

MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

2

ADPORSR-TR- 85-0823

AD-A160 305

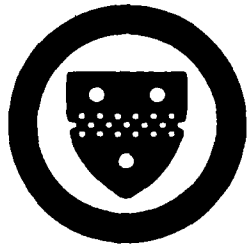
EXPONENTIAL BOUND FOR ERROR PROBABILITY  
IN NN-DISCRIMINATION\*

Z.D. Bai

CENTER FOR MULTIVARIATE ANALYSIS  
UNIVERSITY OF PITTSBURGH

**Center for Multivariate Analysis**  
**University of Pittsburgh**

DTIC FILE COPY



DTIC  
ELECTE  
OCT 11 1985  
S B D

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

85 10 11 040  
MS F

EXPONENTIAL BOUND FOR ERROR PROBABILITY  
IN NN-DISCRIMINATION\*

Z.D. Bai

CENTER FOR MULTIVARIATE ANALYSIS  
UNIVERSITY OF PITTSBURGH

April 1985

Technical Report #85-15

DTIC  
ELECTE  
OCT 11 1985  
S D  
B

Center for Multivariate Analysis  
Thackeray Hall  
University of Pittsburgh  
Pittsburgh, PA 15260

\*This work was partially supported by the Air Force Office of Scientific Research under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes not withstanding any copyright notation herein.

**DISTRIBUTION STATEMENT A**

Approved for public release  
Distribution Unlimited

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)  
NOTICE  
THIS DOCUMENT IS UNCLASSIFIED  
DATE 10/11/85 BY  
EAC  
MATTHEW  
Chief, Technical Information Division

## ABSTRACT

Let  $(\theta_1, X_1), \dots, (\theta_n, X_n)$  be  $n$  simple samples drawn from the population  $(\theta, X)$  which is a  $\{1, 2, \dots, s\} \times R^d$ -valued random vector. Suppose that  $X$  is known. Let  $X'_n$  be the nearest one among  $X_1, \dots, X_n$ , from  $X$ , in the sense of Euclidean norm or  $l_\infty$ -norm, and let  $\theta'_n$  be the  $\theta$ -value paired with  $X'_n$ . The posterior error probability is defined by  $L_n = P(\theta'_n \neq \theta | (\theta_1, X_1), \dots, (\theta_n, X_n))$ . It is well known that  $EL_n = P(\theta'_n \neq \theta)$  has always a limit  $R$ . In this paper it is shown that for any  $\epsilon > 0$ , there exist constants  $c < \infty$  and  $b > 0$  such that  $P(|L_n - R| \geq \epsilon) \leq ce^{-bn}$ , under the only assumption that the marginal distribution of  $X$  is nonatomic.



Accession For	
NSIC SPA&I	<input checked="" type="checkbox"/>
DTIC	<input type="checkbox"/>
Unpublished	<input type="checkbox"/>
Classification	
Distribution/	
Availability Codes	
A1111111111	
Dist	Special
A-1	

$(\theta_{\text{sub } d}, x_{\text{sub } d})$        $(\theta_{\text{sub } 1}, x_{\text{sub } n})$        $(\theta_{\text{sub } 2}, x_{\text{sub } n})$

1. INTRODUCTION

Let  $(\theta, X)$  be a  $(1, 2, \dots, s) \times \mathbb{R}^d$ -valued random vector and  $(\theta_1, X_1), \dots, (\theta_n, X_n)$  be iid. samples drawn from  $(\theta, X)$ . The so-called discrimination problem is to find a function of  $X$ , usually depending upon  $(\theta_1, X_1), \dots, (\theta_n, X_n)$ , which is used to predict the value of  $\theta$ . One of the most frequently used approaches is the nearest neighbour (NN) discrimination rule, which is defined as follows: Let  $\|\cdot\|$  be a norm in  $\mathbb{R}^d$ . Usually we take this norm being the Euclidean one or  $l_\infty$ -norm. When  $X = x$  is given, we rearrange  $X_1, \dots, X_n$  according to the increasing order of the distances  $\|X_i - x\|$ , namely,

$$\|X_{R_1} - x\| \leq \|X_{R_2} - x\| \leq \dots \leq \|X_{R_n} - x\|,$$

ties are broken by comparing indices, for instance, if  $\|X_i - x\| = \|X_j - x\|$  and  $i < j$ , then  $X_i$  is rearranged before  $X_j$ . In view of the rearrangement of  $X$ 's,  $\theta_1, \dots, \theta_n$  are rearranged as  $\theta_{R_1}, \dots, \theta_{R_n}$ . Let  $k \leq n$  be a positive integer. Then  $\theta_n^{(k)}$ , called  $k$ -NN discrimination, is defined to be the value which has the biggest frequency among  $\theta_{R_1}, \dots, \theta_{R_k}$ . If such value is not unique, then we randomly take one such value with equal probability.

The probability of misdiscrimination and the posterior one are defined to be

*> K-NN discrimination (misclassification);*

$$R_n^{(k)} = P(\theta_n^{(k)} \neq \theta) \quad (1.1)$$

and

$$L_n^{(k)} = P(\theta_n^{(k)} \neq \theta | (\theta_1, X_1), \dots, (\theta_n, X_n)), \quad (1.2)$$

which are simply called error probability or posterior error probability, respectively. For  $k = 1$ ,  $X_{R_1}$ ,  $\theta_n^{(1)}$ ,  $R_n^{(1)}$  and  $L_n^{(1)}$  are simply denoted by  $X_n^1$ ,  $\theta_n^1$ ,  $R_n$  and  $L_n$  respectively. For reviews of the literature on this topic, the reader is referred to Cover and Hart (1967), Wagner (1971), Fritz (1975), Cover (1968), Devroye (1981), Bai (1984).

It was shown that whatever the distribution of  $(\theta, X)$  is, there is always a constant  $R^k$ , depending upon  $k$ , such that  $\lim_{n \rightarrow \infty} R_n^{(k)} = R^k$ , and the constant  $R^k$  satisfies  $R^* \leq R^k \leq R \frac{\Delta}{S-1} R^1 \leq R^*(2 - \frac{S}{S-1} R^*)$ , where  $R^*$  is the error probability in Bayesian discrimination. Let

$$P_i(x) = P(\theta=i | X=x), \quad i = 1, 2, \dots, s. \quad (1.3)$$

Then it can be shown that

$$R = 1 - \sum_{i=1}^S E P_i^2(X). \quad (1.4)$$

In Chen and Kong (1983), it was proved that when  $S = 2$ ,  $L_n \xrightarrow{P} C$  for some constant  $C$  if and only if

$$(P(\theta=1, X=x) - P(\theta=2, X=2))^2 P(\theta=1, X=x) P(\theta=2, X=x) = 0, \quad (1.5)$$

and  $C = R$ . This implies that if the marginal distribution

$F$  of  $X$  has no atoms, then  $L_n \xrightarrow{P} R$ . But according to Hewitt-Savage zero-one law we know that  $P(L_n \rightarrow R) = 0$  or  $1$ , if  $F$  has no atoms. The problem is under what conditions  $P(L_n \rightarrow R) = 1$  is true. To make it clear, let us introduce the following notations.

Suppose  $X_1, \dots, X_n$  are given. Split  $R^d$  into  $n$  sets  $V_{n1}, \dots, V_{nn}$ , such that  $x \in V_{nj}$  if and only if  $X_j$  is the nearest neighbor of  $x$ . Note that  $V_{nj}$ ,  $j = 1, 2, \dots, n$ , are random sets. By the definition of  $L_n$ , it is not difficult to compute out that

$$L_n = 1 - \sum_{j=1}^n \sum_{i=1}^s I(\theta_j=i) \int_{V_{nj}} P_i(x) F(dx). \quad (1.6)$$

It is easy to see that when  $X_1, \dots, X_n$  are given, the terms under the first summation in (1.6) is a sum of conditionally independent and bounded random variables, and the coefficients have expectations not exceeding  $\frac{1}{n}$ . Thus there is a strong reason to believe that there should be an exponential bound for  $P(|L_n - R| \geq \epsilon)$ . But up to now, as the author knows, the best result is due to Fritz (1975). It was shown that

$$P(|L_n - R| \geq \epsilon) \leq ce^{-b\sqrt{n}} \quad (1.7)$$

under the assumptions that  $F$  has no atoms and that  $P_i(x)$ ,  $i = 1, 2, \dots, s$ , are a.e.x.F. continuous. There are some extra examples to support the conjecture that  $\sqrt{n}$  in (1.7)



could be improved as to  $n$ . Also, there is much evidence to suggest that the assumption of continuity of  $P_i(x)$ 's could be abandoned (see Z. D. Bai, X. R. Chen and G. J. Chen (1984)). The main purpose of this paper is to solve these two problems. In §2, we shall show some lemmas and the proof of main result will be given in §3. In a further paper, we will give the necessary and sufficient conditions for  $L_n \rightarrow R$  a.s.

## 2. SOME LEMMAS

In the sequel, we shall need the following lemmas: Some of them are known, we will quote them below without proof. We shall only give the proof for new ones.

Lemma 1. Let  $\xi$  be a random variable with distribution  $F(x)$  (for convenience we assume that  $F(x)$  is left continuous), and let  $\eta$  be a random variable uniformly distributed over the interval  $(0,1)$  and being independent of  $\xi$ . Then

$$Z \stackrel{\Delta}{=} F(\xi) + [F(\xi+0) - F(\xi)]\eta$$

is uniformly distributed over  $(0,1)$ .

The proof is not very difficult and is omitted here.

Lemma 2. Let  $F$  be a nonatomic probability measure defined on  $R^d$  and  $A$  be a measurable set in  $R^d$ . Then for any  $C \in [0, F(A)]$ , there is a measurable set  $B \subset A$  such that  $F(B) = C$ .

The proof of this lemma can be found in Fritz (1975) or Loeve (1977) p.101.

For the need in the sequel, we introduce the concept of  $\omega$ -cone. Let  $O$  be a point in  $R^d$  and  $A$  be a measurable set in  $R^d$ . Then  $A$  is called an  $\omega$ -cone with vertex  $O$  if for any  $x, y \in A$ ,  $\|x-y\| < \max(\|O-x\|, \|O-y\|)$ .

Lemma 3. For each  $d$ , there is a positive integer  $m = m(d)$ , depending only upon  $d$ , such that  $R^d$  can be split into  $m$  disjoint  $\omega$ -cones with a common preassigned vertex.

The proof refers to Fritz (1975).

Let  $F$  be a nonatomic probability measure on  $R^d$  and  $A \subset R^d$  be a measurable subset of positive  $F$ -measure. Then, by lemma 3, we can split  $R^d$  into  $m$   $\omega$ -cones with a common preassigned vertex. Write the intersections of  $A$  and each  $\omega$ -cone as  $K_1, \dots, K_m$ , with  $\bigcup_{t=1}^m K_t = A$ . If  $F(K_t) = 0$ ,  $t = 1, 2, \dots, \ell$ , and  $F(K_t) > 0$ ,  $t = \ell + 1, \dots, m$ , by lemma 2, we can split  $K_{\ell+1}$  into  $\ell + 1$  subsets with equal  $F$ -measure. Drop the original  $K_1, \dots, K_{\ell+1}$  and write the new  $\ell + 1$  subsets as  $K_1, \dots, K_{\ell+1}$ . Note that the new  $K_1, \dots, K_{\ell+1}$  are also  $\omega$ -cones with the same vertex. Hence we obtain the following lemma.

Lemma 4. Let  $F$  be nonatomic,  $A$  be of positive  $F$ -measure and  $O$  be a point in  $R^d$ . Then there are  $m$  disjoint  $\omega$ -cones  $K_1, \dots, K_m$ , having the same vertex  $O$  and satisfying

$$\bigcup_{t=1}^m K_t \subset A$$

$$\sum_{t=1}^m F(K_t) = F(A)$$

$$F(K_t) > 0, \quad t = 1, 2, \dots, m.$$

In the last section, we defined the partition  $V_{nj}$  of  $R_d$ , for random points  $X_1, \dots, X_n$ . We point out that for any given points  $X_1, \dots, X_n$ , even they are not random, we can similarly define  $V_{nj}$ ,  $j = 1, 2, \dots, n$ . Also, if  $1 \leq j \leq k \leq n$ , we always have  $V_{nj} \subset V_{kj}$ .

Lemma 5, Let  $X_1, \dots, X_k$  be  $k$  points in  $R^d$  and  $F$  be a nonatomic probability measure defined on  $R^d$ . Suppose that  $F(V_{kj}) = q_j > 0$ ,  $j = 1, 2, \dots, k$  and that  $X_{k+1}, \dots, X_n$  are iid. random points with distribution  $F$ . Write  $U_{nj} = F(V_{nj})$ . Then there are  $mk$  random variables  $\xi_{j\ell}$ ,  $j = 1, 2, \dots, k, \ell = 1, 2, \dots, m$ , satisfying the following conditions,

$$1) P(U_{nj} \leq \sum_{\ell=1}^m \xi_{j\ell}) = 1, \quad j = 1, 2, \dots, k. \quad (2.1)$$

$$2) 0 \leq \xi_{j\ell} \leq q_{j\ell}, \quad j = 1, 2, \dots, k, \ell = 1, 2, \dots, m \quad (2.2)$$

$$3) P(\xi_{j\ell} \geq w_{j\ell}, \quad j = 1, 2, \dots, k, \ell = 1, 2, \dots, m) \\ = (1 - \sum_{j=1}^k \sum_{\ell=1}^m w_{j\ell})^{n-k}, \quad \text{for } 0 \leq w_{j\ell} \leq q_{j\ell}, \quad (2.3)$$

where  $q_{j\ell}$ ,  $j = 1, 2, \dots, k, \ell = 1, 2, \dots, m$ , are positive constants, independent of  $X_{k+1}, \dots, X_n$  (may depend upon  $X_1, \dots, X_k$  and  $F$ ),

and satisfy  $\sum_{\ell=1}^m q_{j\ell} = q_j$  (of course,  $\sum_{j=1}^k \sum_{\ell=1}^m q_{j\ell} = \sum_{j=1}^k q_j = 1$ ).

Proof. According to Lemma 4, split  $V_{kj}$  into  $m$   $\omega$ -cones of positive  $F$  measure, denoted by  $K_{j\ell}$ ,  $j = 1, 2, \dots, k$ ,  $\ell = 1, 2, \dots, m$ . Write  $q_{j\ell} \stackrel{\Delta}{=} F(K_{j\ell}) > 0$ ,  $j = 1, 2, \dots, k$ ,  $\ell = 1, 2, \dots, m$ . It is evident that  $\sum_{\ell=1}^m q_{j\ell} = q_j$ ,  $j = 1, 2, \dots, k$ .

Define

$$\begin{aligned} D_{j0} &= ||X_j - X||, \quad j = 1, 2, \dots, k, \\ D_{jt} &= ||X_j - X_t||, \quad j = 1, 2, \dots, k, \quad t = k+1, \dots, n, \\ I_{j\ell} &= \{t: k+1 \leq t \leq n, X_t \in K_{j\ell}\}, \\ n_{j\ell} &= * \{I_{j\ell}\} = \text{the number of elements of } I_{j\ell}, \\ H_j &= \begin{cases} \min_{t \in I_{j\ell}} D_{jt} & \text{if } I_{j\ell} \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

By the definition of  $\omega$ -cone,  $X \in V_{nj} \cap K_{j\ell}$  and  $X_t \in K_{j\ell}$  imply that  $D_{j0} < D_{jt}$  hence  $D_{j0} < H_{j\ell}$ . Therefore

$$\begin{aligned} U_{nj} &= P(X \in V_{nj} | X_{k+1}, \dots, X_n) \\ &= \sum_{\ell=1}^m P(X \in V_{nj} \cap K_{j\ell} | X_{k+1}, \dots, X_n) \\ &= \sum_{\ell=1}^m q_{j\ell} P(X \in V_{nj} | X \in K_{j\ell}, X_{k+1}, \dots, X_n) \\ &\leq \sum_{\ell=1}^m q_{j\ell} P(D_{j0} < H_{j\ell} | X \in K_{j\ell}, X_{k+1}, \dots, X_n) \\ &= \sum_{\ell=1}^m q_{j\ell} F_{j\ell}(H_{j\ell}) \\ &= \sum_{\ell=1}^m q_{j\ell} \min_{t \in I_{j\ell}} F_{j\ell}(D_{jt}), \end{aligned} \tag{2.4}$$

where  $F_{j\ell}(u) = P(D_{j0} < u | X \in K_{j\ell})$  and the minimum in (2.4) is one if  $I_{j\ell} = \phi$ .

Construct  $n-k$  iid. random variables  $\eta_{k+1}, \dots, \eta_n$ , which are uniformly distributed over the interval  $(0,1)$  and are independent of  $X_{k+1}, \dots, X_n$ . Define for each  $t = k+1, \dots, n$ ,

$$G_t = F_{j\ell}(D_{jt}) + [F_{j\ell}(D_{jt}+0) - F_{j\ell}(D_{jt})] \eta_t, \text{ if } t \in I_{j\ell}. \quad (2.5)$$

Note that when  $I_{j\ell}$ ,  $j = 1, 2, \dots, k$ ,  $\ell = 1, 2, \dots, m$ , are given and  $X_t \in K_{j\ell}$ ,  $D_{jt}$  has the same conditional distribution as  $D_{j0}$  when  $X \in K_{j\ell}$  given. Thus, by lemma 1,  $G_t$  is uniformly distributed over  $(0,1)$ . Also, when  $I_{j\ell}$ ,  $j = 1, 2, \dots, k$ ,  $\ell = 1, 2, \dots, m$ , are given,  $G_t$  depends only upon  $X_t$  and  $\eta_t$ , hence  $G_t$ ,  $t = k+1, \dots, n$ , are conditionally independent. Define

$$\xi_{j\ell} = \begin{cases} q_{j\ell} \min_{t \in I_{j\ell}} G_t & \text{if } I_{j\ell} \neq \phi \\ q_{j\ell} & \text{otherwise,} \end{cases} \quad (2.6)$$

Evidently, (2.2) follows from (2.6). Also, (2.1) follows from (2.4) - (2.6). Finally, we have for  $0 \leq w_{j\ell} \leq q_{j\ell}$ ,  $j = 1, 2, \dots, k$ ,  $\ell = 1, 2, \dots, m$ ,

$$\begin{aligned} & P(\xi_{j\ell} \geq w_{j\ell}, j=1, 2, \dots, k, \ell=1, 2, \dots, m) \\ &= \sum \left[ \frac{(n-k)!}{\prod_{j=1}^k \prod_{\ell=1}^m n_{j\ell}!} \right] \left[ \prod_{j=1}^k \prod_{\ell=1}^m q_{j\ell}^{n_{j\ell}} \right] \\ & \quad \times P(G_t \geq \frac{w_{j\ell}}{q_{j\ell}}, t \in I_{j\ell}, j=1, 2, \dots, k, \ell=1, 2, \dots, m | \{I_{j\ell}\}) \end{aligned}$$

$$\begin{aligned}
&= \sum \left[ \frac{(n-k)!}{\prod_{j=1}^k \prod_{\ell=1}^m n_{j\ell}!} \right] \left[ \prod_{j=1}^k \prod_{\ell=1}^m q_{j\ell}^{n_{j\ell}} \right] \left[ \prod_{j=1}^k \prod_{\ell=1}^m \left(1 - \frac{w_{j\ell}}{q_{j\ell}}\right)^{n_{j\ell}} \right] \\
&= \left(1 - \sum_{j=1}^k \sum_{\ell=1}^m w_{j\ell}\right)^{n-k}, \tag{2.7}
\end{aligned}$$

where the summation unspecified in (2.7) runs over all possibilities that  $\sum_{j=1}^k \sum_{\ell=1}^m n_{j\ell} = n - k$ ,  $n_{j\ell} \geq 0$ , integers. The proof of lemma 5 is completed.

Lemma 6. Let  $\phi$  be an a.e. positive function defined on the line and

$$H(q_1, \dots, q_k) = \int_0^{q_1} \dots \int_0^{q_k} \phi(w_1 + \dots + w_k) dw_1 \dots dw_k.$$

Then we have

$$\begin{aligned}
\sup_{q_1 + \dots + q_k = kq} H(q_1, \dots, q_k) &= H(q, \dots, q), \\
q_1 \geq 0, \dots, q_k \geq 0 &
\end{aligned} \tag{2.8}$$

here  $q > 0$  be a given number.

Proof. Note that  $D = \{(q_1, \dots, q_k), q_1 + \dots + q_k = kq, q_1 \geq 0, \dots, q_k \geq 0\}$  is a bounded and closed set and that  $H(q_1, \dots, q_k)$  is continuous on  $D$ . Hence there exists a point  $q_1^0, \dots, q_k^0 \in D$  such that

$$\sup_D H(q_1, \dots, q_k) = H(q_1^0, \dots, q_k^0).$$

To prove (2.8), we only need to prove  $q_1^0 = q_2^0 = \dots = q_k^0$ . By symmetry in  $q_1^0, \dots, q_k^0$ , we only need to prove that  $q_1^0 = q_2^0$ . Let

$$\psi(x) = \int_0^{q_3^0} \dots \int_0^{q_k^0} \psi(x+w_3+\dots+w_k) dw_3 \dots dw_k.$$

Then

$$\begin{aligned} Q(q_1^0, q_2^0) &\triangleq H(q_1^0, q_2^0, \dots, q_k^0) = \int_0^{q_1^0} \int_0^{q_2^0} \psi(x+y) dx dy \\ &= \int_0^{q_1^0+q_2^0} \psi(u) \min(u, q_1^0, q_2^0, q_1^0+q_2^0-u) du \\ &\leq \int_0^{q_1^0+q_2^0} \psi(u) \min(u, q_1^0+q_2^0-u) du \\ &= \int_0^{\frac{q_1^0+q_2^0}{2}} \int_0^{\frac{q_1^0+q_2^0}{2}} \psi(x+y) dx dy = Q\left(\frac{q_1^0+q_2^0}{2}, \frac{q_1^0+q_2^0}{2}\right). \end{aligned} \quad (2.9)$$

The quality of (2.9) holds if and only if  $q_1^0 = q_2^0 = \frac{q_1^0+q_2^0}{2}$  because  $\psi(u)$  is positive. The proof of lemma 6 is completed.

Lemma 7 (Bennett, 1962, see Hoeffding (1963)). Let  $U_1, \dots, U_n$  be independent and let  $EU_i = 0$ ,  $\sigma_i^2 = EU_i^2$ ,  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$ . Suppose that  $|U_i| \leq b$ ,  $i = 1, 2, \dots, n$ . Then for any  $\epsilon > 0$ ,

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n U_i\right| \geq \epsilon\right) \leq 2 \exp\{-n\epsilon^2/2(\sigma^2+b\epsilon)\}.$$

Lemma 8 (Hoeffding, 1963). Let  $\xi \sim B(n, p)$ , the binomial distribution with parameter  $n$  and  $p$ . Then for any  $\epsilon > 0$ ,

$$P\left(\left|\frac{1}{n}\xi - p\right| \geq \epsilon\right) \leq 2\exp\{-n\epsilon^2/(2p+\epsilon)\}.$$

### 3. THE MAIN RESULT

Theorem 3.1. If  $F$ , the marginal distribution of  $X$ , has no atoms, then for any  $\epsilon > 0$ , there exist constants  $c$  and  $b$ , depending upon  $\epsilon$  and the distribution of  $(\theta, X)$ , such that

$$P(|L_n - R| \geq \epsilon) \leq ce^{-bn}. \quad (3.1)$$

Proof. Define

$$T_n = P(\theta_n \neq \theta | X_1, \dots, X_n) = E(L_n | X_1, \dots, X_n). \quad (3.2)$$

Using the notations defined in §1 and §2, we have

$$L_n = 1 - \sum_{j=1}^n \sum_{i=1}^s I(\theta_j = i) \int_{V_{nj}} P_i(x) F(dx) \quad (3.3)$$

and

$$T_n = 1 - \sum_{j=1}^n \sum_{i=1}^s P_i(X_j) \int_{V_{nj}} P_i(x) F(dx) \quad (3.4)$$

Hence

$$\begin{aligned} & P(|L_n - T_n| \geq 2s\epsilon') \\ & \leq \sum_{i=1}^s P\left(\left|\sum_{j=1}^n (I(\theta_j = i) - P_i(X_j))\right| \int_{V_{nj}} P_i(x) F(dx) \geq 2\epsilon'\right) \end{aligned} \quad (3.5)$$



Define

$$Q_{nj}^{(i)} \triangleq \int_{V_{nj}} P_i(x) F(dx) \leq U_{nj} = \int_{V_{nj}} F(dx). \quad (3.6)$$

$$\phi_n = \{j, j \leq n, U_{nj} \leq \Delta/n\}, \quad \phi_n^c = \{1, 2, \dots, n\} \setminus \phi_n \quad (3.7)$$

and

$$k = [\delta n] \quad (3.8)$$

where  $\Delta$  and  $\delta$  are positive constants to be specified later,

and  $[x]$  denotes the largest integer less than or equal to  $x$ .

By lemma 5 and lemma 6, we have

$$\begin{aligned} P\left(\sum_{j=1}^k U_{nj} \geq \varepsilon' \mid X_1, \dots, X_k\right) &\leq P\left(\sum_{j=1}^k \sum_{\ell=1}^m \xi_{j\ell} \geq \varepsilon'\right) \\ &= \frac{(n-k)!}{(n-k-mk)!} \int_0^1 \dots \int_0^1 I_D \left(1 - \sum_{j=1}^k \sum_{\ell=1}^m w_{j\ell}\right)^{n-k-mk} \prod_{j=1}^k \prod_{\ell=1}^m dw_{j\ell} \\ &\leq \frac{(n-k)!}{(n-k-mk)!} \int_0^{\frac{1}{mk}} \dots \int_0^{\frac{1}{mk}} I_D \left(1 - \sum_{j=1}^k \sum_{\ell=1}^m w_{j\ell}\right)^{n-k-mk} \prod_{j=1}^k \prod_{\ell=1}^m dw_{j\ell} \\ &\leq \frac{(n-k)!}{(n-k-mk)!} \left(\frac{1}{mk}\right)^{mk} \exp\{-\varepsilon'(n-k-mk)\}, \end{aligned} \quad (3.9)$$

where  $D = \{w_{ij}, j=1, 2, \dots, k, \ell=1, 2, \dots, m, \sum_{j=1}^k \sum_{\ell=1}^m w_{j\ell} \geq \varepsilon'\}$  and

$I_D$  is the indicator of the set  $D$ .

Let

$$E_n = \left\{ \max_{1 \leq j_i < \dots < j_k \leq n} \sum_{i=1}^k U_{nj_i} \geq \varepsilon' \right\} \quad (3.10)$$

We have from (3.9)

$$\begin{aligned}
 P(E'_n) &\leq \sum_{1 \leq j_1 < \dots < j_k \leq n} P\left(\sum_{i=1}^k U_{nj_i} \geq \varepsilon'\right) \\
 &= \binom{n}{k} P\left(\sum_{j=1}^k U_{nj} \geq \varepsilon'\right) = \binom{n}{k} EP\left(\sum_{j=1}^k U_{nj} \geq \varepsilon' \mid X_1, \dots, X_k\right) \\
 &\leq \binom{n}{k} \frac{(n-k)!}{(n-k-mk)!} \left(\frac{1}{mk}\right)^{mk} \exp\{-\varepsilon'(n-k-mk)\}. \quad (3.11)
 \end{aligned}$$

Applying Stirling's formula and recalling  $k = [\delta n]$ , we have

$$\begin{aligned}
 P(E'_n) &\leq \frac{n!}{k! (mk)! (n-mk-k)!} \exp\{-\varepsilon'(1-\delta(m+1))n\} \\
 &\leq [\delta^{-\delta} (m\delta)^{-m\delta} (1-(m+1)\delta)^{-(1-(m+1)\delta)}]^{n} \exp\{-\varepsilon'(1-(m+1)\delta)n\} \\
 &\leq \exp\{-bn\}, \quad (3.12)
 \end{aligned}$$

where  $b = \frac{1}{2} \varepsilon' (1-(m+1)\delta)$  and  $\delta \in (0, \min(\varepsilon', \frac{1}{m+1}))$ , being such that

$$\log[\delta^{-\delta} (m\delta)^{-m\delta} (1-(m+1)\delta)^{-(1-(m+1)\delta)}] < \frac{1}{2} \varepsilon' (1-(m+1)\delta). \quad (3.13)$$

Here the reason that such  $\delta$  can be chosen is based upon the fact that when  $\delta \rightarrow 0$ , the left hand side of the above inequality tends to zero and the right hand side to  $\frac{1}{2} \varepsilon' > 0$ .

If we choose  $\Delta > \frac{1}{\delta}$ , then  $\#\{\phi_n^C\} \leq k$ , hence we have

$$\begin{aligned}
& P\left(\left|\sum_{j \in \phi_n^c} (I(\theta_j=i) - P_i(X_j))\right| \int_{V_{nj}} P_i(x) F(dx) \geq \varepsilon'\right) \\
& \leq P\left(\sum_{j \in \phi_n^c} U_{nj} \geq \varepsilon'\right) \\
& \leq P(E_n) \leq \exp\{-bn\}. \tag{3.14}
\end{aligned}$$

Hence

$$\begin{aligned}
& P\left(\left|\sum_{j=1}^n (I(\theta_j=i) - P_i(X_j))\right| \int_{V_{nj}} P_i(x) F(dx) \geq 2\varepsilon'\right) \\
& \leq e^{-bn} + P\left(\left|\sum_{j \in \phi_n} (I(\theta_j=i) - P_i(X_j)) Q_{nj}\right| \geq \varepsilon'\right) \\
& = e^{-bn} + EP\left(\left|\sum_{j \in \phi_n} (I(\theta_j=i) - P_i(X_j)) Q_{nj}\right| \geq \varepsilon' \mid X_1, \dots, X_n\right). \tag{3.15}
\end{aligned}$$

Note that when  $X_1, \dots, X_n$  are given,  $I(\theta_j=i)$ ,  $j = 1, 2, \dots, n$  are conditionally independent and  $P_i(X_j)$ ,  $Q_{nj}$  and  $\phi_n$  depend only upon  $X_1, \dots, X_n$ . Hence applying lemma 7, we have

$$\begin{aligned}
& P\left(\left|\sum_{j \in \phi_n} (I(\theta_j=i) - P_i(X_j)) Q_{nj}\right| \geq \varepsilon' \mid X_1, \dots, X_n\right) \\
& = P\left(\left|\frac{1}{n_*} \sum_{j \in \phi_n} (I(\theta_j=i) - P_i(X_j)) Q_{nj}\right| \geq \frac{\varepsilon'}{n_*} \mid X_1, \dots, X_n\right) \\
& \leq 2 \exp\left\{-n_* \left(\frac{\varepsilon'}{n_*}\right)^2 / 2 \left(\frac{1}{n_*} \sum_{j \in \phi_n} Q_{nj}^2 + \frac{\varepsilon'}{n_*} \max_{j \in \phi_n} Q_{nj}\right)\right\} \\
& \leq 2 \exp\left\{-\varepsilon' / 2(1+\varepsilon') \max_{j \in \phi_n} Q_{nj}\right\} \\
& \leq 2 \exp\left\{-(\varepsilon' / 2(1+\varepsilon') \Delta) n\right\} = 2e^{-bn}, \tag{3.16}
\end{aligned}$$

where  $n_* = \#\{\bar{\phi}_n\}$ , and  $b$  is a positive constant independent of  $n$ , but in its different appearance, it may take different value, e.g. this  $b$  is different from that  $b$  in (3.13).

From (3.5), (3.15) and (3.16), it follows that

$$P(|L_n - T_n| \geq 2s\varepsilon') \leq 4s \exp\{-bn\}. \quad (3.17)$$

Finally, we shall establish the similar bound for  $P(|T_n - R| \geq 3s\varepsilon')$ . According to Lusin's Theorem, for each  $i$ ,  $i = 1, 2, \dots, s$ , there exists a continuous function  $\tilde{P}_i(x)$  satisfying the following two conditions

$$a) \quad 0 \leq \tilde{P}_i(x) \leq \sup_x P_i(x) \leq 1, \quad (3.18)$$

$$b) \quad F(A) < \delta/2, \quad A = \{x \in R^d, P_i(x) \neq \tilde{P}_i(x)\}, \quad (3.19)$$

where  $\delta$  is the constant determined by inequality (3.13).

Define

$$\Psi_n = \{j \leq n, X_j \in A\}.$$

Then we have

$$\begin{aligned} & P\left(\left|\sum_{j=1}^n P_i(X_j) Q_{nj} - \sum_{j=1}^n \tilde{P}_i(X_j) Q_{nj}\right| \geq \varepsilon'\right) \\ & \leq P\left(\sum_{j \in \Psi_n} Q_{nj} \geq \varepsilon'\right) \leq P\left(\sum_{j \in \Psi_n} U_{nj} \geq \varepsilon'\right) \\ & \leq P(\#\{\Psi_n\} \geq \delta n) + P(E_n). \end{aligned} \quad (3.20)$$

By Hoeffding's inequality (see Hoeffding (1963)), we have

$$\begin{aligned} P(\#\{\bar{\Psi}_n\} \geq \delta n) &\leq 2 \exp\{-n(\delta - F(A))^2 / (2F(A) + (\delta - F(A)))\} \\ &\leq 2 \exp\{-\left(\frac{\delta}{6}\right)n\} = 2 \exp\{-bn\}. \end{aligned} \quad (3.21)$$

(3.21), together with (3.14), yields

$$P\left(\left| \sum_{j=1}^n P_i(X_j) Q_{nj} - \sum_{j=1}^n \tilde{P}_i(X_j) Q_{nj} \right| \geq \varepsilon'\right) \leq 3 \exp\{-bn\}. \quad (3.22)$$

Define

$$\tilde{H}_i = \int_{\mathbb{R}^d} \tilde{P}_i(x) P_i(x) F(dx) \quad (3.23)$$

and

$$H_i = \int_{\mathbb{R}^d} P_i^2(x) F(dx). \quad (3.24)$$

It is obvious that

$$|\tilde{H}_i - H_i| \leq F(A) < \delta/2 < \varepsilon'. \quad (3.25)$$

On the other hand, using the same approach as in Fritz (1975), we can prove that

$$P\left(\left| \sum_{j=1}^n \tilde{P}_i(X_j) - \tilde{H}_i \right| \geq \varepsilon'\right) \leq \frac{6}{\varepsilon'} e^{-bn} \quad (3.26)$$

From (3.22), (3.25) and (3.26), it follows that

$$P\left(\left| \sum_{j=1}^n P_i(X_j) Q_{nj} - H_i \right| \geq 3\varepsilon'\right) \leq \left(3 + \frac{6}{\varepsilon'}\right) e^{-bn}. \quad (3.27)$$

Recalling that  $R = 1 - \sum_{i=1}^S H_i$  we obtain that

$$P(|T_n - R| \geq 3S\varepsilon') \leq S(3 + \frac{6}{\varepsilon'})e^{-bn} \quad (3.28)$$

(3.17) and (3.28) yield

$$P(|L_n - R| \geq 5S\varepsilon') \leq S(7 + \frac{6}{\varepsilon'})e^{-bn}, \quad (3.29)$$

which, together with taking  $\varepsilon' = \varepsilon/5S$ , completes the proof of Theorem 3.1.

#### REFERENCES

- [1] Bai, Z.D. (1984). The Strong Consistency of Error Probability Estimates in NN Discrimination. CMA., Univ. of Pitt., Tech. Report 84-43.
- [2] Chen, G.J. and Kong, F.C. (1983). Sufficient and Necessary Condition for Convergence of Conditional Error Probability in NN Discrimination. Submitted by Journal of Mathematical Research and Exposition.
- [3] Bai, Z.D., Chen, X.R. and Chen, G.J. Some Results on Nearest Neighbor Discrimination by Treating Ties Randomly. Submitted by Journal of Mathematical Research and Exposition.
- [4] Cover, T.M. (1968). Estimation by the Nearest Neighbor Rule. IEEE Trans. Inform. Theory 14, 50-55.
- [5] Cover, T.M. and Hart, P.E. (1967). Nearest Neighbor Pattern Classification. IEEE Trans. Inform. Theory 13, 21-27.
- [6] Devroye, L. (1981). On the Asymptotic Probability of Error in Nonparametric Discrimination. Ann. of Statistics, Vol. 9, No. 6, 1320-1327.

- [7] Fritz, J. (1975). Distribution-Free Exponential Error Bound for Nearest Neighbor Pattern Classification. *IEEE Trans. Inform. Theory*, 21, 552-557.
- [8] Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* 58, 13-30.
- [9] Wagner, T.J. (1971). Convergence of Nearest Neighbor Rule. *IEEE Trans. Inform. Theory*, 17, 566-571.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>AFOSR-TR-85-082</b>	2. GOVT ACCESSION NO. <b>AD-A160305</b>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  Exponential Bound for Error Probability in NN-Discrimination		5. TYPE OF REPORT & PERIOD COVERED Technical - April 1985
		6. PERFORMING ORG. REPORT NUMBER 85-15
7. AUTHOR(s) Author - Z.D. Bai Center for Multivariate Analysis 515 Thackeray Hall University of Pittsburgh, PA 15260		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/AS
		12. REPORT DATE April 1985
11. CONTROLLING OFFICE NAME AND ADDRESS		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Let $(\theta_1, X_1), \dots, (\theta_n, X_n)$ be $n$ simple samples drawn from the population $(\theta, X)$ which is a $\{1, 2, \dots, s\} \times \mathbb{R}^d$ -valued random vector. Suppose that $X$ is known. Let $X'_n$ be the nearest one among $X_1, \dots, X_n$ , from $X$ , in the sense of Euclidean norm or $\ell_\infty$ -norm, and let $\theta'_n$ be the $\theta$ -value paired with $X'_n$ . The posterior		



error probability is defined by  $L_n = P(\theta'_n \neq \theta | (\theta_1, X_1), \dots, (\theta_n, X_n))$ . It is well known that  $EL_n = P(\theta'_n \neq \theta)$  has always a limit  $R$ . In this paper it is shown that for any  $\epsilon > 0$ , there exist constants  $c < \infty$  and  $b > 0$  such that  $P(|L_n - R| \geq \epsilon) \leq ce^{-bn}$ , under the only assumption that the marginal distribution of  $X$  is nonatomic.

Unclassified

**END**

**FILMED**

**11-85**

**DTIC**