2,\cdots\\ x_1,x_2,\cdots\end{pmatrix}\bar{\in}A^*$, we have

$$\lim_{n\to\infty}L_n^{(1)}=\sum_{k=1}^{\infty}\left[\eta_0q_0(a_k)-\eta_1q_1(a_k)\right]\theta_{j_k}+\eta_1\sum_{k=1}^{\infty}q_1(a_k) \tag{21}$$

Proof. From (15) in Lemma 1, it follows that

$$L_n^{(1)}=\eta_0\sum_{j=1}^{n}\Big[\theta_j\sum_{a_i\in V_{nj}^{(1)}}q_0(a_i)\Big]-\eta_1\sum_{j=1}^{n}\Big[\theta_j\sum_{a_i\in V_{nj}^{(1)}}q_1(a_i)\Big]+\eta_1\sum_{k=1}^{\infty}q_1(a_k)$$

$$\triangleq\eta_0I_{0n}-\eta_1I_{1n}+\eta_1\sum_{k=1}^{\infty}q_1(a_k). \tag{22}$$

Put $I_i=\sum_{k=1}^{\infty} q_i(a_k)\theta_{j_k}$, $i=0, 1$. In order to prove (21), it is enough to show that $\lim_{n\to\infty} I_{in}=I_i$ for $i=0, 1$. For example when $i=0$, for $\varepsilon>0$ there exists N such that $\sum_{k=N+1}^{\infty} q_0(a_k)<\varepsilon$. From 2° of Lemma 2, when $n\geqslant n_0$, $\triangleq \max\{j_1, j_2, \cdots, j_N\}$, we have

$$|I_{on}-I_0|\leqslant|\sum_{k=1}^{N}\Big[\sum_{a_i\in V_{nj_k}^{(1)}} q_0(a_i)\theta_{j_k}\Big]-\sum_{k=1}^{N} q_0(a_k)\theta_{j_k}|$$

$$+|I_{on}-\sum_{k=1}^{N}\Big[\sum_{a_i\in V_{nj_k}^{(1)}} q_0(a_i)\theta_{j_k}\Big]|+\sum_{k=N+1}^{\infty} q_0(a_k)\theta_{j_k}\triangleq J_{1n}+J_{2n}+J_{3n}, \quad (23)$$

where $J_{3n}\leqslant\sum_{k=N+1}^{\infty} q_0(a_k)<\varepsilon$. From 3° of Lemma 2 there exists $n_1\geqslant n_0$ such that $J_{1n}<\varepsilon$ as $n\geqslant n_1$. Write

$$W_n\triangleq\{j:\ V_{nj}^{(1)}\neq\phi,\quad 1\leqslant j\leqslant n\},$$

then from 2° of Lemma 2 we have

$$J_{2n}\leqslant\sum_{j\in W_n-\{j_1,\cdots,j_N\}}\Big[\sum_{a_i\in V_{nj_k}^{(1)}} q_0(a_i)\Big]$$

$$\leqslant\sum_{k=1}^{\infty} q_0(a_k)-\sum_{k=1}^{N}\Big[\sum_{a_i\in V_{nj_k}^{(1)}} q_0(a_i)\Big]$$

$$\leqslant\sum_{k=N+1}^{\infty} q_0(a_k)<\varepsilon,$$

as $n\geqslant n_1$. Therefore Lemma 3 is proved.

Lemma 4. $\{L_n^{(1)}\to R^{(1)}\ (a.s.)\ as\ n\to\infty\}\Leftrightarrow(8')$ (24)

Proof. By (17), $R^{(1)}=2\sum_{k=1}^{\infty}\frac{\eta_0\eta_1 q_0(a_k)q_1(a_k)}{q(a_k)}$, hence from Lemma 3 we have that $\{L_n^{(1)}\to R^{(1)}\ (a.s.)\ as\ n\to\infty\}\Leftrightarrow\sum_{k=1}^{\infty}\big[\eta_0 q_0(a_k)-\eta_1 q_1(a_k)\big]\theta_{j_k}$

$$=R^{(1)}-\eta_1\sum_{k=1}^{\infty} q_1(a_k),\quad a.s. \quad (25)$$

It is easy to prove that if $(8')$ holds then $R^{(1)}=\eta_1\sum_{k=1}^{\infty} q_1(a_k)$, therefore (25) follows. On the other hand, if (25) is true, then, since the right hand side of (25) is a constant, by computing the variances of two sides of (25) under $X_{j_k}=a_k$, $k=1, 2, \cdots$, it follows that

$$\sum_{k=1}^{\infty}\Big[(\eta_0 q_0(a_k)-\eta_1 q_1(a_k))^2\eta_0 q_0(a_k)\eta_1 q_1(a_k)/q^2(a_k)\Big]=0\Rightarrow(8').$$

Therefore Lemma 4 is proved.

We turn to discuss $L_n^{(2)}$ now.

Lemma 5. Let $\varphi(x)$ be a Borel measurable function on $\mathscr{X}$, $|\varphi(x)|<M<\infty$. Then there exists a set $A\in\mathscr{B}_x$ such that $Q(A)=0$ and for every $x\bar{\in}A$ we have

$$\varphi(X_n'(x))\xrightarrow{P}\varphi(x) \qquad as \quad n\to\infty,$$

where $X_n'(x)$ is the nearest neighbouring point of x in $X_1, \cdots, X_n$.

Proof. Denote by $X_n''(x)$ the second nearest neighbouring point of x in $X_1, \cdots, X_n$, then for every $x\in\mathscr{X}$,

$$X_n''(x)\to x \quad (a.s.) \quad \text{as} \quad n\to\infty. \tag{26}$$

Since φ is a bounded function, there exists a set $A\in\mathscr{B}_x$ such that $Q(A)=0$ and

$$\lim_{\rho\to0}\frac{\int_{S_{x,\rho}}|\varphi(y)-\varphi(x)|Q(dy)}{Q(S_{x,\rho})}=0 \quad \text{for} \quad x\bar{\in}A, \tag{27}$$

by Theorem 2.9.8 of [6], where

$$S_{x,\rho}\triangleq\{y:\ y\in\mathscr{X},\|x-y\|<\rho\}.$$

Denote by $G_{n,x}(d\rho)$ the distribution of $\|X_n''(x)-x\|$, then if $x\bar{\in}A$, we have

$$E[|\varphi(X_n'(x))-\varphi(x)|]=E\{E[|\varphi(X_n'(x))-\varphi(x)|\,\big|\,\|X_n''(x)-x\|]\}$$
$$=\left[\int_0^{\rho_0}+\int_{\rho_0}^{\infty}\right]\frac{\int_{S_{x,\rho}}|\varphi(y)-\varphi(x)|Q(dy)}{Q(S_{x,\rho})}G_{n,x}(d\rho)\triangleq I_1+I_2.$$

For arbitrarily given $\varepsilon>0$, from (27) it follows that there exists $\rho_0>0$ such that $I_1<\varepsilon$. On the other hand,

$$K(x,\rho_0)\triangleq\sup_{\rho_0<\rho}\frac{\int_{S_{x,\rho}}|\varphi(y)-\varphi(x)|Q(dy)}{Q(S_{x,\rho})}<M<\infty.$$

Hence, by (26).

$$I_2<MP[\|X_n''(x)-x\|>\rho_0]\to0 \qquad as \quad n\to\infty.$$

Since $\varepsilon>0$ is arbitrary, the proof of Lemma 5 is concluded.

Lemma 6. $T_n^{(2)}\xrightarrow{P}R^{(2)}$ $\quad as \quad n\to\infty.$ (28)

Proof. From (16) of Lemma 1, we have

$$T_n^{(2)}=\sum_{j=1}^n\eta_1(x_j)\int_{V_{nj}^{(2)}}\eta_0f_0(x)\mu(dx)+\sum_{j=1}^n\eta_0(x_j)\int_{V_{nj}^{(2)}}\eta_1f_1(x)\mu(dx)\triangleq T_{1n}^{(2)}+T_{2n}^{(2)},$$

$$R^{(2)}=2\int_{\mathscr{X}^{(2)}}\eta_1(x)\eta_0f_0(x)\mu(dx).$$

Hence, in order to prove (28), it is enough to prove $T_{1n}^{(2)} \xrightarrow{P} \frac{1}{2} R^{(2)}$ as $n \to \infty$.

Because $0 \leqslant \eta_1(x) \leqslant 1$, from Lemma 5 there exists a set $A \in \mathscr{B}_x$ such that $Q(A) = 0$, and for $x \bar{\in} A$, we have $\eta_1(X_n'(x)) \xrightarrow{P} \eta_1(x)$ as $n \to \infty$. Note that for $x \in \mathscr{X}^{(2)} - A$,

$$\sum_{j=1}^{n} \eta_1(x_j) I_{V_{nj}^{(2)}}(x) = \eta_1(X_n'(x)) \xrightarrow{P} \eta_1(x) \quad as \quad n \to \infty.$$

Then by the dominated convergence theorem and Fubini's theorem, we have

$$E[\ E[| T_{1n}^{(2)} - \frac{1}{2} R^{(2)} |] = E\{| \int_{\mathscr{X}^{(2)}} (\eta_1(X_n'(x)) - \eta_1(x)) \eta_0 f_0(x) \mu(dx) |\}$$

$$\leqslant E\{\int_{\mathscr{X}^{(2)}} | \eta_1(X_n'(x)) - \eta_0 f_0(x) \mu(dx)\}$$

$$= \int_{\mathscr{X}^{(2)} - A} E[| \eta_1(X_n'(x)) - \eta_1(x) |] \eta_0 f_0(x) \mu(dx)$$

$$\to 0 \quad as \quad n \to \infty.$$

It can be shown in the same way that $E[| T_{2n}^{(2)} - \frac{1}{2} R^{(2)} |] \to 0$ as $n \to \infty$, and the proof of Lemma 6 is completed.

Corollary 3. $T_n \xrightarrow{P} R$ as $n \to \infty$.

Proof. This corollary follows from the fact that the proof of Lemma 6 is still valid when $\mathscr{X}^{(2)}$ is replaced by $\mathscr{X}$.

Lemma 7. Let $\tilde{Q}$ be a finite measure without atom on $(R^d, \mathscr{B}^d)$, then we have

$$\lim_{\rho \to 0} \tilde{Q}\{y: \| y - x \| < \rho, \quad y \in R^d\} = 0 \qquad (29)$$

uniformly for $x \in R^d$.

Proof. Obviously it is enough to prove (29) for the case $\| x \| = \max_{l<i<d} | b_i |$ $(x = (b_1, \cdots, b_d)')$. For arbitrarily given $\varepsilon > 0$, there exists $M > 0$ such that $\tilde{Q}(A_M^c) < \varepsilon$, where $A_M \triangleq \{y: \| y \| < M, y \in R^d\}$. For every $k (= 1, 2, 3, \cdots)$, we split A_M into 2^{kd} equal "cuboids" A_{ki}, $i = 1, 2, \cdots, 2^{kd}$ with marginal length $\frac{1}{2^k} M$ such that for every k we have that

$$A_M = \bigcup_{i=1}^{2^{kd}} A_{ki}, \qquad A_{ki} \cap A_{kj} = \phi \quad (i \neq j)$$

and for every (k, k', i, j) we have

$$A_{ki} \cap A_{k'j} = \phi \quad \text{or} \quad A_{ki} \subset A_{k'j} \qquad (k' < k).$$

Then there exists k_0 such that

$$\tilde{Q}(A_{k_0 i}) < \varepsilon, \quad i = 1, 2, \cdots, 2^{k_0 d}. \qquad (30)$$

For, if on the contrary (30) is not true, there will exist a set-sequence $\{A_{ki_k}, \ k=1,2,\cdots\}$ such that

$$\widetilde{Q}(A_{ki_k}) \geqslant \varepsilon, \quad k=1,2,\cdots. \tag{31}$$

By finiteness of $\widetilde{Q}$, there exists only a finite number of sets in $\{A_{ki_k}, \ k=1,2, \cdots\}$ such that they do not intersect with each other, hence there exists k_1 such that

$$A_{k_1 i_{k_1}} \supset A_{(k_1+1)}A_{(k_1+1)} \supset A_{(k_1+2)i_{(k_1+2)}} \supset \cdots.$$

Denote $A_0 = \lim_{k\to\infty} A_{ki_k}$, then $A_0 = \phi$ or $\{x_0\}$ for some $x_0 \in R^d$, and we have $\lim_{k\to\infty} \widetilde{Q}(A_{ki_k}) = 0$ since $\widetilde{Q}$ has no atom. This is contrary to (31) Hence (30) holds. Denote $A_{k_0 o} \triangleq A_M^c$. Then for every $x \in R^d$, there cexist $N(d)$ sets in $\{A_{k_0 i}, \ i=0,1,\cdots, 2^{dk_0}\}$ such that they cover $\{y: \| y-x \| < \rho_0, \ y \in R^d\}$, where $\rho_0 = \frac{1}{2^{k_0-1}} M$, $N(d)$ depends only on d. From (30) we have

$$\widetilde{Q}\{y: \| y-x \| < \rho_0, \ y \in R^d\} < N(d)\varepsilon.$$

Hence (29) follows by arbitrariness of ε.

Lemma 8. $L_n^{(2)} - T_n^{(2)} \xrightarrow{P} 0 \quad as \quad n \to \infty$ (32)

Proof. From Lemma 1,

$$L_n^{(2)} = \sum_{j=1}^n \theta_j \int_{V_{nj}^{(2)}} \eta_0 f_0(x)\mu(dx) + \sum_{j=1}^n (1-\theta_j) \int_{V_{nj}^{(2)}} \eta_1 f_1(x)\mu(dx) \triangleq L_{1n}^{(2)} + L_{2n}^{(2)}.$$

Notations $T_{in}^{(2)}$, $i=1,2$, were defined in Lemma 6. Therefore in order to prove (32), it is enough to prove $L_{in}^{(2)} - T_{in}^{(2)} \xrightarrow{P} 0$ as $n \to \infty$, $i=1,2$.

Let K be a compact set in R^d, and

$$\Gamma_n^1 \triangleq \sup_{x \in K \cap \mathscr{X}^{(2)}} \{\{\inf_{1 \le i \le n} \| X_i - x \|\},$$

then it can be shown that

$$\Gamma_n^1 \to 0 \quad (a.s.) \quad as \quad n \to \infty \tag{33}$$

by the same method of [2]. Note that $\mathscr{X}^{(2)}$ contains no atom of Q. By Lemma 7 we have

$$\lim_{\rho \to 0} Q[\mathscr{X}^{(2)} \cap S_{x,\rho}] = 0 \tag{34}$$

uniformly for $x \in \mathscr{X}^{(2)}$. By (34), (33) it can be shown that

$$\Gamma_n \triangleq \max_{1 \le j \le n} \int_{V_{nj}^{(2)}} f(x)\mu(dx) \to 0 \quad (a.s.) \quad as \quad n \to \infty$$

by the same method of [2]. Note that when $(x_1, \cdots, x_n)$ is given, $Q_j - E[\theta_j | x_j]$, $j = 1, \cdots, n$ are conditionally independent. Hence we have

$$E\{[L_{1n}^{(2)} - T_{1n}^{(2)})^2 | x_1, \cdots, x_n\}$$

$$= E\{[\sum_{j=1}^{n} (\theta_j - E(\theta_j | x_j)) \int_{V_{nj}^{(2)}} \eta_0 f_0(x) \mu(\mathrm{d}x)]^2 | x_1, \cdots, x_n\}$$

$$= \sum_{j=1}^{n} \{E[(\theta_j - E(\theta_j | x_j))^2 | x_j][\int_{V_{nj}^{(2)}} \eta_0 f_0(x) u(\mathrm{d}x)]^2\}$$

$$\leqslant \max_{1 \leqslant j \leqslant n} \int_{V_{nj}^{(2)}} f(x) \mu(\mathrm{d}x) \to 0 \quad as \quad n \to \infty.$$

Therefore, by the dominated convergence theorem we have

$$E(L_{1n}^{(2)} - T_{1n}^{(2)})^2 = E\{E[(L_{1n}^{(2)} - T_{1n}^{(2)})^2 | X_1 \cdots, X_n]\} \to 0 \quad as \quad n \to \infty.$$

This proves that $L_{1n}^{(2)} - T_{1n}^{(2)} \xrightarrow{P} 0$ as $n \to \infty$. Similarily it can be shown that $L_{2n}^{(2)} - T_{2n}^{(2)} \xrightarrow{P} 0$ as $n \to \infty$. Lemma 8 is proved.

Now it is easy to complete the proof of Theorem 1.

Proof of Theorem 1. Denote

$$\Delta\begin{pmatrix} \theta_1, \theta_2, \cdots \\ x_1, x_2, \cdots \end{pmatrix} = \begin{cases} \sum_{k=1}^{\infty} [\eta_0 q_0(a_k) - \eta_1 q_1(a_k)] \theta_{jk} + \eta_1 \sum_{k=1}^{\infty} q_1(a_k) + R^{(2)} & \\ & as \begin{pmatrix} \theta_1, \theta_2, \cdots \\ x_1, x_2, \cdots \end{pmatrix} \bar{\in} A^* \\ R & as \begin{pmatrix} \theta_1, \theta_2, \cdots \\ x_1, x_2, \cdots \end{pmatrix} \in A^*, \end{cases}$$

where A^*, j_k were defined in Lemma 1, 2. Employing the Lemmas given above, we can easily see

$$L_n = L_n^{(1)} + L_n^{(2)} \xrightarrow{P} \triangle \quad as \quad n \to \infty.$$

An argument similar to those used in the proof of Lemma 4 shows that

$$\triangle = R^{(1)} + R^{(2)} = R \quad \Leftrightarrow (8')$$

which ends the proof of Theorem 1.

References

[1] Devroye, L., On the Asymptotic Probability of Error In Nonparametric Discrimination, Ann. Statist., 1981, 1320—1327.

[2] Wagner, T.J., Convergence of the NN Rule, IEEE Trans. Inform.

Theory , 1971, 566 —570 .
〔3〕Jozsef , Fritz , Distribution-Free Exponential Bounded for NN Pattern Classification, IEEE. Trans . Inform. Theory. 1975, 552 —557 .
〔4〕Chen Xiru (陈希孺), On the Theory of Nearest Neighbour Prediction Under Squares Loss.
〔5〕Chen Gui-jing (陈佳景), Kong Fan-chao (孔繁超). Nearest Neighbour Discrimination in Discrete Case .
〔6〕Federer, H., Geometric Measure Theory, Springer-Verlag , 1969.