ISBN 0-315-55575-0

THE UNIVERSITY OF ALBERTA

NEAREST NEIGHBOR CLASSIFICATION RULES FOR MULTIPLE OBSERVATIONS

BY

SUBHASH CHANDRA BAGUI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

FALL 1989

THE UNIVERSITY OF ALBERTA

RELEASE FORM

Name of Author Subhash Chandra Bagui

TITLE OF THESIS Nearest Neighbor Classification Rules for Multiple Observations

DEGREE FOR WHICH THESIS WAS PRESENTED Doctor of Philosophy

YEAR OF DEGREE GRANTED Fall 1989

(Signed) Subhash Chandra Bagui

Permanent Address:

P.O. - Paschim Tajpur
Dt. - Hooghly, W.B.
INDIA, Pin - 712706

DATED Oct. 6, 1989

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Nearest Neighbor Classification Ruled for Multiple Observations submitted by Subhash C. Bagui in partial fulfilment of the requirements for the degree of Doctor of Philosophy

K. L. Mehra
Supervisor

Peter Hooper

Jerome N. Sheahan

J. D. Jobson

M. Saïd Ahmed Raza

M. D. Burke
External Examiner

Date: October 6, 1989

Dedicated to

my late father Paresh Chandra Bagui

and

my mother Kohden Bala Bagui

ABSTRACT

In this thesis, a number of nearest neighbor (NN) type decision rules are proposed for classifying a (small) set of K (≥ 2) multiple observations (that is, independent and identically distributed, (i.i.d.)) into one of s (≥ 2) given populations. The proposed rules are generalizations of certain known NN rules for the case $K = 1$. The rules are based on the information contained in the (available) "training" (i.i.d.) samples from the s populations under consideration. The asymptotic risk functions of the proposed classification rules, as "training" sample sizes tend to infinity, are derived and appropriate bounds for them are studied.

In Chapter 1, the relevant literature on NN classification is reviewed. In Chapter 2, we consider the case $s = 2$ and propose NN classification rules for K multiple multivariate observtions in both the (empirical) Bayes as well as the nonparametric formulations of the problem. It is shown that the asymptotic risks of the proposed NN rules in the above two models are identical. Also, for the case $K = 2$, the asymptotic risk is shown to coincide with that of the 1-NN rule of Cover and Hart (1967). In Chapter 3, we consider again the case $s = 2$, but for K (≥ 2) multiple univariate observations using a two sided Rank nearest neighbor (RNN) rule and derive the asymptotic risks of the proposed so called "M-stage" RNN rule. For the case $K = 2$ and $M = 1$, we derive an upper bound on the risk when the sample sizes are equal. An estimate of the asymptotic risk of first-stage RNN rule is shown to be asymptotically unbiased and consistent. In Chapter 4, we extend the above NN and RNN rules of Chapter 2 and Chapter 3, respectively, to the case $s > 2$. In Chapter 5, we develop the sub-sample approach, namely, that based on all possible sub-samples of size K from the "training" samples. The asymptotic risk of the proposed NN rule in this case is shown to have bounds that are parallel to those obtained by Cover and Hart (1967) for the case $K = 1$. Finally, in Chapter 6, some Monte-Carlo results are given to compare the expected performance of the proposed rules in the small sample situations.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Professor K.L. Mehra for his help, advice and guidance. I am indebted to Professor M.S. Rao for fruitful discussions and helpfulness. My thanks are due to Dr. J.N. Sheahan for his encouragement and help throughout the past year. I appreciate the finanacial support of Professor R.J. Elliott and Dr. D. Wiens towards the typing of the thesis. I also wish to express my gratitude to a fellow student, Mr. D. Rakshit, for his assistance in writing a computer programme. I am most grateful to my family for their continuing love and support.

Last but not the least, I am thankful to Mrs. Chrinstine Fischer and Mrs. Laura Heiland for their professional typing to fthe thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The object of the present thesis is to investigate suitable nonparametric procedures for the classification of an observed random sample into one of two or more populations. We shall, however, limit our investigations to "Nearest Neighbor" (NN) type procedures only, and generalize certain well known single observation "Nearest Neighbor" and "Rank Nearest Neighbor" (RNN) procedures for the classification of multiple independently and identically distributed (i.i.d.) observations.

The basic problem in Discriminant Analysis is to devise appropriate procedures for classifying one or more observations into one of two or more distinct populations on the basis of their values. Specifically, let $\underset{\sim}{Z} = \{Z_1, \ldots, Z_K\} (K \geq 1)$ with $Z_i = (Z_{i1}, \ldots, Z_{id})', i = 1, \ldots, K$, denote a sample of independent and identically distributed d-variate observations from one of the s-populations $\pi_1, \pi_2, \ldots, \pi_s$. The object is to correctly classify $\underset{\sim}{Z}$ into its parent population. In this situation there are two extremes of knowledge which a Statistician may possess: either he may have complete statistical knowledge about the underlying distributions of these populations or he may have no knowledge at all beyond some broad information regarding some general properties of these populations along with identified samples (i.e. training samples) from them. In the first situation an optimum solution, derived by Welch (1939), is implicit in the Neyman-Pearson lemma. In the other extreme, the problem is in the domain of nonparametric Statistics and, quite naturally, no optimal solution exists in this case with respect to all underlying distributions. For the case $K = 1$, that is, when there is only a single observation to be classified, there have been many nonparametric classification procedures proposed in the literature, such as (i) Nearest Neighbor (NN) rules (Fix and Hodges (1951)); (ii) Rules based on density estimates (Van Ryzin (1966)); (iii) Rules based on distance between Empirical c.d.f's (Matusita (1956)), (iv) An Empirical Bayes approach (Johns (1961)), (v) Rules based on ranks (Das Gupta (1964)) (vi) Rules based on tolerance regions (Quesenberry and Gessaman (1968)), etc. (See also Broffitt et. al. (1976), Cacoullos (1973), Cover (1968, 1969), Das Gupta (1962, 1973), Lachenbruch (1975), Peterson (1970) and Randles et. al. (1978)).

While most of the work done in parametric or nonparametric Discriminant Analysis is on the classification of a single observation, the more general problem of classifying a set of multiple observations, assumed to be originating from the same population, has been virtually ignored in the literature. The simplest, and a rather naive, way to classify these multiple observations would be to repeat K times one of the above single observation classification procedures. The drawback of this "repeating" procedure is that it does not utilize the full information available in the data. It can be seen easily (see (2.41) below) that any such procedure would be inevitably inadmissible. Das Gupta (1964) considered the problem of classifying a random sample based on Kolmogorov's distance and Wilcoxon's statistic and showed the consistency of the rules proposed. However, these rules are only appropriate when both the size of the random sample to be classified and the sizes of the training samples are moderately large.

The NN rule for classifying a single observation Z into one of two populations π_1 and π_2 was first introduced by Fix and Hodges (1951). Their proposed k-NN rule may be described as follows:
Let $\{X_{11}, \ldots, X_{n_1}\}$ and $\{X_{21}, \ldots, X_{2n_2}\}$ be the random samples from π_1 and π_2 respectively. Using a distance function $d(X_{ij}, Z)$, order the observed distances $d(x_{ij}, z)$ for $i = 1, 2$ and $j = 1, 2, \ldots, n_i$. For a fixed integer k, the k-NN rule assigns Z to π_1 if $\frac{k_1}{n_1} > \frac{k_2}{n_2}$, where k_i is the number of observations from $\pi_i, i = 1, 2$, among k observations nearest to Z. For $n_1 = n_2 = n$, they showed that the probability of misclassification (PMC) of the NN-rule tends to the PMC of the maximum likelihood rule as $n \to \infty, k \equiv k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$ (for details see Fix and Hodges, 1951). In 1967, Cover and Hart studied an 1-NN rule, again for classifying a single observation, based on an identified training sample from a mixture of s-populations $\pi_1, \ldots, \pi_s$. Their model follows the Bayesian set up

and the proposed 1-NN rule assigns Z to π_i if the NN (measured by a distance function d) of Z belongs to the i-th population $\pi_i, i = 1, \ldots, s$. They were the first to obtain a bound on the NN risk $R_{(s)}(1)$ (i.e. the total probability of misclassification (TPMC)) of the type $R^*_{(s)}(1) \leq R_{(s)}(1) \leq R^*_{(s)}(1)\left(2 - \frac{s}{s-1}R^*_{(s)}(1)\right)$ where $R^*_{(s)}(1)$ is the Bayes risk (see (2.9)). In 1966, Anderson suggested a nonparametric procedure in the case of two populations based on separate training samples. The procedure may be described as follows: Let $\{X_{11}, \ldots, X_{1n_1}\}$ and $\{X_{21}, \ldots, X_{2n_2}\}$ be the training samples from π_1 and π_2 and Z be the single observation to be classified between π_1 and π_2. Combine the X_{1i}'s X_{2j}'s and Z and rank them in increasing order. (i) If the Z is either the smallest or largest, classify it to the class of its Nearest Neighbor, and (ii) if both the left and right hand Nearest Neighbors are of same kind, classify Z to that class. (iii) if Z falls between two different types of observations, classify it into either of the two classes with probability 1/2 and 1/2. Das Gupta and Lin (1980) studied the above rule and termed it a Rank Nearest Neighbor (RNN) rule. They derived the asymptotic total probability of misclassification (TPMC) which turned out to be exactly the same as the limiting risk obtained by Cover and Hart (1967) with their 1-NN rule. They also considered a Multistage (M-stage) version of the RNN rule and showed that the asymptotic TPMC of the M-stage RNN rule decreases as the stage M increases. They suggested estimates based on "runs" to estimate the PMC's of the one-stage RNN rule and showed that these estimates are asymptotically unbiased. Das Gupta and Lin's procedure is especially useful when the observations are available only in terms of their ranks. Wagner (1971) showed that the conditional risk of the 1-NN rule of Cover and Hart (1967), given the identified training sample of n observations, denoted by L_n, converges to $R_{(2)}(1)$, the asymptotic NN risk, with probability 1 under certain mild continuity and moment conditions on the assumed densities. Also an estimate of $R_{(2)}(1)$ was suggested and was shown to be consistent. Fritz in 1975, improved Wagner's result by replacing the assumption regarding the existence and continuity of the density with non-atomicity of the underlying probability distributions. Fritz considered the case of general s and $k = 1$, and showed that $\lim_{n\to\infty} L_n = R_{(s)}(1)$ a.s.. In fact, for any given $\varepsilon > 0$, he established the exponential bound $P(|L_n - R_{(s)}(1)| \geq \varepsilon) = O(e^{-c\sqrt{n}})$. Fritz also indicated that when $k > 1$, the expression for L_n becomes far more complicated. For $k > 1$, Xiru (1985) noticed that L_n is a kind of "weighted U-statistic" and for $s = 2$ and k odd, he showed that $\lim_{n\to\infty} L_n = R_{(2)}(k)$ a.s. still holds by establishing again an exponential bound for $P(|L_n - R_{(2)}(k)| \geq \epsilon)$. In the meantime Devroye (1981) extended the result of Cover and Hart (1967) from the 1-NN to the k-NN rule and obtained the following bound on the asymptotic k-NN risk $R_k(1)$: $R_k(1) \leq (1 + a_k)R^*_{(2)}(1)$, $a_k = \frac{\alpha\sqrt{k}}{k-3.25}(1 + \frac{\alpha}{\sqrt{k-3}})$, k odd, $k \geq 5$ where $R^*_{(2)}(1)$ is the Bayes risk (see (2.10)) and $\alpha = 0.3399\ldots$ and $\beta = 0.074\ldots$ are universal constants. The bound is the best possible in a "certain" sense (see Devroye (1981)). Now we give a brief summary of our results.

A brief summary of the results:

In chapter two, we propose a classification procedure of the Nearest Neighbor type for multiple observations, both in the (empirical) Bayesian and the nonparametric framework. First we consider the problem in Bayesian formulation, where the training sample is from a mixture of populations; then we consider a nonparametric model where the training samples have been drawn separately from the populations. In both cases, we arrive at the same asymptotic risk when observations are classified between two populations. The asymptotic risk for $K = 2$ also coincides with the respective asymptotic risk of the 1-NN rule of Cover & Hart (1967).

Chapter 3 deals with the classification of Multiple univariate observations between two populations using the RNN rule. We obtain asymptotic total probability of misclassification (TPMC) of the proposed first-stage RNN rule and derive an upper bound for it when the sample sizes are equal. Asymptotic TPMC of the M-stage RNN rule is also derived. An estimate of the asymptotic TPMC of first-stage RNN rule is shown to be asymptotically unbiased and consistent.

In chapter 4, we extend our proposed NN and RNN type classification rules of Chapter 2 and Chapter 3 respectively to general s populations. First in Section 4.1, we consider the classification of a single univariate observation ($K = 1$) to one among s univariate populations using RNN rule, i.e. it is an extension of Das Gupta and Lin (1980) single observation RNN classification procedure to s populations. Here we derive its asymptotic TPMC, which turns out to be exactly the same as the limiting risk of the 1-NN rule of Cover and Hart (1967) for s populations. We also derive the asymptotic TPMC of the corresponding multi-stage procedure. A simple estimate of the asymptotic TPMC of the first-stage RNN is proposed and it is shown to be asymptotically unbiased and consistent. In Section 4.2, for general K and s, we propose a classification procedure using the first-stage RNN rule in the univariate case and indicate how to get the limiting TPMC of the proposed rule. Finally in Section 4.3, we propose a classification procedure for Multivariate observations for general K and s. For $K = 2$ and general s we obtain its limiting risk which turns out to be exactly the same as that of Cover and Hart (1967) for general s.

We also consider in Chapter 5, a 1-NN classification rule for classifying Multiple observations based on sub-groupings of size K from the training samples. We derive its asymptotic risk $R_{(s)}(K)$ for the general case of K observations and s populations. The bounds obtained on $R_{(s)}(K)$ come out to be parallel to the bounds derived by Cover and Hart (1967) for the case $K = 1$. An estimate $R_{(2)}(K)$ is also considered and shown to be asymptotically unbiased and consistent.

Finally in Chapter 6, we report the results of a Monte Carlo simulation study to examine the performance of our proposed rules in small sample situations. In section 6.2, a comparison is made between the Cover and Hart procedure and the Das Gupta and Lin procedure for certain different pairs of distributions. The performance of the 1st-stage RNN rule is studied in Section 6.3. We calculate the average proportion of misclassification $R^{(1)}_{(2)}(K)$ for each increasing value of K. Interestingly, we find that $R^{(1)}_{(2)}(K)$ decreases with the increasing value of K.

Finally in Section 6.4, we make a small sample comparison between the 1st-stage RNN and the sub-sample (SS) procedure. In the small sample situations the sub-sample procedure seems to perform very well in most cases.

CHAPTER 2

ONE-NEAREST NEIGHBOR CLASSIFICATION RULES FOR MULTIPLE OBSERVATIONS BETWEEN TWO POPULATIONS.

2.0 **Introduction.** In this chapter a simple Nearest Neighbor (NN) decision rule is proposed for classifying an independent, identically distributed (i.i.d.) sample of K ($K > 1$) observations into one of two populations. The asymptotic average probability of misclassification (PMC) $R_{(2)}(K)$, also called average risk, is derived for the proposed classification rule, both in the Bayesian as well as the nonparametric formulations. For $K = 2$, we obtain bounds on $R_{(2)}(K)$ of the type $R^*_{(2)}(1) \leq R_{(2)}(2) \leq 2R^*_{(2)}(1)(1-R^*_{(2)}(1))$ in both Baysian and nonparametric frameworks, where $R^*_{(2)}(1)$ is the Bayes risk (see (2.10)) corresponding to the single observation classification rule. We make a conjecture regarding the bounds on $R_{(2)}(K)$ for $K \geq 3$ and an example is given in support of the proposed bounds. First we describe the preliminaries of the Bayesian model.

Let $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ be i.i.d. random variables taking value in $\mathbb{R}^d \times \{1, 2, \ldots, s\}$, where $\{\theta_i\}, i = 1, \ldots, n$ are i.i.d. discrete random variables with $P(\theta_i = j) = \xi_j, i = 1, \ldots, n, j = 1, \ldots, s$ and $\sum_{j=1}^{s} \xi_j = 1$. The ξ_j's are called the prior probabilities associated with populations $\pi_j, j = 1, \ldots, s$. If $\theta_i = j$, we say that X_i is from the population π_j with distribution function F_j which is assumed to possess a probability density function (p.d.f.) f_j with respect to Lebesgue measure μ. Let $f_{\theta_i}(x)$ denote the conditional probability density of X_i given $\theta_i, i = 1, \ldots, n, -\infty < x < \infty$. Throughout the thesis, we assume $f_j, j = 1, 2, \ldots, s$ are continuous. Assume now that we have $(Z_1, \theta_1), (Z_2, \theta_2), \ldots, (Z_K, \theta_K)$, K i.i.d. observations taking values in $\mathbb{R}^d \times \{1, 2, \ldots, s\}$, where only $Z_1, Z_2, \ldots, Z_K$ are observable and we assume that $\theta_1 = \ldots = \theta_K = \theta$ (say). The object is to classify $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ into one among $\pi_1, \pi_2, \ldots, \pi_s$. For this, we need to estimate θ using the information contained in the "identified" (training) sample (X_i, θ_i) $i = 1, \ldots, n$. By "identified" sample we mean that the value of each θ_i is assumed to be correctly given or known. Let θ'_n be an estimate of θ and

$L(\theta, \theta'_n)$ denote the loss incurred in estimating θ with θ'_n. We write the posterior probabilities given $Z_\ell = z_\ell$ by Bayes theorem

$$\begin{aligned}\eta_i(z_\ell) &= P(\theta = i / Z_\ell = z_\ell) \\ &= \frac{\xi_i f_i(z_\ell)}{\sum_{i=1}^{s} \xi_i f_i(z_\ell)},\end{aligned} \tag{2.1}$$

$i = 1, 2, \ldots, s; \ell = 1, 2, \ldots, K$, and for given $\underset{\sim}{Z} = \underset{\sim}{z} = (z_1, z_2, \ldots, z_K)$,

$$\eta_i^*(\underset{\sim}{z}) = P(\theta = i / \underset{\sim}{Z} = \underset{\sim}{z}) = \frac{\xi_i \prod_{\ell=1}^{K} f_i(z_\ell)}{\sum_{i=1}^{s} \xi_i \prod_{\ell=1}^{K} f_i(z_\ell)}. \tag{2.2}$$

If we decide to classify $\underset{\sim}{Z}$ to the population j, then the conditional risk given $\underset{\sim}{Z} = \underset{\sim}{z}$ is given by

$$\begin{aligned}r_j(\underset{\sim}{z}) &= E(L(\theta, \theta'_n) / \underset{\sim}{Z} = \underset{\sim}{z}) \\ &= \sum_{i=1}^{s} \eta_i^*(\underset{\sim}{z}) L(i, j).\end{aligned} \tag{2.3}$$

Bayes decision rule δ^* chooses the population π_j for which $r_j(\underset{\sim}{z})$ in (2.3) is minimum. Using δ^*, the conditional Bayes risk $r^*(\underset{\sim}{z})$ can be written as

$$r^*(\underset{\sim}{z}) = \min_j \{\sum_{i=1}^{s} \eta_i^*(\underset{\sim}{z}) L(i, j)\}. \tag{2.4}$$

The overall minimum expected risk $R^*_{(s)}(K)$, called the Bayes risk, is given by

$$R^*_{(s)}(K) = E r^*(\underset{\sim}{z}) = \int r^*(\underset{\sim}{z}) f(\underset{\sim}{z}) d\underset{\sim}{z}, \tag{2.5}$$

where

$$f(\underset{\sim}{z}) = \sum_{i=1}^{s} \xi_i \prod_{\ell=1}^{K} f_i(z_\ell). \tag{2.6}$$

Throughout the thesis, we take our loss function $L(\theta, \theta'_n)$ as

$$\begin{aligned}L(\theta, \theta'_n) &= 0 \text{ if } \theta = \theta'_n \\ &= 1 \text{ if } \theta \neq \theta'_n.\end{aligned} \tag{2.7}$$

Using the (0-1) loss function (2.7) above, the expressions (2.4) and (2.5) reduce to the following expressions respectively:

$$r^*(\underset{\sim}{z}) = min(1-\eta_1^*(\underset{\sim}{z}),\ldots,1-\eta_s^*(\underset{\sim}{z})), \tag{2.8}$$

and

$$\begin{aligned} R^*_{(s)}(K) &= \int \min(1-\eta_1^*(\underset{\sim}{z}),\ldots,1-\eta_s^*(\underset{\sim}{z}))f(\underset{\sim}{z})d\,\underset{\sim}{z} \\ &= \int \min(\sum_{i\neq 1}\xi_i\prod_{\ell=1}^{K} f_i(z_\ell),\ldots,\sum_{i\neq s}\xi_i\prod_{\ell=1}^{K} f_i(z_\ell))d\,\underset{\sim}{z}\,. \end{aligned} \tag{2.9}$$

For $K = 1$ and $s = 2$, one may write (2.9) as

$$\begin{aligned} R^*_{(2)}(1) &= \int \min(\eta_1(z_1), 1-\eta_1(z_1))f(z_1)dz_1 \\ &= \int \min(\xi_1 f_1(z_1), \xi_2 f_2(z_1))dz_1. \end{aligned} \tag{2.10}$$

Suppose our proposed NN decision rule δ_n (say) classifies $\underset{\sim}{Z}$ to the population θ_n'. Then we define the NN risk $R_{(s)}(K;\delta_n)$ by

$$R_{(s)}(K;\delta_n) = EL(\theta,\theta_n'), \tag{2.11}$$

and the large sample NN risk by

$$R_{(s)}(K) = \lim_{n\to\infty} EL(\theta,\theta_n'). \tag{2.12}$$

We describe now the preliminaries of the nonparametric model. Let $X_{ij}, j = 1,2,\ldots,n_i$ be a training sample from the i-th population $\pi_i, i = 1,2,\ldots,s$. Let $\underset{\sim}{Z}= (Z_1,\ldots,Z_K)$ be a random sample from π_0 where it is known that $\pi_0 = \pi_i$ for exactly one i. The problem is to classify $\underset{\sim}{Z}$ to the correct population i for which $\pi_0 = \pi_i, i = 1,\ldots,s$. We shall denote the c.d.f. of population π_i by $F_i, i = 1,2,\ldots,s$. We shall assume that each F_i possesses a density $f_i, i = 1,2,\ldots,s$ with respect to the Lebesgue measure μ. Let $\alpha_{\ell j}, \ell \neq j = 1,2,\ldots,s$ be the probability of classifying $\underset{\sim}{Z}$ as being from π_ℓ when in fact, it is really from π_j. In notation we write it as

$$\alpha_{\ell j} = P(\text{ Decide } \underset{\sim}{Z}\in \pi_\ell / \underset{\sim}{Z}\in \pi_j). \tag{2.13}$$

Now we define the total probability of misclassification (TPMC) by

$$R'_{(s)}(K) = \sum_{\ell=1}^{s} \sum_{\substack{j=1 \\ \ell \neq j}}^{s} \xi_j \alpha_{\ell j}. \tag{2.14}$$

The quantities $R_{(s)}(K)$ and $R'_{(s)}(K)$ (limiting value) would be equivalent, if we use (0-1) loss function in (2.12).

For convenience of presentation we shall first consider the case for $K = 2$ and $s = 2$, then in section 2.2, we deal with the situation for general K and $s = 2$.

2.1 Asymptotic risk and its bounds in the Bayesian model when $K = 2$ observations are classified.

Let $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ be the identified training sample taking values in $\mathbb{R}^d \times \{1, 2\}$. Let (Z_1, θ_1) and (Z_2, θ_2) be a random sample taking values in $\mathbb{R}^d \times \{1, 2\}$ where it is known that $\theta_1 = \theta_2 = \theta$. Now to classify the pair $\underset{\sim}{Z} = (Z_1, Z_2)$ into π_1 or π_2, we need to estimate θ using the information contained in the preceding identified training sample.

We call $X'_{nj} \in \{X_1, X_2, \ldots, X_n\}$ a Nearest Neighbor (NN) of Z_j

$$\text{if} \quad \min_{i \in (1,2,\ldots,n)} \|X_i - Z_j\| = \|X'_{nj} - Z_j\|, \quad j = 1, 2. \tag{2.15}$$

Let X'_{n1} and X'_{n2} be the NN of Z_1 and Z_2, respectively, and suppose that X'_{n1} and X'_{n2} are identified as being from the category θ'_{n1} and θ'_{n2} respectively. We now define a Nearest Neighbor classification rule which we shall call 1-Nearest Neighbor (1-NN) rule and denote it by $\delta^{(2)}_{2B}$ (the subscript B signifies Bayesian framework):

$$\left.\begin{array}{l} \text{Classify } (Z_1, Z_2) \text{ to } \pi_j \text{ with probability one} \\ \text{if both the NN's of } Z_1 \text{ and } Z_2 \text{ are from } \pi_j, j = 1, 2; \text{ and} \\ \text{classify } (Z_1, Z_2) \text{ to } \pi_j, j = 1, 2 \text{ with probability each } 1/2 \\ \text{if the NN's of } Z_1 \text{ and } Z_2 \text{ are from different populations.} \end{array}\right\} \tag{2.16}$$

Now we state a lemma concerning the a.s. convergence of the NN X'_{nj} to $Z_j, j = 1, 2$.

LEMMA 2.1.1. *Let Z_1, Z_2 and $X_1, X_2, \ldots, X_n$ be the independent and identically distributed random variables. Suppose X'_{nj} is the NN of $Z_j, j = 1, 2$. Then*

$$X'_{nj} \longrightarrow Z_j \quad \text{a.s. as } n \to \infty.$$

PROOF. In view of continuity of f_j for $j = 1, 2$, we have for every $\varepsilon > 0$

$$\begin{aligned} P\{\|X'_{nj} - Z_j\| > \varepsilon\} &= P\{\min_i \|X_i - Z_j\| > \varepsilon\} \\ &= P\{\|X_1 - Z_j\| > \varepsilon, \|X_2 - Z_j\| > \varepsilon, \ldots, \|X_n - Z_j\| > \varepsilon\} \\ &= [P\{\|X_1 - Z_j\| > \varepsilon\}]^n \\ &= [1 - P\{\|X_1 - Z_j\| \leq \varepsilon\}]^n \\ &\to 0 \text{ as } n \to \infty. \end{aligned} \tag{2.17}$$

Since $\|X'_{nj} - Z_j\|$ is monotonically decreasing in n, it follows that

$$\lim_{n\to\infty} P[\cup_{k\geq n}\{\|X'_{kj} - Z_j\| > \varepsilon\}] = \lim_{n\to\infty} P[\|X'_{nj} - Z_j\| > \varepsilon] \tag{2.18}$$

so that by (2.17) we have

$$X'_{nj} \longrightarrow Z_j \quad \text{a.s. as } n \to \infty.$$

The proof is complete. □

Let θ'_n be an estimate of θ defined by (2.16). Then the conditional risk given $\underset{\sim}{Z} = \underset{\sim}{z} = (z_1, z_2)$ and $\underset{\sim}{X}'_n = \underset{\sim}{x}'_n = (x'_{n1}, x'_{n2})$ is given by

$$\begin{aligned} r(\underset{\sim}{z}; \underset{\sim}{x}'_n) &= E\{L(\theta, \theta'_n)/ \underset{\sim}{z}, \underset{\sim}{x}'_n\} \\ &= P(\theta \neq \theta'_n / \underset{\sim}{z}, \underset{\sim}{x}'_n) \\ &= P\{(\theta'_n = 1 \cap \theta = 2) \cup (\theta'_n = 2 \cap \theta = 1)/ \underset{\sim}{z}, \underset{\sim}{x}'_n\} \\ &= P(\theta'_n = 1 \cap \theta = 2/ \underset{\sim}{z}, \underset{\sim}{x}'_n) + P(\theta'_n = 2 \cap \theta = 1/ \underset{\sim}{z}, \underset{\sim}{x}'_n). \end{aligned}$$

In the above expression, using the conditional independence of θ and θ'_n, we get

$$\begin{aligned} r(\underset{\sim}{z};\underset{\sim}{x}'_n) &= P(\theta = 2/\underset{\sim}{z}) \cdot P(\theta'_n = 1/\underset{\sim}{x}'_n) + P(\theta = 1/\underset{\sim}{z}) \cdot P(\theta'_n = 2/\underset{\sim}{x}'_n) \\ &= \eta_2^*(\underset{\sim}{z}) \cdot \pi_1(\underset{\sim}{z};\underset{\sim}{x}'_n) + \eta_1^*(\underset{\sim}{z}) \cdot \pi_2(\underset{\sim}{z};\underset{\sim}{x}'_n), \end{aligned} \tag{2.19}$$

where $\eta_i^*(\underset{\sim}{z})$ for $K = 2$ is as given in (2.2),

$$\begin{aligned} \pi_1(\underset{\sim}{z};\underset{\sim}{x}'_n) &= P(\theta'_n = 1/\underset{\sim}{x}'_n) \\ &= \frac{\xi_1^2 f_1(x'_{n1}) f_1(x'_{n2})}{Bn} \\ &+ \frac{1}{2}\frac{\xi_1\xi_2(f_1(x'_{n1})f_2(x'_{n2}) + f_1(x'_{n2})f_2(x'_{n1}))}{Bn} \end{aligned} \tag{2.20}$$

with

$$\begin{aligned} Bn &= \xi_1^2 f_1(x'_{n1}) f_1(x'_{n2}) + \xi_2^2 f_2(x'_{n1}) f_2(x'_{n2}) \\ &\quad + \xi_1\xi_2(f_1(x'_{n1})f_2(x'_{n2}) + f_1(x'_{n2})f_2(x'_{n1})) \\ &= (\xi_1 f_1(x'_{n1}) + \xi_2 f_2(x'_{n1}))(\xi_1 f_1(x'_{n2}) + \xi_2(x'_{n2})), \\ &= \prod_{j=1}^{2}(\xi_1 f_1(x'_{nj}) + \xi_2 f_2(x'_{nj})), \end{aligned} \tag{2.21}$$

and $\pi_2(\underset{\sim}{z} : \underset{\sim}{x}'_n) = P(\theta'_n = 2/\underset{\sim}{x}'_n)$

$$\begin{aligned} &= \frac{\xi_2^2 f_2(x'_{n1}) f_2(x'_{n2})}{Bn} \\ &+ \frac{1}{2}\frac{\xi_1\xi_2(f_1(x'_{n1})f_2(x'_{n2}) + f_1(x'_{n2})f_1(x'_{n1}))}{Bn}. \end{aligned} \tag{2.22}$$

Now we prove a theorem on limiting NN risk.

THEOREM 2.1.1. *Suppose z_1 and z_2 are continuity points of both f_1 and f_2. Then the limiting conditional risk given $\underset{\sim}{Z} = (z_1, z_2)$ and the unconditional risk for the rule $\delta_{2B}^{(2)}$ are given respectively by*

$$\begin{aligned} r(\underset{\sim}{z}) &= \lim_{n\to\infty} r(\underset{\sim}{z};\underset{\sim}{x}'_n) \\ &= \eta_2^*(\underset{\sim}{z}) \cdot \pi_1(\underset{\sim}{z}) + \eta_1^*(\underset{\sim}{z}) \cdot \pi_2(\underset{\sim}{z}) \qquad \text{a.s.,} \end{aligned}$$

and

$$R_{(2)}(2) = \lim_{n\to\infty} E\{r(\underset{\sim}{z}; x'_n)\} = E\{r(\underset{\sim}{z})\},$$

where

$$\pi_1(\underset{\sim}{z}) = \frac{\xi_1^2 f_1(z_1) f_1(z_2)}{\prod_{i=1}^{2}(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))}$$

$$+ \frac{1}{2}\frac{\xi_1\xi_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^{2}(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))} \tag{2.23}$$

and

$$\pi_2(\underset{\sim}{z}) = \frac{\xi_2^2 f_2(z_1) f_2(z_1)}{\prod_{i=1}^{2}(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))}$$

$$+ \frac{1}{2}\frac{\xi_1\xi_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^{2}(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))}. \tag{2.24}$$

PROOF. By Lemma 2.1.1, we know that, as $n \to \infty$,

$$X'_{n1} \to z_1 \text{ and } X'_{n2} \to z_2 \quad \text{a.s.}$$

so that by the continuity of the f_i's $i = 1, 2$, we get

$$f_i(X'_{n1}) \longrightarrow f_i(z_1) \quad a.s. \tag{2.25}$$

and

$$f_i(X'_{n2}) \longrightarrow f_i(z_2) \quad a.s. \text{ as } n \to \infty. \tag{2.26}$$

Thus, by (2.20), (2.21), (2.22), (2.25) and (2.26), we have

$$\pi_1(\underset{\sim}{z}; \underset{\sim}{x}'_n) \longrightarrow \pi_1(\underset{\sim}{z}) \qquad \text{a.s.} \tag{2.27}$$

and

$$\pi_2(\underset{\sim}{z}; \underset{\sim}{x}'_n) \longrightarrow \pi_2(\underset{\sim}{z}) \qquad \text{a.s.} \qquad \text{as } n \to \infty. \tag{2.28}$$

Therefore, by (2.19), (2.27) and (2.28)

$$\begin{aligned} r(\underset{\sim}{z}) &= \lim_n r(\underset{\sim}{z}; \underset{\sim}{x}'_n) \\ &= \eta_2^*(\underset{\sim}{z}) \cdot \lim_n \pi_1(\underset{\sim}{z}; x'_n) + \eta_1^*(\underset{\sim}{z}) \cdot \lim_n \pi_2(\underset{\sim}{z}; x'_n) \\ &= \eta_2^*(\underset{\sim}{z}) \cdot \pi_1(\underset{\sim}{z}) + \eta_1^*(\underset{\sim}{z}) \cdot \pi_2(\underset{\sim}{z}) \qquad a.s.. \end{aligned} \tag{2.29}$$

Again, by the Lebesque Dominated Convergence Theorem (D.C.T.), we get

$$R_{(2)}(2) = \lim_n E\{r(\underset{\sim}{z}; x'_n)\} = Er(\underset{\sim}{z}). \tag{2.30}$$

The proof is complete. □

Now, we state and prove a theorem regarding the bounds on $R_{(2)}(2)$.

THEOREM 2.1.2. *The asymptotic risk $R_{(2)}(2)$ of the rule $\delta_{2B}^{(2)}$ has the following bounds*

$$R^*_{(2)}(1) \leq R_{(2)}(2) \leq 2R^*_{(2)}(1)(1 - R^*_{(2)}(1)), \tag{2.31}$$

where $R^*_{(2)}(1)$ is the Bayes risk (see 2.10) based on a single observation.

PROOF. By (2.29), (2.23) and (2.24) we write

$$\begin{aligned} r(\underset{\sim}{z}) = \eta_1^*(\underset{\sim}{z}) \Bigg[&\frac{\xi_2^2 f_2(z_1) f_2(z_2)}{(\xi_1 f_1(z_1) + \xi_2 f_2(z_1))(\xi_1 f_1(z_2) + \xi_2 f_2(z_2))} \\ &+ \frac{1}{2} \frac{\xi_1 \xi_2 (f_1(z_1) f_2(z_2) + f_1(z_2) f_2(z_1))}{(\xi_1 f_1(z_1) + \xi_2 f_2(z_1))(\xi_1 f_1(z_2) + \xi_2 f_2(z_2))} \Bigg] \\ + \eta_2^*(\underset{\sim}{z}) \Bigg[&\frac{\xi_1^2 f_1(z_1) f_1(z_2)}{(\xi_1 f_1(z_1) + \xi_2 f_2(z_1))(\xi_1 f_1(z_2) + \xi_2 f_2(z_2))} \\ &+ \frac{1}{2} \frac{\xi_1 \xi_2 (f_1(z_1) f_2(z_2) + f_1(z_2) f_2(z_1))}{(\xi_1 f_1(z_1) + \xi_2 f_2(z_1))(\xi_1 f_1(z_2) + \xi_2 f_2(z_2))} \Bigg]. \end{aligned}$$

Using (2.1) and (2.2) in above, we get

$$r(\underset{\sim}{z}) = \eta_1^*(\underset{\sim}{z})\left[\eta_2(z_1)\cdot\eta_2(z_2) + \frac{1}{2}(\eta_1(z_1)\cdot\eta_2(z_2) + \eta_1(z_2)\cdot\eta_2(z_1))\right]$$

$$+ \eta_2^*(\underset{\sim}{z})\left[\eta_1(z_1)\cdot\eta_1(z_2) + \frac{1}{2}(\eta_1(z_1)\cdot\eta_2(z_2) + \eta_1(z_2)\cdot\eta_2(z_1))\right]$$

$$= \eta_1^*(\underset{\sim}{z})\left[\frac{1}{2}\eta_2(z_1)\eta_2(z_2) + \frac{1}{2}\eta_1(z_1)\eta_2(z_2) + \frac{1}{2}\eta_2(z_1)\eta_2(z_2) + \frac{1}{2}\eta_1(z_2)\eta_2(z_1)\right]$$

$$+ \eta_2^*(\underset{\sim}{z})\left[\frac{1}{2}\eta_1(z_1)\eta_1(z_2) + \frac{1}{2}\eta_1(z_1)\eta_2(z_2) + \frac{1}{2}\eta_1(z_1)\eta_1(z_2) + \frac{1}{2}\eta_1(z_2)\eta_2(z_1)\right]$$

$$= \eta_1^*(\underset{\sim}{z})\left[\frac{1}{2}\eta_2(z_2)(\eta_1(z_1) + \eta_2(z_1)) + \frac{1}{2}\eta_2(z_1)(\eta_1(z_2) + \eta_2(z_2))\right]$$

$$+ \eta_2^*(\underset{\sim}{z})\left[\frac{1}{2}\eta_1(z_1)(\eta_1(z_2) + \eta_2(z_2)) + \frac{1}{2}\eta_1(z_2)(\eta_1(z_1) + \eta_2(z_1))\right]$$

$$= \eta_1^*(\underset{\sim}{z})\cdot\frac{1}{2}\left[\eta_2(z_1) + \eta_2(z_2)\right] + \eta_2^*(\underset{\sim}{z})\cdot\frac{1}{2}\left[\eta_1(z_1) + \eta_1(z_2)\right]. \tag{2.32}$$

Note that

$$r(\underset{\sim}{z}) \geq \min\{\frac{1}{2}(\eta_1(z_1) + \eta_1(z_2)), \frac{1}{2}(\eta_2(z_1) + \eta_2(z_2))\}$$

$$= \frac{1}{2}\min\{\eta_1(z_1) + \eta_1(z_2), 1 - \eta(z_1) + 1 - \eta(z_2)\}$$

$$\geq \frac{1}{2}\{\min(\eta_1(z_1), 1 - \eta_1(z_1)) + \min(\eta_1(z_2), 1 - \eta(z_2))\}. \tag{2.33}$$

Taking expections both sides of (2.33) with respect to the distribution of $\underset{\sim}{Z} = (z_1, z_2)$ and using (2.30), we have

$$R_{(2)}(2) = E(r(\underset{\sim}{z})) \geq \frac{1}{2} \int\int \min(\eta_1(z_1), 1 - \eta_1(z_1))$$

$$\times (\xi_1 f_1(z_1) f_1(z_2) + \xi_2 f_2(z_1) f_2(z_2)) dz_1 dz_2$$

$$+ \frac{1}{2} \int\int \min(\eta_1(z_2), 1 - \eta_1(z_2))$$

$$\times (\xi_1 f_1(z_1) f_1(z_2) + \xi_2 f_2(z_1) f_2(z_2)) dz_1 dz_2$$

$$= \frac{1}{2} [\int \min(\eta_1(z_1), 1 - \eta_1(z_1)) \sum_{i=1}^{2} \xi_i f_i(z_1) dz_1$$

$$+ \int \min(\eta_1(z_2), 1 - \eta_1(z_2)) \sum_{i=1}^{2} \xi_i f_i(z_2) dz_2].$$

Now using (2.10) in above we get

$$R_{(2)}(2) \geq \frac{1}{2} \left[R^*_{(2)}(1) + R^*_{(2)}(1) \right] = R^*_{(2)}(1). \tag{2.34}$$

Again, taking expectations both sides of (2.32) w.r.t. the distribution of $\underset{\sim}{z} = (z_1, z_2)$ and using (2.30), we write

$$R_{(2)}(2) = E(r(\underset{\sim}{z})) = \frac{1}{2} [\int\int \xi_1 f_1(z_1) f_1(z_2) \eta_2(z_1) dz_1 dz_2$$

$$+ \int\int \xi_1 f_1(z_1) f_1(z_2) \eta_2(z_2) dz_1 dz_2 + \int\int \xi_2 f_2(z_1) f_2(z_2) \eta_1(z_1) dz_1 dz_2$$

$$+ \int\int \xi_2 f_2(z_1) f_2(z_2) \eta_1(z_2) dz_1 dz_2]$$

$$= \frac{1}{2} [\int \xi_1 f_1(z_1) \eta_2(z_1) dz_1 + \int \xi_1 f_1(z_2) \eta_2(z_2) dz_2$$

$$+ \int \xi_2 f_2(z_1)\eta_1(z_1)dz_1 + \int \xi_2 f_2(z_2)\eta_1(z_2)dz_1 dz_2]$$

$$= \frac{1}{2}[4 \int \eta_1(z)\eta_2(z)(\xi_1 f_1(z) + \xi_2 f_2(z))dz]$$

$$= 2E[\eta_1(z)(1 - \eta_1(z))]. \tag{2.35}$$

Recall that the conditional Bayes risk from (2.8) for $s = 2$ is given by

$$r^*(z) = \min\{\eta_2(z), \eta_1(z)\} \tag{2.36}$$

which is symmetric in η_1 and η_2; since $\eta_1 + \eta_2 = 1$, thus by (2.35), (2.36) and Jensen's inequality, we have

$$\begin{aligned} R_{(2)}(2) &= 2E[r^*(z)(1 - r^*(z))] \\ &= 2[Er^*(z) - E(r^*(z))^2] \\ &\leq 2[Er^*(z) - (Er^*(z))^2] \\ &= 2(R^*_{(2)}(1) - R^{*2}_{(2)}(1)) \\ &= 2R^*_{(2)}(1)(1 - R^*_{(2)}(1)). \end{aligned} \tag{2.37}$$

From (2.34) and (2.37), we get

$$R^*_{(2)}(1) \leq R_{(2)}(2) \leq 2R^*_{(2)}(1)(1 - R^*_{(2)}(1)). \tag{2.38}$$

The proof is complete. □

REMARKS 2.1.1. The bounds obtained for the risk $R_{(2)}(2)$ of the rule $\delta^{(2)}_{2B}$ in (2.16) are in terms of the Bayes Risk corresponding to a single observation classification rule and therefore are not the ones we had hoped for. These bounds are, in fact, exactly the same as those Cover and Hart (1967) obtained corresponding to their rule for the classification of a single observation. However the asymptotic risk $R_{(2)}(2)$ is less than the asymptotic risk of the "repeating" procedure (that is, when observations are classified one by one). Now we prove this last assertion.

Let $R(1)$ be the asymptotic NN risk when a single observation is classified. Thus, when two observations are classified repeatedly, the risk is given by

$$\begin{aligned} R\text{ (repeated) } &= 1-(1-R(1))^2 \\ &= 2R(1)-R^2(1); \end{aligned} \tag{2.39}$$

but by (2.35), we have

$$R_{(2)}(2) = R(1). \tag{2.40}$$

Consequently

$$\begin{aligned} R\text{ (repeated) } - R_{(2)}(2) &= R(1)-R^2(1) \\ &= R(1)(1-R(1)) \geq 0, \end{aligned}$$

which implies,

$$R_{(2)}(2) \leq R\text{ (repeated) }. \tag{2.41}$$

Thus, from (2.41), we conclude that $R_{(2)}(2)$ has a lower risk compared to the repeated case.

Now we propose a classification rule which generalizes the rule $\delta_{2B}^{(2)}$ in (2.16) to K observations and we obtain its limiting risk.

2.2. Asymptotic risk in a Bayesian model when K observations are classified.

Let $(X_1,\theta_1),(X_2,\theta_2),\ldots,(X_n,\theta_n)$ be a random training sample taking values in $\mathbb{R}^d \times \{1,2\}$.

Let $(\underset{\sim}{Z}= (Z_1,Z_2,\ldots,Z_K),\theta)$ be a random sample of size K to be classified between π_1 and π_2. We describe a classification rule $\delta_{KB}^{(2)}$ as follows:

First, we find the NN's (defined in (2.15)) of all Z_ℓ's for fall (or each) $\ell = 1, 2, \ldots, K$ and we identify them. Then,

$$\begin{aligned} &\text{Classify } \underset{\sim}{Z} \text{ into } \pi_1 \text{ with probability } 1 \\ &\qquad \text{if } \# \text{ of NN from } \pi_1 > \# \text{ of NN from } \pi_2; \\ &(\text{and of course to } \pi_2 \text{ with prob. } 1 \text{ if } \# \text{ of NN from} \\ &\quad \pi_2 > \# \text{ of NN from } \pi_1); \\ &\text{classify } \underset{\sim}{Z} \text{ into } \pi_1 \text{ with probability } \frac{1}{2}, \\ &\qquad \text{if } \# \text{ of NN from } \pi_1 = \# \text{ of NN from } \pi_2. \end{aligned} \tag{2.42}$$

Applying the rule $\delta_{KB}^{(2)}$ described in (2.42), we estimate θ. Suppose the estimate of θ is θ'_n.

Now we state a lemma on almost sure convergence of NN X'_{nj} to $Z_j, j = 1, 2, \ldots, K$.

LEMMA 2.2.1. *Let $Z_1, Z_2, \ldots, Z_K$ and $X_1, X_2, \ldots, X_n$ be i.i.d random variables. Let X'_{nj} be the NN to $Z_j, j = 1, 2, \ldots, K$. Then*

$$X'_{nj} \longrightarrow Z_j \text{ a.s. as } n \to \infty.$$

PROOF. The proof is similar to Lemma 2.1.1 and hence it is omitted.

We denote $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ and $\underset{\sim n}{X'} = (\underset{\sim n1}{X'}, \underset{\sim n2}{X'}, \ldots, \underset{\sim nK}{X'})$. Let S_{K-i} be the set of permutations of $\left\{\underbrace{1, 1, \ldots, 1}_{(K-i)}, \underbrace{2, 2, \ldots, 2}_{i}\right\}$ and also let $\underset{\sim}{r} = (r_1, r_2, \ldots, r_K)$ be an element of S_{K-i}. The conditional risk given $\underset{\sim}{Z} = \underset{\sim}{z} = (z_1, z_2, \ldots, z_K)$ and $\underset{\sim n}{X'} = (x'_{n1}, x'_{n2}, \ldots, x'_{nK}) = x'_n$ is given by

$$r(\underset{\sim}{z}; \underset{\sim n}{x'}) = E(L(\theta, \theta'_n) / \underset{\sim}{z}, \underset{\sim n}{x'})$$

which, by using the conditional independence of θ and θ'_n, can be written as

$$\begin{aligned} r(\underset{\sim}{z}; \underset{\sim n}{x'}) &= P(\theta = 2 / \underset{\sim}{z}) \cdot P(\theta'_n = 1 / \underset{\sim n}{x'}) + P(\theta = 1 / \underset{\sim}{z}) \cdot P(\theta'_n = 2 / \underset{\sim n}{x'}) \\ &= \eta_2^*(\underset{\sim}{z}) \cdot \pi_1(\underset{\sim}{z}; \underset{\sim n}{x'}) + \eta_1^*(\underset{\sim}{z}) \cdot \pi_2(\underset{\sim}{z}; \underset{\sim n}{x'}). \end{aligned} \tag{2.43}$$

Expressions for π_1 and π_2 are given below for the cases K even and K odd

Case I. when K, is even, say, $K = 2m$, then

$$\pi_1(\underset{\sim}{z}; \underset{\sim n}{x'}) = P(\theta'_n = 1/ \underset{\sim n}{x'})$$
$$= \{\sum_{i=0}^{m-1} C_{K-i}(n) + \frac{1}{2}C_m(n)\}\frac{1}{B_n} \tag{2.44}$$

$$\pi_2(\underset{\sim}{z}; \underset{\sim n}{x'}) = P(\theta'_n = 2/ \underset{\sim n}{x'})$$
$$= \{\sum_{i=m+1}^{2m} C_{K-i}(n) + \frac{1}{2}C_m(n)\}\frac{1}{B_n}, \tag{2.45}$$

with

$$C_K(n) = \xi_1^K \prod_{j=1}^{K} f_1(x'_{nj}),$$

$$C_{K-1}(n) = \xi_1^{K-1}\xi_2\{\prod_{j=1}^{K-1} f_1(x'_{nj})f_2(x'_{nK}) + \prod_{\substack{j=1\\ j\neq K-1}}^{K} f_1(x'_{nj})f_2(x'_{n\overline{K-1}}) +$$
$$\cdots + \prod_{j=2}^{K} f_1(x'_{nj})f_2(x'_{n1})\}, \tag{2.46}$$

$$\vdots$$

$$C_{K-i}(n) = \xi_1^{K-i}\xi_2^{i} \sum_{\underset{\sim}{r}\in S_{K-i}} f_{r_1}(x'_{n1})f_{r_2}(x'_{n2})\cdots f_{r_K}(x'_{nK}), \tag{2.47}$$

$$\vdots$$

$$C_0(n) = \xi_2^K \prod_{j=1}^{K} f_2(x'_{nj}), \tag{2.48}$$

and

$$B_n = \sum_{\ell=0}^{K} C_\ell(n). \tag{2.49}$$

THEOREM 2.2.1. *Let $z_1, z_2, \ldots, z_K$ be the continuity points of f_1 and f_2. Then the limiting conditional risk given $\underset{\sim}{Z} = (z_1, \ldots, z_K)$ and the unconditional risk for the rule $\delta_{KB}^{(2)}$ are given respectively by*

$$\begin{aligned} r(\underset{\sim}{z}) &= \lim_{n\to\infty} r(\underset{\sim}{z}; \underset{\sim}{x}'_n) \\ &= \eta_2^*(\underset{\sim}{z})\pi_1(\underset{\sim}{z}) + \eta_1^*(\underset{\sim}{z})\pi_2(\underset{\sim}{z}) \quad \text{a.s.} \end{aligned}$$

and

$$R_{(2)}(K) = \lim_{n\to\infty} Er(\underset{\sim}{z}; \underset{\sim}{x}'_n) = E(r(\underset{\sim}{z})),$$

where

$$\pi_1(\underset{\sim}{z}) = \{\sum_{i=0}^{m-1} C_{K-i} + \frac{1}{2}C_m\} \cdot \frac{1}{B} \tag{2.50}$$

and

$$\pi_2(\underset{\sim}{z}) = \{\sum_{i=m+1}^{2m-1} C_{K-i} + \frac{1}{2}C_m\} \cdot \frac{1}{B} \tag{2.51}$$

$$C_K = \xi_1^K \prod_{j=1}^{K} f_1(z_j)$$

$$\vdots$$

$$C_{K-i} = \xi_1^{K-i}\xi_2^i \sum_{\mathbf{r}\in S_{K-i}} f_{r_1}(z_1)f_{r_2}(z_2)\cdots f_{r_K}(z_K),$$

$$\vdots$$

$$C_0 = \xi_2^K \prod_{j=1}^{K} f_2(z_j)$$

and

$$B = \sum_{\ell=0}^{K} C_\ell.$$

PROOF. By Lemma 2.2.1, we know

$$X'_{nj} \to Z_j \quad \text{a.s.} \quad \text{as } n \to \infty, \quad j = 1, 2, \ldots K.$$

Since both f_1 and f_2 are continuous at $z_1, z_2, \ldots, z_K$, we have

$$\begin{aligned} C_{K-i}(n) &\xrightarrow{\text{a.s.}} \xi_1^{K-i}\xi_2 \sum_{r \in S_{K-i}} f_{r_1}(z_1) f_{r_2}(z_2) \ldots f_{r_K}(z_K) \\ &= C_{K-i} \qquad \forall i = 1, 2, \ldots, K, \end{aligned} \tag{2.52}$$

and

$$B_n \xrightarrow{\text{a.s.}} \sum_{i=0}^{K} C_{K-i} = B. \tag{2.53}$$

Therefore by (2.44)-(2.53), we get

$$\pi_1(\underset{\sim}{z}; \underset{\sim}{x}'_n) \xrightarrow{\text{a.s.}} \pi_1(\underset{\sim}{z}), \tag{2.54}$$

and

$$\pi_2(\underset{\sim}{z}; \underset{\sim}{x}'_n) \xrightarrow{\text{a.s.}} \pi_2(\underset{\sim}{z}). \tag{2.55}$$

So by (2.43), (2.54) and (2.55) we have

$$\begin{aligned} r(\underset{\sim}{z}) &= \lim_{n\to\infty} r(\underset{\sim}{z}, \underset{\sim}{x}'_n) \\ &= \eta_2^*(\underset{\sim}{z})\pi_1(\underset{\sim}{z}) + \eta_1^*(\underset{\sim}{z})\pi_2(\underset{\sim}{z}). \quad \text{a.s.} \end{aligned}$$

By D.C.T. we also get

$$R_{(2)}(K) = \lim_{n\to\infty} Er(\underset{\sim}{z}; \underset{\sim}{x}'_n) = E(r(\underset{\sim}{z})).$$

The proof is complete . □

Case II. When $K = 2m + 1$, we replace $\pi_1(\underset{\sim}{z})$ and $\pi_2(\underset{\sim}{z})$ in (2.50) and (2.51) by

$$\pi_1(\underset{\sim}{z}) = \sum_{i=0}^{m} C_{K-i}/B, \tag{2.56}$$

and

$$\pi_2(\underset{\sim}{z}) = \sum_{i=m+1}^{2m+1} C_{K-i}/B, \text{ respectively.} \tag{2.57}$$

REMARK 2.2.1. We have not been able to extend (2.31) for general $K > 2$; however, we expect that the following bounds hold for any $K > 2$:

$$R^*_{(2)}(K/2) \le R_{(2)}(K) \le 2R^*_{(2)}(K/2)(1 - R^*_{(2)}(K/2)) \text{ if } K \text{ is even,} \quad (2.58)$$

$$R^*_{(2)}(\frac{K+1}{2}) \le R_{(2)}(K) \le 2R^*_{(2)}(\frac{K+1}{2})(1 - R^*_{(2)}(\frac{K+1}{2})) \text{ if } K \text{ is odd} \quad (2.59)$$

where $R^*_{(s)}(K)$ is defined in (2.9). Thus, we notice that when K increases the upper bound of $R_{(2)}(K)$ decreases and that when $K = 2$, the bounds in (2.58) become the bounds in (2.31). The bounds mentioned in (2.58) and (2.59) have been verified in specific cases and an example of which is given below:

Example 2.1. Let $f_1(x) = 2x, \quad 0 < x < 1$, $f_2(y) = 2 - 2y, \quad 0 < y < 1$ and $\xi_1 = \xi_2 = \frac{1}{2}$. By (2.35), we write $R_{(2)}(2)$ as

$$\begin{aligned} R_{(2)}(2) &= 2E(\eta_1(z)\eta_2(z)) \\ &= 2\int_0^1 \frac{\xi_1 f_1(z)\xi_2 f_2(z)}{\xi_1 f_1(z) + \xi_2 f_2(z)} dz \\ &= 2\int_0^1 z(1-z)dz \\ &= 2\int_0^1 (z - z^2)dz \\ &= 2(\frac{1}{2} - \frac{1}{3}) = \frac{1}{3} = \frac{8}{24}. \end{aligned} \quad (2.60)$$

The Bayes risk from (2.9) for $K = 2$ and $s = 2$, is

$$
\begin{aligned}
R^*_{(2)}(2) &= \int\int \min(\xi_1 f_1(z_1)f_1(z_2), \xi_2 f_2(z_1)f_2(z_2))dz_1dz_2 \\
&= 2\int_0^1\int_0^1 \min(z_1z_2, (1-z_1)(1-z_1))dz_1dz_2 \\
&= 2[\iint_{z_1z_2<(1-z_1)(1-z_2)} z_1z_2dz_1dz_2 + \iint_{z_1z_2>(1-z_1)(1-z_2)} (1-z_1)(1-z_2)dz_1dz_2] \\
&= 2[\iint_{z_1+z_2<1} z_1z_2dz_1dz_2 + \iint_{z_1+z_2>1} (1-z_1)(1-z_2)dz_1dz_2] \\
&= 2[\int_0^1 z_1\int_0^{z_1} z_2dz_2dz_1 + \int_0^1(1-z_1)\int_{1-z_1}(1-z_2)dz_2dz_1] \\
&= 2[\frac{1}{2}\int_0^1 z_1(1-z_1)^2dz_1 + \int_0^1(1-z_1)\int_0^{z_1} z_2dz_2dz_1] \\
&= 2[\frac{1}{2}\int_0^1 z_1^2(1-z_1)dz_1 + \int_0^1 z_1^2(1-z_1)dz_2] \\
&= 2\int_0^1 z(1-z)dz = 2[\frac{1}{3}-\frac{1}{4}] = \frac{1}{6}.
\end{aligned}
$$

Thus

$$
R^*_{(2)}(2) = \frac{1}{6} = \frac{9}{54}. \tag{2.61}
$$

The Bayes risk from (2.10) is

$$
\begin{aligned}
R^*_{(2)}(1) &= \int \min(\xi_1 f_1(z), \xi_2 f_2(z))dz \\
&= \int_0^1 \min(z, 1-z)dz \\
&= \int_{z<1-z} zdz + \int_{z>1-z}(1-z)dz \\
&= \int_0^{1/2} zdz + \int_{1/2}^1 (1-z)dz \\
&= \frac{1}{8}+\frac{1}{8} = \frac{1}{4} = \frac{6}{24}.
\end{aligned}
$$

Thus

$$
R^*_{(2)}(1) = \frac{6}{24}. \tag{2.62}
$$

Therefore, by (2.60) and (2.62), we have

$$R^*_{(2)}(1) = \frac{6}{24} \le R_{(2)}(2) = \frac{8}{24} \le 2R^*_{(2)}(1)(1 - R^*_{(2)}(1)) = \frac{9}{24}.$$

Thus, the inequality (2.58) is satisfied for $K = 2$. Now using Theorem 2.2.1 and (2.56) and (2.57) for $K = 3$, we get

$$R_{(2)}(3) = \int\int\int \xi_1 f_1(z_1)f_1(z_2)f_1(z_3)[\eta_2(z_1)\eta_2(z_2) + \eta_2(z_1)\eta_2(z_3)\eta_1(z_2) + \eta_2(z_2\eta_2(z_3)\eta_1(z_1)]$$
$$+ \int\int\int \xi_2 f_2(z_1)f_2(z_2)f_2(z_3)[\eta_1(z_1)\eta_1(z_2) + \eta_1(z_1)\eta_1(z_3)\eta_2(z_2) + \eta_1(z_2\eta_1(z_3)\eta_2(z_1)]$$
$$= I_1 + I_2.$$

$$I_1 = 4[\int_0^1\int_0^1\int_0^1 z_1z_2z_3[(1-z_1)(1-z_2) + (1-z_1)z_2(1-z_3) + (1-z_2)(1-z_3)z_1]dz_1dz_2dz_3$$
$$= 4[\int_0^1 z_1(1-z_1)\int_0^1 z_2(1-z_2)\int_0^1 z_3 + \int_0^1 z_1(1-z_1)\int_0^1 z_2^2\int_0^1 z_3(1-z_3) + \int_0^1 z_1^2\int_0^1 z_2(1-z_2)\int_0^1 z_3(1-z_3)]$$
$$= 4[\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{2} + \frac{1}{6}\cdot\frac{1}{3}\cdot\frac{1}{6} + \frac{1}{3}\cdot\frac{1}{6}\cdot\frac{1}{6}]$$
$$= \frac{7}{54}.$$

$$I_2 = 4[\int_0^1\int_0^1\int_0^1 (1-z_1)(1-z_2)(1-z_3)[z_1z_2) + z_1z_3(1-z_2) + z_2z_3(1-z_1)]dz_1dz_2dz_3$$
$$= 4[\int_0^1 z_1(1-z_1)\int_0^1 z_2(1-z_2)\int_0^1 (1-z_3) + \int_0^1 z_1(1-z_1)\int_0^1 (1-z_2)^2\int_0^1 z_3(1-z_3) + \int_0^1 (1-z_1)^2 \times \int_0^1 z_2(1-z_2)\int_0^1 z_3(1-z_3)]$$
$$= 4[\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{3} + \frac{1}{6}\cdot\frac{1}{3}\cdot\frac{1}{6} + \frac{1}{3}\cdot\frac{1}{6}\cdot\frac{1}{6}]$$
$$= \frac{7}{54}.$$

Thus,

$$R_{(2)}(3) = \frac{7}{54} + \frac{7}{54} = \frac{7}{27} = \frac{14}{54}. \tag{2.63}$$

Therefore, by (2.61) and (2.63), we have

$$R^*_{(2)}(2) = \frac{9}{54} \leq R_{(2)}(3) = \frac{14}{54} \leq 2R^*_{(2)}(2)(1 - R^*_{(2)}(2)) = \frac{15}{54}.$$

Thus, for $K = 3$ the inequality (2.59) is satisfied.

2.3 Asymptotic TPMC and its bounds in a nonparametric model when $K \geq 2$ observations are classified.

Let π_1 and π_2 be two distinct multivariate populations taking values in $\mathbb{R}^d$ with density functions f_1 and f_2 respectively. Assume that a random sample (Z_1, Z_2) comes from one or the other of the two populations. Now the problem is to classify (Z_1, Z_2) to π_1 or π_2.

Let $(X_{11}, X_{12}, \ldots X_{1n_1})$ and $(X_{21}, X_{22}, \ldots X_{2n_2})$ be the separate training samples from π_1 and π_2 respectively.

Consider the following 1-NN procedure:

Let U_1 and U_2 be the two Nearest Neighbors (using some distance function) of Z_1 and Z_2 respectively from the combined samples X_{1i}'s and X_{2j}'s and denote by $\phi_1(Z_1, Z_2;\ X_{1i}\text{'s}, X_{2j}\text{'s}\ i = 1, \ldots, n_1, j = 1, \ldots, n_2)$ the discriminant function, namely, the probability with which one decides to choose π_1 (over π_2). Then the 1-NN nonparametric classification procedure $\delta^{(2)}_{2N}$ is defined as (N denotes the non-parametric model)

$$\phi_1 = \begin{cases} 1 & \text{if } U_1 \text{ and } U_2 \text{ both are from } \pi_1 \\ 1/2 & \text{if one of } U_1 \text{ and } U_2 \text{ is from } \pi_1 \text{ and other is from } \pi_2 \\ 0 & \text{otherwise.} \end{cases} \tag{2.64}$$

We assume that $P(U_1 = U_2) = 0$, which holds when F_1 and F_2 are continuous. Let

$$T_{N1} = \|U_1 - Z_1\| = \min_{1 \leq i \leq N} \|W_i - Z_1\|, \tag{2.65}$$

and

$$T_{N2} = \|U_2 - Z_2\| = \min_{1 \le i \le N} \|W_i - Z_2\|, \tag{2.66}$$

where $\{W_1, W_2, \ldots W_N\}$, $(N = n_1 + n_2)$ stands for the pooled samples of X_{1i}'s and X_{2j}'s. Let $n = \min(n_1, n_2)$.

LEMMA 2.3.1. *If Z_1 and Z_2 are i.i.d. with density f_1, as $n \to \infty$, we have*

$$T_{N1} \to 0 \text{ a.s.} \tag{2.67}$$

and

$$T_{N2} \to 0 \text{ a.s.} \quad \text{as} \quad n \to \infty. \tag{2.68}$$

PROOF. For any $t_1 > 0$

$$\begin{aligned} P(T_{N1} > t_1) &= P(\|U_1 - Z_1\| > t_1) \\ &= P(\min_{1 \le i \le n} \|W_i - Z_1\| > t_1) \\ &= [P(\|X_{11} - Z_1\| > t_1)]^{n_1} \cdot [P\|X_{21} - Z_1\| > t_1]^{n_2} \\ &\to 0 \text{ as } n \to \infty. \end{aligned} \tag{2.69}$$

Since, T_{N1} is monotonically decreasing in N, by (2.69) and (2.18)

$$T_{N1} \to 0 \text{ a.s.}, \quad \text{as } n \to \infty.$$

Similarly, one can prove

$$T_{N2} \to 0 \text{ a.s.}, \quad \text{as } n \to \infty.$$

The proof is complete. □

Let E_{ij} denote the event that U_1 and U_2 come respectively from i-th and j-th populations $i = 1, 2$; $j = 1, 2$. We shall first find the conditional joint distribution of T_{N1} and T_{N2} given E_{ij} and $\underset{\sim}{Z} = (z_1, z_2)$ $i = 1, 2$ and $j = 1, 2$.

First we find joint c.d.f for T_{N1} and T_{N2} , when both U_1 and U_2 are from π_1. We have

$$G_{11}(t_1,t_2) = P[T_{N1} \le t_1, T_{N2} \le t_2/E_{11}, \underset{\sim}{z}]$$
$$= 1 - P_{11}[T_{N1} > t_1] - P_{11}[T_{N2} > t_2]$$
$$+ P_{11}[T_{N1} > t_1, T_{N2} > t_2]. \tag{2.70}$$

Now

$$P_{11}(T_{N1} > t_1, T_{N2} > t_2)$$
$$= \int_{\|u_1 - z_1\| > t_1} \int_{\|u_2 - z_2\| > t_2} n_1(n_1-1)[1 - P_1(S_{z_1}(\|u_1 - z_1\|)) - P_1(S_{z_2}(\|u_2 - z_2\|))]^{n_1-2}$$
$$\times [1 - P_2(S_{z_1}(\|u_1 - z_1\|)) - P_2(S_{z_2}(\|u_2 - z_2\|))]^{n_2}$$
$$\times dP_1(u_1)dP_1(u_2),$$

which by the transformation $u_1 - z_1 = v_1$ and $u_2 - z_2 = v_2$, equals

$$\int_{\|v_1\| > t_1} \int_{\|v_2\| > t_2} n_1(n_1-1)[1 - P_1(S_{z_1}(\|v_1\|)) - P_1(S_{z_1}(\|v_2\|))]^{n_1-2}$$
$$\times [1 - P_2(S_{z_1}(\|v_1\|)) - P_2(S_{z_2}(\|v_2\|))]^{n_2}$$
$$\times f_1(v_1 + z_1) \cdot f_1(v_2 + z_2) dv_1 dv_2 , \tag{2.71}$$

where $S_{z_i}(\|v_i\|)$, $i = 1,2$ is the sphere centered at z_i and radius $\|v_i\|$. Let $v_1 = (v_{11}, v_{12}, \ldots, v_{1d})'$; $v_2 = (v_{21}, v_{22}, \ldots, v_{2d})'$,

$$z_1 = (z_{11}, z_{12}, \ldots, z_{1d})' \text{ and } \quad z_2 = (z_{21}, z_{22}, \ldots, z_{2d})'.$$

Now consider the following transformations:

$$v_{11} = r_1 cos\theta_1$$
$$v_{22} = r_2 sin\theta_1 cos\theta_2$$
$$\vdots$$
$$v_{1\overline{d-1}} = r_1 \sin\theta_1 cos\theta_2 \ldots \sin\theta_{d-2} cos\theta_{d-1}$$
$$v_d = r_1 \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{d-2} \sin\theta_{d-1},$$

$$0 \le r_1 < \infty,\ 0 \le \theta_i \le \pi,\ i = 1,2,\ldots,d-2;\ 0 \le \theta_{d-1} \le 2\pi,$$

with jacobian $= |J_1| = r_1^{d-1} \sin^{d-2}\theta_1 \sin^{d-3}\theta_2 \ldots \sin\theta_{d-2}$, and $r_1 = \|v_1\|$, and

$$v_{21} = r_2 cos\alpha_1$$

$$v_{22} = r_2 sin\alpha_1 cos\alpha_2$$

$$\vdots$$

$$v_{2\overline{d-1}} = r_2 \sin\alpha_1 cos\alpha_2 \ldots \sin\alpha_{d-2} cos\alpha_{d-1}$$

$$v_{2d} = r_2 \sin\alpha_1 \sin\alpha_2 \ldots \sin\alpha_{d-2} \sin\alpha_{d-1},$$

$$0 \le r_2 < \infty,\ 0 \le \alpha_i \le \pi,\ i = 1,2,\ldots,d-2;\ 0 \le \alpha_{d-1} \le 2\pi,$$

with jacobian $= |J_2| = r_2^{d-1} \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2 \ldots \sin\alpha_{d-2}$, and $r_2 = \|v_2\|$.

Applying the above polar transformations in (2.71), we substitute the value of (2.71) in (2.70) and then differentiate (2.70) with respect to t_1 and t_2 to get the conditional joint p.d.f. of t_1 and t_2. Thus we have

$$\begin{aligned} f^{(1,1)}{}_{n_1,n_2}(t_1,t_2) &= n_1(n_1-1)[1 - P_1(S_{z_1}(t_1)) - P_1(S_{z_2}(t_2))]^{n_1-2} \\ &\times [1 - P_2(S_{z_1}(t_1)) - P_2(S_{z_2}(t_2))]^{n_2} \cdot t_1^{d-1} t_2^{d-1} \\ &\times \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} f_1(t_1 cos\theta_1 + z_{11}, t_1 \sin\theta_1 cos\theta_2 + z_{12}, \ldots, \\ &t_1 \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{d-1} + z_{1d}) \sin^{d-1}\theta_1 \sin^{d-3}\theta_2 \\ &\ldots \sin\theta_{d-2} d\theta_1 d\theta_2 \ldots d\theta_{d-1} \\ &\times \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} f_1(t_1 cos\alpha_1 + z_{21}, t_1 \sin\alpha_1 cos\alpha_2 + z_{22}, \ldots, \\ &t_2 \sin\alpha_1 \sin\alpha_2 \ldots sin\alpha_{d-1} + z_{2d}) \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2 \\ &\ldots \sin\alpha_{d-2} d\alpha_1 d\alpha_2 \ldots d\alpha_{d-1}. \end{aligned} \tag{2.72}$$

Similarly, one can obtain

$$\begin{aligned} f^{(2,2)}{}_{n_1,n_2}(t_1,t_2) &= n_2(n_2-1)[1 - P_2(S_{z_1}(t_1)) - P_2(S_{z_2}(t_2))]^{n_2-2} \\ &\times [1 - P_1(S_{z_1}(t_1)) - P_1(S_{z_2}(t_2))]^{n_1} \cdot t_1^{d-1} t_2^{d-1} \\ &\times \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} f_2(t_1 cos\theta_1 + z_{11}, t_1 \sin\theta_1 cos\theta_2 + z_{12}, \ldots, \end{aligned}$$

$$t_1 \sin\theta_1 \sin\theta_2 \dots sin\theta_{d-1} + z_{1d}) \sin^{d-2}\theta_1 \sin^{d-3}\theta_2$$
$$\dots \sin\theta_{d-2} d\theta_1 d\theta_2 \dots d\theta_{d-1}$$
$$\times \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} f_2(t_2 cos\alpha_1 + z_{21}, t_2 \sin\alpha_1 cos\alpha_2 + z_{22}, \dots,$$
$$t_2 \sin\alpha_1 \sin\alpha_2 \dots sin\alpha_{d-1} + z_{2d}) \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2$$
$$\dots \sin\alpha_{d-2} d\alpha_1 d\alpha_2 \dots d\alpha_{d-1}, \tag{2.73}$$

$$f^{(1,2)}{}_{n_1,n_2}(t_1,t_2) = n_1 n_2 [1 - P_1(S_{z_1}(t_1)) - P_1(S_{z_2}(t_2))]^{n_1-1}$$
$$\times [1 - P_2(S_{z_1}(t_1)) - P_2(S_{z_2}(t_2))]^{n_2-1} \cdot t_1^{d-1} t_2^{d-1}$$
$$\times \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} f_1(t_1 cos\theta_1 + z_{11}, t_1 \sin\theta_1 cos\theta_2 + z_{12}, \dots,$$
$$t_1 \sin\theta_1 \sin\theta_2 \dots sin\theta_{d-1} + z_{1d}) \sin^{d-2}\theta_1 \sin^{d-3}\theta_2$$
$$\dots \sin\theta_{d-2} d\theta_1 d\theta_2 \dots d\theta_{d-1}$$
$$\times \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} f_2(t_2 cos\alpha_1 + z_{21}, t_2 \sin\alpha_1 cos\alpha_2 + z_{22}, \dots,$$
$$t_2 \sin\alpha_1 \sin\alpha_2 \dots sin\alpha_{d-1} + z_{2d}) \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2$$
$$\dots \sin\alpha_{d-2} d\alpha_1 d\alpha_2 \dots d\alpha_{d-1}, \tag{2.74}$$

and

$$f^{(2,1)}{}_{n_1,n_2}(t_1,t_2) = n_1 n_2 [1 - P_1(S_{z_1}(t_1)) - P_1(S_{z_2}(t_2))]^{n_1-1}$$
$$\times [1 - P_2(S_{z_1}(t_1)) - P_2(S_{z_2}(t_2))]^{n_2-1} \cdot t_1^{d-1} t_2^{d-1}$$
$$\times \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} f_1(t_2 cos\alpha_1 + z_{11}, t_1 \sin\alpha_1 cos\alpha_2 + z_{22}, \dots,$$
$$t_2 \sin\alpha_1 \sin\alpha_2 \dots sin\alpha_{d-1} + z_{2d}) \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2$$
$$\dots \sin\alpha_{d-2} d\alpha_1 d\alpha_2 \dots d\alpha_{d-1}$$
$$\times \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} f_2(t_1 cos\theta_1 + z_{21}, t_1 \sin\theta_1 cos\theta_2 + z_{12}, \dots,$$
$$t_2 \sin\theta_1 \sin\theta_2 \dots sin\theta_{d-1} + z_{2d}) \sin^{d-2}\theta_1 \sin^{d-3}\theta_2$$
$$\dots \sin\theta_{d-2} d\theta_1 d\theta_2 \dots d\theta_{d-1}. \tag{2.75}$$

By Lemma 2.3.1 as $n \to \infty$

$$t_1 \to 0 \text{ a.s} \quad \text{and } t_2 \to 0 \text{ a.s..}$$

So that

$$[1 - P_1(S_{z_1}(t_1)) - P_1(S_{z_2}(t_2))]^2 \to 1, \text{ a.s.} \tag{2.76}$$

$$[1 - P_2(S_{z_1}(t_1)) - P_2(S_{z_2}(t_2))]^2 \to 1, \text{ a.s.} \tag{2.77}$$

and

$$\begin{aligned}
\int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} & f_i(t_1 cos\theta_1 + z_{11}, t_1 \sin\theta_1 cos\theta_2 + z_{22}, \dots, \\
& t_2 \sin\theta_1 \sin\theta_2 \dots sin\theta_{d-1} + z_{1d}) \\
& \times \sin^{d-2}\theta_1 \sin^{d-3}\theta_2 \dots \sin\theta_{d-2} d\theta_1 d\theta_2 \dots d\theta_{d-1} \\
& \to f_i(z_1) \int_0^{\pi} \sin^{d-2}\theta_1 d\theta_1 \int_0^{\pi} \sin^{d-3}\theta_2 d\theta_2 \cdots \\
& \int_0^{\pi} \sin\theta_{d-2} d\theta_{d-2} \cdot \int_0^{2\pi} d\theta_{d-1} \\
& = f_i(z_1) \frac{\sqrt{\pi}\sqrt{\frac{d-1}{2}}}{\sqrt{\frac{d}{2}}} \cdot \frac{\sqrt{\pi}\sqrt{\frac{d-2}{2}}}{\sqrt{\frac{d-1}{2}}} \cdots \frac{\sqrt{\pi}\sqrt{\frac{2}{2}}}{\sqrt{\frac{3}{2}}} \times 2\pi \\
& = f_i(z_1) \frac{2(\pi)^{\frac{d}{2}}}{\sqrt{\frac{d}{2}}}, \quad i = 1, 2, \text{ a.s. .}
\end{aligned} \tag{2.78}$$

Similarly

$$\begin{aligned}
\int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} & f_i(t_2 cos\alpha_1 + z_{21}, t_2 \sin\alpha_1 cos\alpha_2 + z_{22}, \dots, \\
& t_2 \sin\alpha_1 \sin\alpha_2 \dots sin\alpha_{d-1} + z_{2d}) \\
& \times \sin^{d-2}\alpha_1 \sin^{d-3}\alpha_2 \dots \sin\alpha_{d-2} d\alpha_1 d\alpha_2 \dots d\alpha_{d-1} \\
& \to f_i(z_2) \cdot \frac{2(\pi)^{\frac{d}{2}}}{\sqrt{\frac{d}{2}}}, \quad i = 1, 2. \text{ a.s. .}
\end{aligned} \tag{2.79}$$

Let

$$p_i = \lim_n \frac{n_i}{n_1 + n_2}, \quad i = 1, 2. \tag{2.80}$$

The conditional probability of deciding that z_1 and z_2 come from π_1, given $Z_1 = z_1$, $Z_2 = z_2$, is

$$\pi(z_1, z_2; n_1, n_2) = E(\phi_1 / Z_1 = z_1, Z_2 = z_2) \tag{2.81}$$

Now we state and prove the main theorem of this section.

THEOREM 2.3.1. *Let z_1 and z_2 be continuity points of both f_1 and f_2. Then the asymptotic conditional probability of deciding $\underset{\sim}{Z} = (z_1, z_2)$ from π_1, given $Z_1 = z_1$ and $Z_2 = z_2$ of rule $\delta_{2N}^{(2)}$ is given by*

$$\pi(z_1, z_2) = \lim_n \pi(z_1, z_2; n_1, n_2) = \widehat{\eta}_1 + \frac{1}{2}\widehat{\eta}_0,$$

where

$$\widehat{\eta}_1 = \frac{p_1^2 f_1(z_1) f_1(z_2)}{\prod_{i=1}^{2}\{p_1 f_1(z_i) + p_2 f_2(z_i)\}} \tag{2.82}$$

and

$$\widehat{\eta}_0 = \frac{p_2 p_1(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^{2}\{p_1 f_1(z_i) + p_2 f_2(z_i)\}}. \tag{2.83}$$

PROOF. From (2.64) and (2.81) we have

$$\begin{aligned}\pi(z_1, z_2; n_1, n_2) &= P(\phi_1 = 1/Z_1 = z_1, Z_2 = z_2) + \frac{1}{2}P(\phi_1 = \frac{1}{2}/Z_1 = z_1, Z_2 = z_2)\\ &= EP(\phi_1 = 1/U_1, U_2; z_1, z_2) + \frac{1}{2}EP(\phi_1 = \frac{1}{2}/U_1, U_2; z_1, z_2).\end{aligned} \tag{2.84}$$

It can be seen that

$$\begin{aligned}P(\phi_1 = 1/U_1, U_2; z_1, z_2) &= P(\phi_1 = 1/\|U_1 - z_1\| = t_1, \|U_2 - z_2\| = t_2; z_1, z_2)\\ &= \frac{f_{n_1,n_2}^{(1,1)}(t_1, t_2)}{f_{n_1,n_2}(t_1, t_2)},\end{aligned} \tag{2.85}$$

where

$$f_{n_1,n_2}(t_1, t_2) = f_{n_1,n_2}^{(1,1)}(t_1, t_2) + f_{n_1,n_2}^{(2,2)}(t_1, t_2) + f_{n_1,n_2}^{(1,2)}(t_1, t_2) + f_{n_1,n_2}^{(2,1)}(t_1, t_2). \tag{2.86}$$

Now using (2.72) - (2.80), (2.85) and (2.86) we get

$$\begin{aligned}&\lim_n P(\phi_1 = 1/U_1, U_2; z_1, z_2)\\ &= \frac{p_1^2 f_1(z_1) f_1(z_2)}{p_1^2 f_1(z_1) f_1(z_2) + p_2^2 f_2(z_1) f_2(z_2) + p_1 p_2(f_1(z_1) f_2(z_2) + f_1(z_2) f_2(z_1))}\\ &= \frac{p_1^2 f_1(z_1) f_1(z_2)}{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_2 f_1(z_2) + p_2 f_2(z_2))}.\end{aligned} \tag{2.87}$$

Similarly, one would get

$$\lim_n P(\phi_1 = \frac{1}{2}/U_1, U_2; z_1, z_2) = \lim_n \frac{f^{(1,2)}_{n_1,n_2}(t_1,t_2) + f^{(2,1)}_{n_1,n_2}(t_1,t_2)}{f_{n_1,n_2}(t_1,t_2)}$$
$$= \frac{p_1p_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{(p_1f_1(z_1) + p_2f_2(z_1))(p_1f_1(z_2) + p_2f_2(z_2))}. \quad (2.88)$$

Now using (2.84), (2.87), (2.88) and by DCT, we get

$$\begin{aligned}\pi(z_1, z_2) &= \lim_n \pi(z_1, z_2; n_1, n_2)\\ &= \frac{p_1^2 f_1(z_1 f_1(z_2)}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}\\ &\quad + \frac{1}{2}\frac{p_1p_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}\\ &= \widehat{\eta}_1 + \frac{1}{2}\widehat{\eta}_0.\end{aligned}$$

The proof is complete. □

Now we derive the TPMC for the rule $\delta^{(2)}_{2N}$. The limiting PMC's are given as follows:

$$\begin{aligned}\alpha_1 &= \lim_{n\to\infty} P(\text{Decide}(z_1, z_2) \in \pi_2/(z_1, z_2) \in \pi_1)\\ &= \int\int (1 - \pi(z_1, z_2))f_1(z_1)f_1(z_2)dz_1dz_2\\ &= \int\int [\frac{p_2^2 f_2(z_1)f_2(z_2)}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}\\ &\quad + \frac{1}{2}\frac{p_1p_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}]\\ &\quad \times f_1(z_1)f_1(z_2)dz_1dz_2 \end{aligned} \quad (2.89)$$

and

$$\begin{aligned}\alpha_2 &= \lim_{n\to\infty} P(\text{Decide}(z_1, z_2) \in \pi_1/(z_1, z_2) \in \pi_2)\\ &= \int\int \pi(z_1, z_2)f_2(z_1)f_2(z_2)dz_1dz_2\\ &= \int\int [\frac{p_1^2 f_1(z_1)f_1(z_2)}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}\\ &\quad + \frac{1}{2}\frac{p_1p_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^2 \{p_1f_1(z_i) + p_2f_2(z_i)\}}]\\ &\quad \times f_2(z_1)f_2(z_2)dz_1dz_2. \end{aligned} \quad (2.90)$$

Suppose ξ_1 and ξ_2 are the prior probabilities, then we take $p_i = \xi_i \quad (i = 1, 2)$ for comparison of the limiting TPMC. Now the limiting TPMC for the rule $\delta_{2N}^{(2)}$ is given by

$$\begin{aligned} R_{(2)}(2) &= \xi_1\alpha_1 + \xi_2\alpha_2 \\ &= \int\int r(\underset{\sim}{z})(\xi_1 f_1(z_1)f_1(z_2) + \xi_2 f_2(z_1)f_2(z_2))dz_1dz_2 \\ &= Er(\underset{\sim}{z}) \end{aligned} \tag{2.91}$$

where

$$\begin{aligned} r(\underset{\sim}{z}) = \eta_1^*(\underset{\sim}{z}) & \frac{\xi_2^2 f_2(z_1)f_2(z_2)}{\prod_{i=1}^2\{\xi_1 f_1(z_i) + \xi_2 f_2(z_i)\}} \\ & + \frac{1}{2}\frac{\xi_1\xi_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^2\{\xi_1 f_1(z_i) + \xi_2 f_2(z_i)\}}] \\ + \eta_2^*(\underset{\sim}{z})[& \frac{\xi_1^2 f_1(z_1)f_1(z_2)}{\prod_{i=1}^2\{\xi_1 f_1(z_i) + \xi_2 f_2(z_i)\}} \\ & + \frac{1}{2}\frac{\xi_1\xi_2(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{\prod_{i=1}^2\{\xi_1 f_1(z_i) + \xi_2 f_2(z_i)\}}]. \end{aligned}$$

Using (2.23) and (2.24) in above, we write

$$r(\underset{\sim}{z}) = \eta_1^*(\underset{\sim}{z})\pi_2(\underset{\sim}{z}) + \eta_2^*(\underset{\sim}{z})\pi_1(\underset{\sim}{z}) \tag{2.92}$$

where η_i^*, $i = 1, 2$ for $K = 2$ are defined in (2.2). Thus the total limiting PMC in (2.91) is exactly equal to the total limiting risk in (2.30) in the Bayesian model. Therefore (2.91) will have same bounds that are obtained in Theorem 2.1.2. In the nonparametric set-up when K observations are to be classified using rule $\delta_{KB}^{(2)}$, the same asymptotic TPMC $R_{(2)}(K)$ can be obtained as in the Bayesian model. It is routine and lengthy, so it is omitted.

CHAPTER 3.

CLASSIFICATION OF MULTIPLE UNIVARIATE OBSERVATIONS USING MULTISTAGE RANK NEAREST NEIGHBOR RULE

3.0 Introduction. We consider the problem of classifying multiple univariate observations from one of the two given populations π_1 and π_2, thus generalizing the work of Das Gupta and Lin (1980) from one to K observations. Let $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ be training samples from π_1 and π_2 respectively. Suppose $\underset{\sim}{Z} = (Z_1, \ldots, Z_K)$ is a random sample from either π_1 or π_2 and is to be classified into one or the other. We shall propose a left-right Multi-Stage Rank Nearest Neighbor rule. We now describe the first stage RNN rule as follows: Combine X_i's, Y_j's and Z_ℓ's and arrange them in increasing order, and look for the first left and right hand rank nearest neighbors of each $Z_\ell \; \forall \ell = 1, 2, \ldots, K$ and identify them. Then the first stage Rank Nearest Neighbor (RNN)rule is:

Classify $\underset{\sim}{Z}$ into π_1 with probability 1,

$$\left.\begin{array}{l} \quad \text{if } \#(1^{st} \text{ stage) RNN's from } \pi_1 > \#(1^{st} \text{ stage) RNN's from } \pi_2 \,, \\ \text{classify } \underset{\sim}{Z} \text{ into } \pi_2 \text{ with probability } 1 \\ \quad \text{if } \#(1^{st} \text{ stage) RNN's from } \pi_1 < \#(1^{st} \text{ stage) RNN's from } \pi_2 \,, \\ \text{classify } \underset{\sim}{Z} \text{ into } \pi_1 \text{ or } \pi_2 \text{ with probability } 1/2 \text{ each} \\ \quad \text{if } \#(1^{st} \text{ stage) RNN's from } \pi_1 = \#(1^{st} \text{ stage) RNN's from } \pi_2 \,. \end{array}\right\} \quad (3.1)$$

To reduce the chance of randomization in the above RNN rule defined by (3.1) we consider a multi-stage version as follows:

If the 1st-stage RNN rule leads to a tie, delete those tied observations and apply the first-stage rule to the remaining observations. Proceed in this way, moving to the next stage whenever a tie occurs. The M-stage RNN rule denoted by MRNN is defined to be the one which terminates at the M-th stage (and allows for a tie in the final stage). We derive the asymptotic total PMC $R^{(1)}_{(2)}(K)$ for the 1st-stage RNN rule where $K = 2$. We explicitly derive the total PMC $R^{(1)}_{(2)}(2)$

and show that $R_{(2)}^{(1)}(2) \leq R_{(2)}^{(1)}(1)$ when $\xi_1 = \xi_2$. We also illustrate this by an example.

In the case of M-stage, we prove that the asymptotic total PMC $R_{(2)}^{(M)}(2)$ decreases as the stage level (M) increases. We also suggest an estimate for $R_{(2)}^{(1)}(2)$ which is asymptotically unbiased and consistent. We shall denote the c.d.f's of π_1 and π_2 by F_1 and F_2, respectively and assume that F_i possesses a density function f_i w.r.t. Lebesque measure μ, i=1,2.

3.1 Asymptotic PMC's of the First-stage RNN rule when $K = 2$.

We assume $\underset{\sim}{Z} = (Z_1, Z_2)$ is either from π_1 or π_2. The following lemma shows that asymptotically the right-hand and the left-hand neighbors of Z_ℓ ($\ell = 1, 2$) exist atthe M-th stage. We denote the right-hand and left-hand M^{th}RNN's of Z_ℓ by $V_\ell^{(M)}$ and $U_\ell^{(M)}$ respectively, $\ell = 1, 2$. Let $n = \min(n_1, n_2)$ and assume $Z_1 < Z_2$.

LEMMA 3.1.1. *If $M/n \to 0$ as $n \to \infty$, the probability that there are at least $2M$ observations between Z_1 and Z_2 and also M observations to the left of Z_1 and to right of Z_2 is one.*

PROOF. Since F_1 is absolutely continuous, it is sufficient to prove the lemma conditional on $Z_1 = z_1$ and $Z_2 = z_2$, with $z_1 < z_2$ and $F_1(z_2) - F_1(z_1) > 0$. Define

$$W_i = I_{(z_1, z_2)}(X_i) \quad i = 1, 2, \ldots n_1, \; W = \sum_{i=1}^{n_1} W_i, \tag{3.2}$$

where I is an indicator function. Then $E(W_i) = F_1(z_2) - F_1(z_1) > 0$ for each i. By the strong law of large numbers, we have from (3.2) that

$$P\Big[\frac{\sum_{i=1}^{n_1} W_i}{n_1} \to E(W_1) \text{ as } n_1 \to \infty\Big] = 1 \tag{3.3}$$

Now since $\frac{M}{n_1} \to 0$ as $n_1 \to \infty$ and $E(W_1) > 0$, these exists an integer N such that $E(W_1) > \frac{2M}{n_1}$ for $n_1 \geq N$, so that from (3.3)

$$P\Big(\sum_{i=1}^{n_1} W_i \geq 2M \text{ for all sufficiently large } n\Big) = 1.$$

The corresponding results for left-hand RNN of Z_1 and right-hand RNN of Z_2 can be proved similarly by defining $W_i = I_{(-\infty,z_1)}(X_i)$ and $W_i = I_{(z_2,\infty)}(X_i)$ respectively. The proof is complete. □

Next we prove $U_\ell^{(M)}$ and $V_\ell^{(M)}$ converge to Z_ℓ almost surely as $n \to \infty$.

LEMMA 3.1.2. *If Z_1 and Z_2 are distributed as F_1, and $\frac{M}{n_1} \to 0$, as $n_1 \to \infty$, then*

$$U_1^{(M)},\ V_1^{(M)} \to Z_1 \quad a.s.$$

and

$$U_2^{(M)},\ V_2^{(M)} \to Z_2 \quad a.s.\ \text{as}\ n \to \infty.$$

PROOF. Let $S_1 = \{z : F_1(z+\varepsilon) - F_1(z) > 0, F_1(z) - F_1(z-\varepsilon) > 0 \quad \forall \varepsilon > 0\}$. Then $P(Z_1 \in S_1) = 1$ and $P(Z_2 \in S_1) = 1$. This follows from the fact that the set of intervals in which F_1 is constant is at most countable which implies in view of absolute continuity of F_1 that the set of endpoints of these intervals has F_1-measure zero with probability one. Accordingly, for every $\varepsilon > 0, F_1(Z_1) - F_1(Z_1 - \varepsilon) > 0$ and $F_1(Z_2) - F_1(Z_2 - \varepsilon) > 0$. a.s.. If W is as defined in (3.2) with $(z_1 - \varepsilon, z_1)$, in place of (z_1, z_2), we have for almost all given $Z_1 = z_1$, $\{U_1^{(M)} < z_1 - \varepsilon\} \subset \{W < M\}$ which implies

$$P(U_1^{(M)} < z_1 - \varepsilon) \leq P(W < M). \tag{3.4}$$

Letting $q_1 = F_1(z_1) - F_1(z_1 - \varepsilon)$ so that $E(W) = n_1 q_1 > 0$ and choosing n_1 sufficiently large such that for some $t > 0$

$$q_1 - M/n_1 \geq t > 0 \quad \text{i.e.} \quad n_1(q_1 - t) \geq M, \tag{3.5}$$

we obtain by (3.4) and (3.5)

$$P(W < M) \leq P(W < n_1(q_1 - t)) < e^{-2n_1 t^2}, \tag{3.6}$$

where the last inequality follows in view of the well-known Hoeffding inequality for sums of independent random variables (See Hoeffding (1963) p.15). From (3.4) to (3.6), we have

$$\sum_{n_1=1}^{\infty} P(U_1^{(M)} < z_1 - \varepsilon) < \sum_{n_1=1}^{\infty} e^{-2n_1 t^2} < \infty. \tag{3.7}$$

By the Borel-Cantelli lemma and (3.7), we can then conclude that $U_1^{(M)} \to z_1$ a.s.. Similary, we can show that $U_2^{(M)} \to z_2$ a.s., $V_1^{(M)} \to z_1$ and $V_2^{(M)} \to z_2$ a.s. as $n \to \infty$ for almost all $z_1, z_2 \in S_1$. The proof is complete. □

Define

$$\phi^{(M)} = \begin{cases} 1 & \text{if the total no. of X-MRNN of } Z_1,\ldots,Z_K \text{ is greater than the total no. of Y-MRNN of } Z_1,\ldots,Z_K. \\ 1/2 & \text{if they are equal} \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

Let $A_\ell^{(M)}$ be the event that both $U_\ell^{(M)}$ and $V_\ell^{(M)}$ $\ell = 1, 2, \ldots, K$ are well-defined at the M-th stage.

First consider the case $K = 2$ i.e., when two observations Z_1 and Z_2 are to be classified between π_1 and π_2. We assume without loss of generality $Z_1 < Z_2$. Then the conditional probability, given $Z_1 = z_1$ and $Z_2 = z_2$, of deciding Z_1 and Z_2 are from π_1 using 1st-stage RNN rule, is given by

$$\pi^{(1)}(z_1, z_2; n_1, n_2) = E(\phi^{(1)}/Z_1 = z_1, Z_2 = z_2)$$

$$= E(\phi^{(1)} I_{(A_1^{(1)} \cap A_2^{(1)})}/Z_1 = z_1, Z_2 = z_2)$$

$$+ E(\phi^{(1)} I_{(A_1^{(1)} \cap A_2^{(1)})^c}/Z_1 = z_1, Z_2 = z_2). \tag{3.9}$$

However

$$E(\phi^{(1)}I_{(A_1^{(1)}\cap A_2^{(1)})^c}/Z_1 = z_1, Z_2 = z_2) = E(\phi^{(1)}I_{(A_1^{(1)c}\cup A_2^{(1)c})}/Z_1 = z_1, Z_2 = z_2)$$
$$\leq P\{(A_1^{(1)^c} \cup A_2^{(1)^c})/Z_1 = z_1, Z_2 = z_2\}$$
$$\leq P(A^{(1)^c}/Z_1 = z_1) + P(A^{(1)^c}/Z_2 = z_2)$$
$$\longrightarrow 0, \quad \text{as} \quad n \to \infty, \qquad (3.10)$$

the last convergence holding by lemma 3.1.1. Now we write

$$E(\phi^{(1)}I_{(A_1^{(1)}A_2^{(1)})}/Z_1 = z_1, Z_2 = z_2)$$

$$= P(\{\phi^{(1)} = 1\} \cap A_1^{(1)}A_2^{(1)}/Z_1 = z_1, Z_2 = z_2)$$

$$+ \frac{1}{2}P(\{\phi^{(1)} = 1/2\} \cap A_1^{(1)}A_2^{(1)}/Z_1 = z_1, Z_2 = z_2)$$

$$= EP^{(4,3)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, U_2^{(1)}, V_2^{(1)}; z_1, z_2)$$

$$+ \frac{1}{2}EP^{(2)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, U_2^{(1)}, V_2^{(1)}; z_1, z_2), \qquad (3.11)$$

where

$$P^{(4,3)}_{n_1,n_2}(u_1, v_1, u_2, v_2; z_1, z_2)$$
$$= P(\phi^{(1)} = 1/U_1^{(1)} = u_1, V_1^{(1)} = v_1, U_2^{(1)} = u_2, V_2^{(1)} = v_2; A_1^{(1)}A_2^{(1)}), \quad (3.12)$$

$$P^{(2)}_{n_1,n_2}(u_1, v_1, u_2, v_2; z_1, z_2)$$
$$= P(\phi^{(1)} = \frac{1}{2}/U_1^{(1)} = u_1, V_1^{(1)} = v_1, U_2^{(1)} = u_2, V_2^{(1)} = v_2; A_1^{(1)}A_2^{(1)}) \quad (3.13)$$

and

$$P^{(1,0)}_{n_1,n_2}(u_1,v_1,u_2,v_2;z_1,z_2)$$
$$= P(\phi^{(1)} = 0/U_1^{(1)} = u_1, V_1^{(1)} = v_1, U_2^{(1)} = u_2, V_2^{(1)} = v_2; A_1^{(1)}A_2^{(1)}). \quad (3.14)$$

It can be seen from (3.12) to (3.14) that

$$P^{(4,3)}_{n_1,n_2}(u_1,v_1,u_2,v_2;z_1,z_2) = \frac{C_4(n_1,n_2)}{B(n_1,n_2)} + \frac{C_3(n_1,n_2)}{B(n_1,n_2)} \quad (3.15)$$

$$P^{(2)}_{n_1,n_2}(u_1,v_1,u_2,v_2;z_1,z_2) = \frac{C_2(n_1,n_2)}{B(n_1,n_2)} \quad (3.16)$$

$$P^{(1,0)}_{n_1,n_2}(u_1,v_1,u_2,v_2;z_1,z_2) = \frac{C_1(n_1,n_2)}{B(n_1,n_2)} + \frac{C_0(n_1,n_2)}{B(n_1,n_2)}, \quad (3.17)$$

where, by setting $p_1^{(1)} = F_1(v_1) - F_1(u_1)$, $p_1^{(2)} = F_1(v_2) - F_1(u_2), p_2^{(1)} = F_2(v_1) - F_2(u_1)$ and $p_2^{(2)} = F_2(v_2) - F_2(u_2)$ we have

$$C_0(n_1,n_2) = n_2(n_2-1)(n_2-2)(n_2-3)\left(1-\overline{p_2^{(1)}+p_2^{(2)}}\right)^{n_2-4}$$
$$\times \left(1-\overline{p_1^{(1)}+p_1^{(2)}}\right)^{n_1} f_2(u_1)f_2(v_1)f_2(u_2)f_2(v_2),$$

$$C_1(n_1,n_2) = n_2(n_2-1)(n_2-2)(n_1-1)\left(1-\overline{p_2^{(1)}+p_2^{(2)}}\right)^{n_2-3}$$
$$\times \left(1-\overline{p_1^{(1)}+p_1^{(2)}}\right)^{n_1-1} [f_2(u_1)f_2(v_1)f_2(u_2)f_1(v_2)$$
$$+ f_2(u_1)f_2(v_1)f_1(u_2)f_2(v_2)$$
$$+ f_2(u_1)f_1(v_1)f_2(u_2)f_2(v_2)$$
$$+ f_1(u_1)f_2(v_1)f_2(u_2)f_2(v_2)],$$

$$C_2(n_1,n_2) = n_1(n_1-1)n_2(n_2-1)\left(1-\overline{p_2^{(1)}+p_2^{(2)}}\right)^{n_2-2}$$
$$\times \left(1-\overline{p_1^{(1)}+p_1^{(2)}}\right)^{n_1-2} [f_1(u_1)f_1(v_1)f_2(u_2)f_2(v_2)$$
$$+ f_1(u_1)f_2(v_1)f_1(u_2)f_2(v_2)$$
$$+ f_2(u_1)f_1(v_1)f_2(u_2)f_1(v_2)$$
$$+ f_1(u_1)f_2(v_1)f_2(u_2)f_1(v_2)$$
$$+ f_2(u_1)f_1(v_1)f_1(u_2)f_2(v_2)$$
$$+ f_2(u_1)f_2(v_1)f_1(u_2)f_1(v_2)],$$

$$C_3(n_1,n_2) = n_1(n_1-1)(n_1-2)(n_2-1)\left(1-\overline{p_2^{(1)}+p_2^{(2)}}\right)^{n_2-1}$$
$$\times\left(1-\overline{p_1^{(1)}+p_1^{(2)}}\right)^{n_1-3}[f_1(u_1)f_1(v_1)f_1(u_2)f_2(v_2)$$
$$+f_1(u_1)f_1(v_1)f_2(u_2)f_1(v_2)$$
$$+f_1(u_1)f_2(v_1)f_1(u_2)f_1(v_2)$$
$$+f_2(u_1)f_1(v_1)f_1(u_2)f_1(v_2)],$$
$$C_4(n_1,n_2) = n_1(n_1-1)(n_1-2)(n_1-3)\left(1-\overline{p_2^{(1)}+p_2^{(2)}}\right)^{n_2}$$
$$\times\left(1-\overline{p_1^{(1)}+p_1^{(2)}}\right)^{n_1-4} f_1(u_1)f_1(v_1)f_1(u_2)f_1(v_2)$$

and

$$B(n_1,n_2) = C_0(n_1,n_2)+C_1(n_1,n_2)+C_2(n_1,n_2)+C_3(n_1,n_2)+C_4(n_1,n_2).$$

Note that $C_0(n_1,n_2)$ and $C_1(n_1,n_2)$ are proportional to the conditional probability that all four and any three RNN's are from π_2 respectively. $C_2(n_1,n_2)$ is proportional to the conditional probability that any two RNN's are from π_1 (or π_2). $C_3(n_1,n_2)$ and $C_4(n_1,n_2)$ are proportional to the conditional probability that any three and all four RNN are from π_1 respectively. Now assume that

$$p_i = \lim_{n\to\infty}\frac{n_i}{n_1+n_2},\, i=1,2, \text{ exist and are positive }. \tag{3.18}$$

THEOREM 3.1.1. *Suppose z_1 and z_2 are continuity points of both f_1 and f_2. Then the limiting conditional probability of deciding that z_1 and z_2 come from π_1, given $Z_1 = z_1, Z_2 = z_2$, is given by*

$$\Pi^{(1)}(z_1,z_2) = \lim_{n\to\infty}\pi^{(1)}(z_1,z_2;n_1,n_2)$$

$$=\widehat{\eta}_4+\widehat{\eta}_3+\frac{1}{2}\widehat{\eta}_2,$$

where

$$\widehat{\eta}_4 = \frac{(p_1^2 f_1(z_1)f_1(z_2))^2}{\{(p_1f_1(z_1)+p_2f_2(z_1))(p_1f_1(z_2)+p_2f_2(z_2))\}^2}, \tag{3.19}$$

$$\widehat{\eta}_3 = \frac{2p_1^3p_2(f_1^2(z_1)f_1(z_2)f_2(z_2) + f_1^2(z_2)f_1(z_1)f_2(z_1))}{\{(p_1f_1(z_1) + p_2f_2(z_1))(p_1f_1(z_2) + p_2f_2(z_2))\}^2} \tag{3.20}$$

and

$$\widehat{\eta}_2 = \frac{p_1^2p_2^2(f_1^2(z_1)f_2^2(z_2) + 4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2) + f_1^2(z_2)f_2^2(z_1))}{\{(p_1f_1(z_1) + p_2f_2(z_1))(p_1f_1(z_2) + p_2f_2(z_2))\}^2}. \tag{3.21}$$

PROOF. First note that by lemma 3.1.2 $U_1^{(1)} \xrightarrow{a.s} z_1$, $V_1^{(1)} \xrightarrow{a.s} z_1$ and $U_2^{(1)} \xrightarrow{a.s} z_2$, $V_2^{(1)} \xrightarrow{a.s} z_2$ as $n \to \infty$, so that dividing the numerator and denominator of $\frac{C_4(n_1,n_2)}{B((n_1,n_2)}$ by $(n_1+n_2)^4$, $\left(1-\overline{p_2^{(1)} + p_2^{(2)}}\right)^{n_2}$ and $\left(1-\overline{p_1^{(1)} + p_1^{(2)}}\right)^{n_1-4}$ we obtain,

$$\lim_{n\to\infty} \frac{C_4(n_1,n_2)}{B(n_1,n_2)} = \lim_{n\to\infty} \frac{\frac{n_1(n_1-1)(n_1-2)(n_1-3)}{(n_1+n_2)^4} f_1(u_1)f_1(v_1)f_1(u_2)f_1(v_2)}{[Denominator]} \tag{3.22}$$

with $[Denominator] = \frac{n_2(n_2-1)(n_2-2)n_2-3)}{(n_1+n_2)^4} \cdot \frac{(1-\overline{p_1^{(1)}+p_1^{(2)}})^4}{(1-\overline{p_2^{(1)}+p_2^{(2)}})^4}$

$\times f_2(u_1)f_2(v_1)f_2(u_2)f_2(v_2)$

$+ \frac{n_2(n_2-1)(n_2-2)(n_1-1)}{(n_1+n_2)^4} \cdot \frac{(1-\overline{p_1^{(1)}+p_1^{(2)}})^3}{(1-\overline{p_2^{(1)}+p_2^{(2)}})^3}$

$\times [f_2(u_1)f_2(v_1)f_2(u_2)f_1(v_2)$

$+ f_2(u_1)f_2(v_1)f_1(u_2)f_2(v_2) + f_2(u_1)f_1(v_1)f_2(u_2)f_2(v_2)$

$+ f_1(u_1)f_2(v_1)f_2(u_2)f_2(v_2)]$

$+ \frac{n_1(n_1-1)n_2(n_2-1)}{(n_1+n_2)^4} \cdot \frac{(1-\overline{p_1^{(1)}+p_1^{(2)}})^2}{(1-\overline{p_2^{(1)}+p_2^{(2)}})^2}$

$\times [f_1(u_1)f_1(v_1)f_2(u_2)f_2(v_2)$

$+ f_1(u_1)f_2(v_1)f_1(u_2)f_2(v_2) + f_2(u_1)f_1(u_1)f_2(u_2)f_1(v_2)$

$$+ f_1(u_1)f_2(v_1)f_2(u_2)f_1(v_2)$$

$$+ f_2(u_1)f_1(v_1)f_1(u_2)f_1(v_2) + f_2(u_1)f_2(v_1)f_1(u_2)f_1(v_2)]$$

$$+ \frac{n_1(n_1-1)(n_1-2)(n_2-1)}{(n_1+n_2)^4} \frac{(1-\overline{p_1^{(1)}+p_1^{(2)}})}{(1-\overline{p_2^{(1)}+p_2^{(2)}})}[f_1(u_1)f_1(v_1)f_1(u_2)f_2(v_2)$$

$$+ f_1(u_1)f_1(v_1)f_2(u_2)f_1(v_2) + f_1(u_1)f_2(v_1)f_1(u_2)f_1(v_2) + f_2(u_1)f_1(v_1)f_1(u_2)f_1(v_2)]$$

$$+ \frac{n_1(n_1-1)(n_1-2)(n_1-3)}{(n_1+n_2)^4} f_1(u_1)f_1(v_1)f_1(u_2)f_1(v_2)]. \tag{3.23}$$

Since $p_j^{(i)} \xrightarrow{\text{a.s}} 0$ as $n \to \infty$, therefore, by (3.18) and (3.23), we get

$$\lim_n [Denominator] \overset{a.s}{=} [(p_1^2(f_1(z_1)f_1(z_2))^2 + (p_2^2 f_2(z_1)f_2(z_2))^2$$

$$+ 2p_1^3 p_2(f_1^2(z_1)f_1(z_2)f_2(z_2) + f_1^2(z_2)f_1(z_1)f_2(z_1)) + p_1^2 p_2^2(f_1^2(z_1)f_2^2(z_2)$$

$$+ 4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2) + f_1^2(z_2)f_2^2(z_1))$$

$$+ 2p_1 p_2^3(f_1(z_2)f_2^2(z_1)f_2(z_2) + f_1(z_1)f_2^2(z_2)f_2(z_1))]$$

$$= (p_1^2 f_1(z_1)f_1(z_2))^2 + (p_2^2 f_2(z_1)f_2(z_2))^2 + p_1^2 p_2^2(f_1(z_1)f_2(z_2)$$

$$+ f_1(z_2)f_2(z_1))^2 + 2p_1^2 p_2^2 f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2)$$

$$+ 2p_1^3 p_2 f_1(z_1)f_1(z_2)(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1)) + 2p_1 p_2^3 f_2(z_1)f_2(z_2)(f_1(z_1)f_2(z_2)$$

$$+ f_1(z_2)f_2(z_1))$$

$$= \{p_1^2 f_1(z_1) f_1(z_2) + p_2^2 f_2(z_1) f_2(z_2) + p_1 p_2 (f_1(z_1) f_2(z_2) + f_1(z_2) f_2(z_1))\}^2$$

$$= \{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2) + p_2 f_2(z_2))\}^2. \qquad (3.24)$$

Thus by (3.22) and (3.24) we get

$$\lim_n \frac{C_4(n_1, n_2)}{B(n_1, n_2)} \overset{\text{a.s}}{=} \frac{(p_1^2 f_1(z_1) f_1(z_2))^2}{\{p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2) + p_2 f_2(z_2))\}^2}$$

$$= \widehat{\eta}_4. \qquad (3.25)$$

Similarly,

$$\lim_n \frac{C_3(n_1, n_2)}{B(n_1, n_2)} = \frac{2p_1^3 p_2 [f_1^2(z_1) f_1(z_2) f_2(z_2) + f_1^2(z_2) f_1(z_1) f_2(z_1)]}{\{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2) + p_2 f_2(z_2))\}^2}$$

$$= \widehat{\eta}_3 \quad a.s., \qquad (3.26)$$

and

$$\lim_n \frac{C_2(n_1, n_2)}{B(n_1, n_2)}$$

$$= \frac{p_1^2 p_2^2 [f_1^2(z_1) f_2^2(z_2) + 4 f_1(z_1) f_1(z_2) f_2(z_1) f_2(z_2) + f_1^2(z_2) f_2^2(z_1)]}{\{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2) + p_2 f_2(z_2))\}^2}$$

$$= \widehat{\eta}_2 \quad a.s.. \qquad (3.27)$$

Therefore by (3.15), (3.16) and (3.25) to (3.27) as $n \to \infty$

$$P_{n_1,n_2}^{(4,3)}(u_1, v_1, u_2, v_2, z_1, z_2)$$

$$\xrightarrow{\text{a.s.}} \frac{(p_1^2 f_1(z_1) f_1(z_2))^2}{\{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2 + p_2 f_2(z_2))\}^2}$$

$$+ \frac{2p_1^3 p_2 [f_1^2(z_1) f_1(z_2) f_2(z_2) + f_1^2(z_2) f_1(z_1) f_2(z_1)]}{\{(p_1 f_1(z_1) + p_2 f_2(z_1))(p_1 f_1(z_2) + p_2 f_2(z_2))\}^2}$$

$$= \widehat{\eta}_4 + \widehat{\eta}_3. \qquad (3.28)$$

and

$$P^{(2)}_{n_1,n_2}(u_1, v_1, u_2, v_2, z_1, z_2)$$

$$\xrightarrow{\text{a.s}} \frac{p_1^2 p_2^2[f_1^2(z_1)f_2^2(z_2) + 4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2) + f_1^2(z_2)f_2^2(z_1)]}{\{(p_1f_1(z_1) + p_2f_2(z_1))(p_1f_1(z_2) + p_2f_2(z_2))\}^2}$$

$$= \widehat{\eta}_2. \tag{3.29}$$

Using (3.9), (3.11) and (3.10), we get

$$\Pi^{(1)}(z_1, z_2) = \lim_n \pi^{(1)}(z_1, z_2; n_1, n_2)$$

$$= \lim_n EP^{(4,3)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, U_2^{(1)}, V_2^{(1)}; z_1, z_2)$$

$$+ \frac{1}{2}\lim_n EP^{(2)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, U_2^{(1)}, V_2^{(1)}; z_1, z_2), \tag{3.30}$$

so that by (3.28) to (3.30) and the Dominated Convergence theorem, we have

$$\Pi^{(1)}(z_1, z_2) = \widehat{\eta}_4 + \widehat{\eta}_3 + \frac{1}{2}\widehat{\eta}_2.$$

The proof is complete. □

Now we will derive the asymptotic TPMC for the 1st-stage RNN rule. The limiting PMC's of the first-stage RNN rule are given as follows:

$$\alpha_1^{(1)}(2) = \lim_{n\to\infty} P(Decide(z_1, z_2) \in \pi_2/(z_1, z_2) \in \pi_1)$$

$$= \int\int [1 - \Pi^{(1)}(z_1, z_2)]f_1(z_1)f_1(z_2)dz_1dz_2$$

$$= \int\int \left[\frac{(p_2^2 f_2(z_1)f_2(z_2))^2 + 2p_1p_2^3(f_1(z_2)f_2^2(z_1)f_2(z_2) + f_1(z_1)f_2^2(z_2)f_2(z_1))}{\{(p_1f_1(z_1) + p_2f_2(z_1))(p_1f_1(z_2) + p_2f_2(z_2))\}^2}\right.$$

$$+\frac{1}{2}\frac{p_1^2p_2^2(f_1^2(z_1)f_2^2(z_2)+4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2)+f_1^2(z_2)f_2^2(z_1))}{\{(p_1f_1(z_1)+p_2f_2(z_1))(p_2f_1(z_2)+p_2f_2(z_2))\}^2}\Bigg]$$

$$\times f_1(z_1)f_1(z_2)dz_1dz_2 \tag{3.31}$$

$$\alpha_2^{(1)}(2)=\lim_{n\to\infty}P(Decide(z_1,z_2)\in\pi_1/(z_1,z_2)\in\pi_2)$$

$$=\int\int\Pi^{(1)}(z_1,z_2)f_2(z_1)f_2(z_2)dz_1dz_2$$

$$=\int\int\Bigg[\frac{(p_1^2f_1(z_1)f_1(z_2))^2+2p_2^3p_2(f_1^2(z_1)f_1(z_2)f_2(z_2)+f_1^2(z_2)f_1(z_1)f_2(z_1))}{\{(p_1f_1(z_1)+p_2f_2(z_1))(p_1f_1(z_2)+p_2f_2(z_2))\}^2}$$

$$+\frac{1}{2}\frac{p_1^2p_2^2(f_1^2(z_1)f_2^2(z_2)+4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2)+f_1^2(z_2)f_2^2(z_1))}{\{(p_1f_1(z_1)+p_2f_2(z_1))(p_2f_1(z_2)+p_2f_2(z_2))\}^2}\Bigg]$$

$$\times f_2(z_1)f_2(z_2)dz_1dz_2. \tag{3.32}$$

By letting $p_i=\xi_i$, $i=1,2$, the prior probabilities, we get the limiting value of the total PMC of 1st-stage RNN rule as

$$R_{(2)}^{(1)}(2)=\xi_1\alpha_1^{(1)}(2)+\xi_2\alpha_2^{(1)}(2), \tag{3.33}$$

which, using (3.31) and (3.32), can be written as

$$R_{(2)}^{(1)}(2)=Er(\underset{\sim}{z})=\int\int r(\underset{\sim}{z})(\xi_1f_1(z_1)f_1(z_2)+\xi_2f_2(z_1)f_2(z_2))dz_1dz_2, \tag{3.34}$$

where $r(\underset{\sim}{z})$ can be written as

$$r(\underset{\sim}{z})=\frac{\xi_1f_1(z_1)f_1(z_2)}{(\xi_1f_1(z_1)f_1(z_2)+\xi_2f_2(z_1)f_2(z_2))}$$

$$\times\Bigg[\frac{(\xi_2^2f_2(z_1)f_2(z_2))^2+2\xi_1\xi_2^3(f_1(z_2)f_2^2(z_1)f_2(z_2)+f_1(z_1)f_2^2(z_2)f_2(z_1))}{\left\{\prod_{i=1}^2(\xi_1f_1(z_i)+\xi_2f_2(z_i))\right\}^2}$$

$$+\frac{1}{2}\frac{\xi_1^2\xi_2^2(f_1^2(z_1)f_2^2(z_2)+4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2)+f_1^2(z_2)f_2^2(z_1))}{\left\{\prod_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\right\}^2}\Bigg]$$

$$+\frac{\xi_2 f_2(z_1)f_2(z_2)}{\xi_1 f_1(z_1)f_1(z_2)+\xi_2 f_2(z_1)f_2(z_2)}$$

$$\Bigg[\frac{(\xi_1^2 f_1(z_1)f_1(z_2))^2+2\xi_1^3\xi_2(f_1^2(z_1)f_1(z_2)f_2(z_2)+f_1^2(z_2)f_1(z_1)f_2(z_1))}{\left\{\prod_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\right\}^2}$$

$$+\frac{1}{2}\frac{\xi_1^2\xi_2^2(f_1^2(z_1)f_2^2(z_2)+4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2)+f_1^2(z_2)f_2^2(z_1))}{\left\{\prod_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\right\}^2}\Bigg]. \quad (3.35)$$

We define

$$\eta_i^*(\underset{\sim}{z})=\frac{\xi_i f_i(z_1)f_i(z_2)}{\xi_1 f_1(z_1)f_1(z_2)+\xi_2 f_2(z_1)f_2(z_2)}, \quad (3.36)$$

and

$$\eta_i(z_j)=\frac{\xi_i f_i(z_j)}{\xi_1 f_1(z_j)+\xi_2 f_2(z_j)}\; i=1,2; \quad j=1,2. \quad (3.37)$$

Using (3.36) and (3.37), we can write (3.35) as

$$r(\underset{\sim}{z})=\eta_1^*(\underset{\sim}{z})[\eta_2^2(z_1)\eta_2^2(z_2)+2(\eta_1(z_2)\eta_2^2(z_1)\eta_2(z_2)+\eta_1(z_1)\eta_2^2(z_2)\eta_2(z_1))$$
$$+\frac{1}{2}(\eta_1^2(z_1)\eta_2^2(z_2)+4\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)+\eta_1^2(z_2)\eta_2^2(z_1))]$$

$$+\eta_2^*(\underset{\sim}{z})[\eta_1^2(z_1)\eta_1^2(z_2)+2(\eta_1^2(z_1)\eta_1(z_2)\eta_2(z_2)+\eta_1^2(z_2)\eta_1(z_1)\eta_2(z_1))$$

$$+\frac{1}{2}(\eta_1^2(z_1)\eta_2^2(z_2)+4\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)+\eta_1^2(z_2)\eta_2^2(z_1)]. \quad (3.38)$$

Now,

$$\eta_2^2(z_1)\eta_2^2(z_2)+2(\eta_1(z_2)\eta_2^2(z_1)\eta_2(z_2)+\eta_1(z_1)\eta_2^2(z_2)\eta_2(z_1))$$
$$+\frac{1}{2}(\eta_1^2(z_1)\eta_2^2(z_2)+4\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)+\eta_1^2(z_2)\eta_2^2(z_1))$$
$$=2(\eta_1(z_2)\eta_2^2(z_1)\eta_2(z_2)+\eta_1(z_1)\eta_2^2(z_2)\eta_2(z_1))$$

$$+\frac{1}{2}[\eta_2^2(z_1)(\eta_1^2(z_2)+\eta_2^2(z_2))+\eta_2^2(z_2)(\eta_1^2(z_1)+\eta_2^2(z_1))]$$
$$+2\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)$$
$$=\eta_2(z_1)\eta_2(z_2)(\eta_1(z_2)\eta_2(z_1)+\eta_1(z_1)\eta_2(z_2))$$
$$+\frac{1}{2}[\eta_2^2(z_1)(\eta_1^2(z_2)+\eta_2^2(z_2)+2\eta_1(z_2)\eta_2(z_2))$$
$$+\eta_2^2(z_1)(\eta_1^2(z_1)+\eta_2^2(z_1)+2\eta_1(z_1)\eta_2(z_1))]$$
$$+2\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)$$
$$=\eta_2(z_1)\eta_2(z_2)(\eta_1(z_2)\eta_2(z_1)+\eta_1(z_1)\eta_2(z_2))$$
$$+2\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_2(z_2)$$
$$+\frac{1}{2}[\eta_2^2(z_2)+\eta_2^2(z_1)]$$
$$=\eta_2(z_1)\eta_2(z_2)(\eta_1(z_2)\eta_2(z_1)+\eta_1(z_1)\eta_1(z_2)$$
$$+\eta_1(z_1)\eta_2(z_2)+\eta_1(z_1)\eta_1(z_2))$$
$$+\frac{1}{2}[\eta_2^2(z_2)+\eta_2^2(z_1)]$$
$$=\eta_2(z_1)\eta_2(z_2)[\eta_1(z_2)+\eta_1(z_1)]+\frac{1}{2}[\eta_2^2(z_2)+\eta_2^2(z_1)]; \qquad (3.39)$$

similarly,

$$\eta_1^2(z_1)\eta_1^2(z_2)+2(\eta_1^2(z_1)\eta_1(z_2)\eta(z_2)+\eta_1^2(z_2)\eta_1(z_1)\eta_2(z_1))$$
$$+\frac{1}{2}(\eta_1^2(z_1)\eta_2^2(z_2)+4\eta_1(z_1)\eta_1(z_2)\eta_2(z_1)\eta_1(z_2)+\eta_1^2(z_2)\eta_2^2(z_1))$$
$$=\eta_1(z_1)\eta_1(z_2)[\eta_2(z_2)+\eta_2(z_1)]+\frac{1}{2}[\eta_1^2(z_1)+\eta_1^2(z_2)]. \qquad (3.40)$$

We write $r(\underset{\sim}{z})$ from (3.38), (3.39) and (3.40) as

$$r(\underset{\sim}{z})=\eta_1^*(\underset{\sim}{z})[\eta_2(z_1)\eta_2(z_2)[\eta_1(z_1)+\eta_1(z_2)]+\frac{1}{2}(\eta_2^2(z_1)+\eta_2^2(z_2))]$$

$$+\eta_2^*(\underset{\sim}{z})[\eta_1(z_1)\eta_1(z_2)[\eta_2(z_1)+\eta_2(z_2)]+\frac{1}{2}(\eta_1^2(z_1)+\eta_1^2(z_2))]. \qquad (4.41)$$

Now

$$\eta_2(z_1)\eta_2(z_2)[\eta_1(z_1)+\eta_1(z_2)]+\frac{1}{2}[\eta_2^2(z_1)+\eta_2^2(z_2)]$$
$$=\frac{1}{2}[\eta_2(z_1)(\eta_2(z_1)+\eta_2(z_2)\eta_1(z_1))+\eta_2(z_2)(\eta_2(z_2)+\eta_2(z_1)\eta_1(z_2))]$$
$$+\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)$$
$$=\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))(1-\eta_1(z_1)\eta_1(z_2))+\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_2)\eta_2(z_2)$$
$$=\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))+\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)$$
$$-\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2); \tag{3.42}$$

similarly,

$$\eta_1(z_1)\eta_1(z_2)[\eta_2(z_1)+\eta_2(z_2)]+\frac{1}{2}[\eta_1^2(z_1)+\eta_1^2(z_2)]$$
$$=\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))+\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2)$$
$$-\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2). \tag{3.43}$$

Therefore, in view of (3.41) to (3.43), we get

$$r(\underset{\sim}{z})=\eta_1^*(\underset{\sim}{z})\cdot\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))+\eta_2^*(\underset{\sim}{z})\cdot\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))$$

$$+\eta_1^*(\underset{\sim}{z})[\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))-\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2)]$$

$$+\eta_2^*(\underset{\sim}{z})[\frac{1}{2}(\eta_2(z_1)+\eta_2(z_2))-\frac{1}{2}(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)]. \tag{3.44}$$

Thus, (3.34) and (3.44) implies

$$R^{(1)}_{(2)}(2)=Er(\underset{\sim}{z})$$
$$=\frac{1}{2}\Big[\int\int\xi_1 f_1(z_1)f_1(z_2)(\eta_2(z_1)+\eta_2(z_2))dz_1dz_2$$
$$+\int\int\xi_2 f_2(z_1)f_2(z_2)(\eta_1(z_1)+\eta_1(z_2))dz_1dz_2\Big]$$

$$+\frac{1}{2}\int\int(\xi_1 f_1(z_1)f_1(z_2)-\xi_2 f_2(z_1)f_2(z_2))$$
$$\times((\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)-(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2)))$$
$$\times dz_1dz_2. \tag{3.45}$$

Now

$$(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)=\left[\frac{\xi_1 f_1(z_1)}{\xi_1 f_1(z_1)+\xi_2 f_2(z_1)}+\frac{\xi_1 f_1(z_2)}{\xi_1 f_1(z_2)+\xi_2 f_2(z_2)}\right]$$
$$\times\frac{\xi_2^2 f_2(z_1)f_2(z_2)}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))(\xi_1 f_1(z_2)+\xi_2 f_2(z_2))}$$

$$=\frac{[2\xi_1^2 f_1(z_1)f_1(z_2)+\xi_1\xi_2(f_1(z_1)f_2(z_2)+f_1(z_2)f_2(z_1))]\xi_2^2 f_2(z_1)f_2(z_2)}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))^2(\xi_1 f_1(z_2)+\xi_2 f_2(z_1))^2},$$

$$(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2)$$
$$=\left[\frac{\xi_2 f_2(z_1)}{\xi_1 f_1(z_1)+\xi_2 f_2(z_1)}+\frac{\xi_2 f_2(z_2)}{\xi_1 f_1(z_2)+\xi_2 f_2(z_2)}\right]$$
$$\times\frac{\xi_1^2 f_1(z_1)f_1(z_2)}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))(\xi_1 f_1(z_2)+\xi_2 f_2(z_2))}$$

$$=\frac{[2\xi_2^2 f_2(z_1)f_2(z_2)+\xi_1\xi_2(f_1(z_1)f_2(z_2)+f_1(z_2)f_2(z_1))]\xi_1^2 f_1(z_1)f_1(z_2)}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))^2(\xi_1 f_1(z_2)+\xi_2 f_2(z_1))^2}.$$

Therefore,

$$(\eta_1(z_1)+\eta_1(z_2))\eta_2(z_1)\eta_2(z_2)-(\eta_2(z_1)+\eta_2(z_2))\eta_1(z_1)\eta_1(z_2)$$

$$=\frac{1}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))^2(\xi_1 f_1(z_2)+\xi_2 f_2(z_2))^2}[\xi_1\xi_2^3 f_1(z_1)f_2(z_1)f_2^2(z_2)$$

$$+\xi_1\xi_2^3 f_1(z_2)f_2^2(z_1)f_2(z_2)-\xi_1^3\xi_2 f_1^2(z_1)f_1(z_2)f_2(z_2)-\xi_1^3\xi_2 f_1(z_1)f_1^2(z_2)f_2(z_1)]$$

$$=\frac{1}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))^2(\xi_1 f_1(z_2)+\xi_2 f_2(z_2))^2}[\xi_1\xi_2 f_1(z_1)f_2(z_2)\times$$

$$(\xi_2^2 f_2(z_1)f_2(z_2)-\xi_1^2 f_1(z_1)f_1(z_2))+\xi_1\xi_2 f_1(z_2)f_2(z_1)$$
$$\times(\xi_2^2 f_1(z_1)f_2(z_2)-\xi_1^2 f_1(z_1)f_1(z_2))]$$

$$=-\frac{\xi_1\xi_2(\xi_1^2 f_1(z_1)f_1(z_2)-\xi_2^2 f_2(z_1)f_2(z_2))(f_1(z_1)f_2(z_2)+f_1(z_2)f_2(z_1))}{(\xi_1 f_1(z_1)+\xi_2 f_2(z_1))^2(\xi_1 f_1(z_2)+\xi_2 f_2(z_2))^2}. \tag{3.46}$$

Also note that

$$\begin{aligned}
&\frac{1}{2}\Big[\int\int \xi_1 f_1(z_1)f_1(z_2)(\eta_2(z_1)+\eta_2(z_2))dz_1dz_2\\
&\quad+\int\int \xi_2 f_2(z_1)f_2(z_2)(\eta_1(z_1)+\eta_1(z_2))dz_1dz_2\Big]\\
&\quad=\frac{1}{2}[\int \xi_1 f_1(z_2)\eta_2(z_1)dz_1+\int \xi_1 f_1(z_2)\eta_2(z_2)dz_2\\
&\quad+\int \xi_2 f_2(z_1)\eta_1(z_1)dz_1+\int \xi_2 f_2(z_2)\eta_1(z_2)dz_2]\\
&\quad=\frac{1}{2}[4\int \eta_1(z)\eta_2(z)(\xi_1 f_1(z)+\xi_2 f_2(z))dz]\\
&\quad=2E\eta_1(z)\eta_2(z)\\
&\quad=R_{(2)}^{(1)}(1).
\end{aligned} \tag{3.47}$$

So from (3.45), (3.46) and (3.47) we obtain

$$\begin{aligned} R^{(1)}_{(2)}(2) = R^{(1)}_{(2)}(1) \\ - \frac{\xi_1\xi_2}{2}\int\int(\xi_1 f_1(z_1)f_1(z_2) - \xi_2 f_2(z_1)f_2(z_2)) \\ \times \frac{(\xi_1^2 f_1(z_1)f_1(z_2) - \xi_2^2 f_2(z_1)f_2(z_2))(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{(\xi_1 f_1(z_1) + \xi_2 f_2(z_1))^2(\xi_1 f_1(z_2) + \xi_2 f_2(z_2))^2} \\ \times dz_1 dz_2. \end{aligned} \tag{3.48}$$

We shall now obtain a bound for the limiting TPMC $R^{(1)}_{(2)}(2)$.

THEOREM 3.1.2. *Let the Bayesian model with prior probability ξ_i for the population $\pi_i, i = 1,2$ hold. Then under the condition $\xi_1 = \xi_2$, we have*

$$R^{(1)}_{(2)}(2) \le R^{(1)}_{(2)}(1) \le 2R^{*}_{(2)}(1)(1 - R^{*}_{(2)}(1)), \tag{3.49}$$

where $R^{*}_{(2)}(1)$ is the Bayes risk given by (2.10).

PROOF. When $\xi_1 = \xi_2 = 1/2$, (3.48) becomes

$$\begin{aligned} R^{(1)}_{(2)}(2) = R^{(1)}_{(2)}(1) - \frac{1}{4}\int\int(f_1(z_1)f_1(z_2) - f_2(z_1)f_2(z_2))^2 \\ \frac{(f_1(z_1)f_2(z_2) + f_1(z_2)f_2(z_1))}{(f_1(z_1) + f_2(z_1))^2(f_1(z_2) + f_2(z_2))^2} \\ \times dz_1 dz_2 \end{aligned} \tag{3.50}$$

which implies $R^{(1)}_{(2)}(2) \le R^{(1)}_{(2)}(1)$; (see (3.50)).

By (3.47) and using the fact that $r^*(z) = \min\{\eta_1(z), \eta_2(z)\}$ which is symmetric in η_1 (see (2.8)) and Jensen's inequality we have

$$\begin{aligned} R^{(1)}_{(2)}(1) &= 2E\eta_1(z)\eta_2(z) \\ &= 2Er^*(z)(1 - r^*(z)) \\ &= 2[R^{*}_{(2)}(1) - E(r^{*2}(z))] \\ &\le 2[R^{*}_{(2)}(1) - R^{*2}_{(2)}(1)] \\ &= 2R^{*}_{(2)}(1)(1 - R^{*}_{(2)}(1)). \end{aligned}$$

This completes the proof. □

We now give an example to illustrate the inequality (3.49) and another possible upper bound for $R^{(1)}_{(2)}(2)$.

Example 3.1. Let $f_1(x) = 2x, 0 < x < 1$; $f_2(y) = 2 - 2y, 0 < y < 1$; and $\xi_1 = \xi_2 = 1/2$
write (3.34) as

$$R^{(1)}_{(2)}(2)$$

$$= \int\int \frac{\xi_1[(\xi_2^2 f_2(z_1)f_2(z_2))^2 + 2\xi_1\xi_2^3(f_1(z_2)f_2^2(z_1)f_2(z_2) + f_1(z_1)f_2^2(z_2)f_2(z_1)]}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}$$

$$\times f_1(z_1)f_2(z_2)dz_1dz_2$$

$$+ \int\int \frac{\xi_2[(\xi_1^2 f_1(z_1)f_1(z_2))^2 + 2\xi_1^3\xi_2(f_1^2(z_1)f_1(z_2)f_2(z_2) + f_1^2(z_2)f_2(z_1)f_2(z_1)]}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}$$

$$\times f_2(z_1)f_2(z_2)dz_1dz_2$$

$$+ \frac{1}{2}\int\int \frac{\xi_1^2\xi_2^2(f_1^2(z_1)f_2^2(z_2) + 4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_2) + f_1^2(z_2)f_2^2(z_1))}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}$$

$$\times (\xi_1 f_1(z_1)f_1(z_2) + \xi_2 f_2(z_1)f_2(z_2))dz_1dz_2,$$

$$= I_1 + I_2 + I_3. \tag{3.51}$$

Now

$$\xi_1 f_1(z_1) + \xi_2 f_2(z_1) = \frac{1}{2}2z_1 + \frac{1}{2}\cdot 2\cdot(1 - z_1) = 1,$$

$$\xi_1 f_1(z_2) + \xi_2 f_2(z_2) = \frac{1}{2}\cdot 2z_2 + \frac{1}{2}\cdot 2\cdot(1 - z_2) = 1.$$

Calculations of I_3:

$$\xi_1 f_1(z_1)f_1(z_2)+\xi_2 f_2(z_1)f_2(z_2)=2(z_1z_2+(1-z_1)(1-z_2))$$

$$\xi_1^2\xi_2^2(f_1^2(z_1)f_2^2(z_2)+4f_1(z_1)f_1(z_2)f_2(z_1)f_2(z_3)+f_1^2(z_2)f_2^2(z_1))$$

$$=\frac{1}{4}\cdot\frac{1}{4}(4z_1^2\cdot 4(1-z_2)^2+4\cdot 4z_1z_2\cdot 4(1-z_1)(1-z_2)+4z_2^2\cdot 4(1-z_1)^2)$$

$$=\left[z_1^2(1-z_2)^2+4z_1z_2(1-z_1)(1-z_2)+z_1z^2(1-z_1)^2\right],$$

$$I_3=\int_0^1\int_0^1\left[z_1^3z_3(1-z_2)^2+4z_1^2z_2^2(1-z_1)(1-z_2)+z_1z_2^3(1-z_1)^2\right.$$

$$\left.+z_1^2(1-z_1)(1-z_2)^3+4z_1z_2(1-z_1)^2(1-z_2)^2+(1-z_1)^3z_2^2(1-z_2)\right]$$

$$=\left[\frac{1}{12.4}+\frac{1}{12.3}+\frac{1}{12.4}+\frac{1}{12.4}+\frac{1}{12.3}+\frac{1}{12.4}\right]$$

$$=\left[\frac{4}{12.4}+\frac{2}{12.3}\right]=\frac{1}{2}\left[\frac{1}{12}+\frac{1}{6.3}\right]=\frac{1}{12}\left[\frac{1}{2}+\frac{1}{3}\right]$$

$$=\frac{5}{36}.$$

Calculation of I_1:

$$\xi_1\left[\left(\xi_2^2 f_2(z_1)f_2(z_2)\right)^2+2\xi_1\xi_2^3\left(f_1(z_2)f_2^2(z_1)f_2(z_2)+f_1(z_1)f_2^2(z_2)f_2(z_1)\right)\right]$$

$$\times f_1(z_1)f_1(z_2)$$

$$=\frac{1}{2}\left[\frac{1}{4}\cdot\frac{1}{4}4(1-z_1)^2\cdot 4(1-z_2)^2+2\cdot\frac{1}{2}\cdot\frac{1}{8}(2z_2\cdot 4(1-z_1)^2\cdot 2(1-z_2)\right.$$

$$\left.+2z_1\cdot 4(1-z_2)\cdot 2(1-z_1))\right]\times 4z_1z_2$$

$$=2\left[z_1(1-z_1)^2\cdot z_2(1-z_2)^2+2z_1(1-z_1)^2\cdot z_2^2(1-z_2)\right.$$

$$\left.+2z_1^2(1-z_1)\cdot z_2(1-z_2)^2\right]$$

$$I_1=2\int_0^1\int_0^1\left[z_1(1-z_1)^2\cdot z_2(1-z_2)^2+2z_1(1-z_1)^2\cdot z_2^2(1-z_2)\right.$$

$$\left.+2z_1^2(1-z_1)\cdot z_2(1-z_2)^2\right]dz_1dz_2$$

$$=2\left(\frac{1}{12.12}+\frac{1}{12.6}+\frac{1}{12.6}\right)=\frac{2}{12.12}[1+2+2]=\frac{5}{72}.$$

Calculation of I_2:

$$\xi_2\Big[\Big(\xi_1^2 f_1(z_1)f_1(z_2)\Big)^2 + 2\xi_1^3\xi_2\Big(f_1^2(z_1)f_1(z_2)f_2(z_2) + f_1^2(z_2)f_1(z_1)f_2(z_1)\Big)\Big]$$
$$\times f_1(z_1)f_2(z_2)$$
$$= \frac{1}{2}\Big[\frac{1}{16}\cdot 4z_1^2\cdot 4z_2^2 + 2\cdot\frac{1}{8}\cdot\frac{1}{2}(4z_1^2\cdot 2z_2\cdot 2(1-z_2)$$
$$+ 4z_2^2\cdot 2(1-z_1)\cdot 2z_1)\Big]4(1-z_1)(1-z_2)$$
$$= 2\Big[z_1^2(1-z_1)z_2^2(1-z_2) + 2z_1^2(1-z_1)\cdot z_2(1-z_2)^2$$
$$+ 2z_1(1-z_1)^2\cdot z_2(1-z_2)^2\Big]$$
$$I_2 = 2\int_0^1\int_0^1 \Big[z_1^2(1-z_1)\cdot z_2^2(1-z_2) + 2z_1^2(1-z_1)\cdot z_2(1-z_2)^2$$
$$+ 2z_1(1-z_1)^2\cdot z_2^2(1-z_2)\Big]dz_1dz_2$$
$$= 2\Big[\frac{1}{12.12} + \frac{2}{12.12} + \frac{2}{12.12}\Big] = \frac{5}{72}.$$

Therefore by (3.51) the limiting total PMC is

$$R_{(2)}^{(1)}(2) = I_1 + I_2 + I_3$$
$$= \frac{5}{72} + \frac{5}{72} + \frac{5}{36} = \frac{10}{72} + \frac{10}{72} = \frac{5}{18}. \qquad (3.52)$$

The Bayes risk for $(K = 2)$ from (2.61) is

$$R_{(2)}^{*}(2) = \frac{1}{6}.$$

The Bayes risk for $(K = 1)$ from (2.62) is

$$R_{(2)}^{*}(1) = \frac{1}{4}.$$

The limiting total PMC $(k = 1)$ is given by

$$
\begin{aligned}
R_{(2)}^{(1)}(1) &= 2\int \frac{\xi_1\xi_2 f_1(z) f_2(z)}{(\xi_1 f_1(z) + \xi_2 f_2(z))} dz \\
&= 2\int_0^1 z(1-z)dz = 2\int_0^1 (z - z^2)dz \\
&= 2[1/2 - 1/3] = 1/3.
\end{aligned} \tag{3.53}
$$

Therefore, $R_{(2)}^{(1)}(2) = \frac{5}{18} < R_{(2)}^{(1)}(1) = \frac{1}{3} = \frac{6}{18}$,. Further it is easy to see that $R_{(2)}^{*}(2) = \frac{6}{36}$ so that $R_{(2)}^{(1)}(2) = 2R_{(2)}^{(*)}(2)(1 - R_{(2)}^{*}(2))$.

3.2 Asymptotic PMC's of the 1st-stage RNN Rule when K observations are classified between two populations.

Suppose, we have K-observations $(Z_1, Z_2, \ldots, Z_K)$ to be classified between π_1 and π_2. Let $Z_{(i)}$, $i = 1, \ldots K$ be the order statistics of Z_i's.

LEMMA 3.2.1. *If $M/n \to 0$ as $n \to \infty$, the probability that there are at least $2M$ observations between $Z_{(\ell)}$ and $Z_{(\ell+1)}$ $\forall \ell = 1, 2, \ldots, K$ and at least M observations to the left of $Z_{(1)}$ and to the right of $Z_{(K)}$ is one.*

LEMMA 3.2.2. *Given that $Z_1, \ldots, Z_K$ are distributed as f_1, then $U_\ell^{(M)}$ and $V_\ell^{(M)}$ converges to Z_ℓ as $\forall\, \ell = 1, 2, \ldots, K$ as $n_1 \to \infty$ and $M/n \to 0$.*

The proofs of Lemma 3.2.1 and Lemma 3.2.2 are similar to those of Lemma 3.1.1 and Lemma 3.1.2 respectively and hence are omitted.

Let $A_\ell^{(1)}$ be the event that both $U_\ell^{(1)}$ and $V_\ell^{(1)}$, the left-hand and the right hand neighbors of Z_ℓ, are well defined at the 1st stage, $\forall\, \ell = 1, 2, \ldots, K$.

The conditional probability that $\underset{\sim}{Z} = (Z_1, \ldots Z_K)$ is from π_1 using the one-stage RNN rule, given $Z_1 = z_1, \ldots, Z_K = z_K$, is given by

$$
\Pi^{(1)}(z_1, \ldots z_K; n_1, n_2) = E(\phi^{(1)}/Z_1 = z_1, Z_2 = z_2, \ldots, Z_K = z_K) \tag{3.54}
$$

$$
\begin{aligned}
&= E(\phi^{(1)} I_{(A_1^{(1)} A_2^{(1)} \ldots A_K^{(1)})}/Z_1 = z_1, \ldots Z_K = z_K) \\
&+ E(\phi^{(1)} I_{(A_1^{(1)} A_2^{(1)} \ldots A_K^{(1)})^c}/Z_1 = z_1, \ldots, Z_K = z_K),
\end{aligned} \tag{3.55}
$$

where we have used (3.8) to get the second equality above. However,

$$\begin{aligned}
&E(\phi^{(1)}I_{(A_1^{(1)}A_2^{(1)}\ldots A_K^{(1)})^c}/Z_1=z_1,\ldots,Z_K=z_K)\\
&\qquad = E(\phi^{(1)}I_{(\bigcup_{i=1}^K A_i^{(1)c})}/Z_1=z_1,\ldots,Z_K=z_K)\\
&\qquad \le P(\bigcup_{i=1}^K A_i^{(1)c}/Z_1=z_1,\ldots,Z_K=z_K)\\
&\qquad \le \sum_{i=1}^K P(A_i^{(1)c}/Z_i=z_i)\to 0,
\end{aligned} \tag{3.56}$$

the above convergence following from Lemma 3.2.1.

Also, note that

$$E(\phi^{(1)}I_{(A_1^{(1)}A_2^{(1)}\ldots A_K^{(1)})}/Z_1=z_1,\ldots,Z_K=z_K)$$

$$=P(\{\phi^{(1)}=1, \quad A_1^{(1)}A_2^{(1)}\ldots A_K^{(1)}/Z_1=z_1,\ldots,Z_K=z_K)$$

$$+\frac{1}{2}P(\{\phi^{(1)}=1/2\}\cap A_1^{(1)}A_2^{(1)}\ldots A_K^{(1)}/Z_1=z_1,\ldots,Z_K=z_K)$$

$$=EP_{n_1,n_2}^{(2K,\ldots,K+1)}(U_1^{(1)},V_1^{(1)},\ldots,U_K^{(1)},V_K^{(1)};z_1,\ldots,z_K)$$

$$+\frac{1}{2}EP_{n_1,n_2}^{(K)}(U_1^{(1)},V_2^{(1)},\ldots,U_K^{(1)}V_K^{(1)};z_1,\ldots z_K), \tag{3.57}$$

where

$$\begin{aligned}
&P_{n_1,n_2}^{(2K,\ldots,K+1)}(u_1,v_1,\ldots u_K,v_K;z_1,\ldots,z_K)\\
&=P(\phi^{(1)}=1/U_1^{(1)}=u_1,V_1^{(1)}=v_1,\ldots U_K^{(1)}=u_K,V_K^{(1)}=v_K;A_1^{(1)},\ldots A_K^{(1)}),
\end{aligned}$$

$$\begin{aligned}
&P_{n_1,n_2}^{(K)}(u_1,v_1,\ldots u_K,v_K;z_1,\ldots,z_K)\\
&=P(\phi^{(1)}=\frac{1}{2}/U_1^{(1)}=u_1,V_1^{(1)}=v_1,\ldots U_K^{(1)}=u_K,V_K^{(1)}=v_K;A_1^{(1)},\ldots A_K^{(1)}),
\end{aligned}$$

$$P^{(2K,\ldots,K+1)}_{n_1,n_2}(u_1,v_1,\ldots u_K,v_K;z_1,\ldots,z_K)$$

$$= \frac{C_{2K}(n_1,n_2)+C_{2K-1}(n_1,n_2)+\ldots+C_{K+1}(n_1,n_2)}{B(n_1,n_2)}, \qquad (3.58)$$

$$P^{(K)}_{n_1,n_2}(u_1,v_1,\ldots u_K,v_K;z_1,\ldots,z_K) = \frac{C_K(n_1,n_2)}{B(n_1,n_2)}, \qquad (3.59)$$

$$P^{(K-1,\ldots,0)}_{n_1,n_2}(u_1,v_1,\ldots u_K,v_K;z_1,\ldots,z_K)$$

$$= \frac{C_{K-1}(n_1,n_2)+C_{K-2}(n_1,n_2)+\ldots+C_0(n_1,n_2)}{B(n_1,n_2)} \qquad (3.60)$$

with

$$B(n_1,n_2) = \sum_{i=0}^{2K} C_{2K-i}(n_1,n_2). \qquad (3.61)$$

Let $p_1^{(i)} = F_1(v_i) - F_1(u_i)$ and $p_2^{(i)} = F_2(v_i) - F_2(u_i)$ for $i = 1,2,\ldots,K$. Now the C_{2K-i}'s for $i = 0,1,\ldots,2K$ are given by

$$C_{2K}(n_1,n_2) = n_1(n_1-1)\ldots(n_1-\overline{2K-1})\cdot(1-\sum_{i=1}^{K} p_1^{(i)})^{n_1-2K}$$
$$\left(1-\sum_{i=1}^{k} p_2^{(1)}\right)^{n_2} \times \prod_{i=1}^{K} f_1(u_i)f_1(v_i),$$

$$C_{2K-1}(n_1,n_2) = n_1(n_1-1)\dots(n_1-\overline{2K-2})n_2\times$$

$$(1-\sum_{i=1}^{K}p_1^{(i)})^{n_1-\overline{2K-1}}(1-\sum_{i=1}^{K}p_2^{(i)})^{n_2-1}[\prod_{i=2}^{K}f_1(u_i)f_1(v_i)(f_2(u_1f_1(v_1)+f_1(u_1)f_2(v_1))$$

$$+\prod_{\substack{i=1\\ i\neq 2}}^{K}f_1(u_i)f_1(v_j)(f_2(u_2)f_1(v_2)+f_1(u_2)f_2(v_2))+\dots$$

$$+\prod_{\substack{i=1\\ i\neq j}}^{K}f_1(u_i)f_1(v_i)(f_2(u_j)f_1(v_j)+f_1(u_j)f_2(v_j))$$

$$+\dots+\prod_{i=1}^{K-1}f_1(u_i)f_1(v_i)(f_2(u_K)f_1(v_K)+f_1(u_K)f_2(v_K)),$$

$$C_{2K-i}(n_1,n_2) = n_1(n_1-1)\dots(n_1-(2K-\overline{i+1}))n_2(n_2-1)\dots(n_2-\overline{i-1})$$

$$\times(1-\sum_{j=1}^{K}p_1^{(j)})^{n_1-(2K-i)}(1-\sum_{j=1}^{K}p_2^{(j)})^{n_2-i}$$

$$\times\sum_{r\in S_{2K-i}}f_{r_1}(u_1)f_{r_2}(v_2)f_{r_3}(u_2)f_{r_4}(v_2)\dots f_{r_{2K-1}}(u_K)f_{r_{2K}}(v_K)$$

where $r=(r_1,r_2,\dots r_{2K})$ is an element of S_{2K-i} and

S_{2K-i} = set of all possible permuations of $(\underbrace{1,1,\dots 1}_{2K-i},\underbrace{2,2,\dots 2}_{i})$

$$\sum_{r\in S_{2K-i}} = \text{sum over } \frac{2K!}{(2K-i)!i!} \text{ terms} \quad i=0,1,\dots,2K.$$

$$C_1(n_1,n_2) = n_1n_2(n_2-1)\dots(n_2-\overline{2K-2})$$

$$\times(1-\sum_{i=1}^{K}p_1^{(i)})^{n_1-1}(1-\sum_{i=1}^{K}p_2^{(i)})^{n_2-\overline{2K-1}}$$

$$\times[\prod_{i=2}^{K}f_2(u_i)f_2(v_i)(f_1(u_1)f_2(v_1)+f_2(u_1)f_1(v_1))+$$

$$\prod_{\substack{i=1\\ i\neq 2}}^{K}f_2(u_i)f_2(v_i)(f_1(u_2)f_2(v_2)+f_2(u_2)f_1(u_2))$$

$$+\dots+\prod_{i=1}^{K-1}f_2(u_i)f_2(v_i)(f_1(u_K)f_2(v_K)+f_2(u_K)f_1(u_K)]$$

and

$$C_0(n_1,n_2) = n_2(n_2-1)\ldots(n_2-\overline{2K-1})(1-\sum_{i=1}^{K} p_1^{(i)})^{n_1}$$

$$(1-\sum_{i=1}^{K} p_2^{(i)})^{n_2-2K} \times \prod_{i=1}^{K} f_2(u_i)f_2(v_i).$$

Let

$$p_i = \lim_{n\to\infty} \frac{n_i}{(n_1+n_2)}, \quad n = min(n_1,n_2) \text{ and} \tag{3.62}$$

assume $0 < p_i < 1$ for $i = 1,2$.

THEOREM 3.2.1. *Suppose $z_1, z_2, \ldots z_K$ are continuity points of both f_1 and f_2. Then*

$$\Pi^{(1)}(z_1,z_2,\ldots,z_K) = \lim_{n\to\infty} \Pi^{(1)}(z_1,z_2,\ldots,z_K;n_1,n_2)$$
$$= \sum_{i=0}^{K-1} \widehat{\eta}_{2K-i} + \frac{1}{2}\widehat{\eta}_K$$

where

$$\widehat{\eta}_{2K} = p_i^{2K} \prod_{i=1}^{K} f_1^2(z_i)/\{\prod_{i=1}^{K}(p_1f_1(z_i)+p_2f_2(z_i))\}^2, \tag{3.63}$$

$$\widehat{\eta}_{2K-1} = \frac{2p_1^{2K-1}p_2}{\{\prod_{i=1}^{K}(p_1f_1(z_i)+p_2f_2(z_i))\}^2}\Big[f_1(z_1)f_2(z_1)\prod_{i=2}^{K} f_1^2(z_i)$$
$$+ f_1(z_2)f_2(z_2)\prod_{\substack{i=1\\ i\neq 2}}^{K} f_1^2(z_i) + \ldots + f_1(z_K)f_2(z_K)\prod_{i=1}^{K-1} f_1^2(z_i)\Big] \tag{3.64}$$

$$\widehat{\eta}_{2K-i} = \frac{p_1^{2K-i}p_2^i \sum_{r\in S_{2K-i}} f_{r_1}(z_1)f_{r_2}(z_1)\ldots f_{r_{2K-1}}(z_K)f_{r_{2K}}(z_K)}{\{\prod_{i=1}^{K}(p_1f_1(z_i)+p_2f_2(z_i))\}^2} \tag{3.65}$$

and

$$\widehat{\eta}_K = \frac{p_1^K p_2^K \sum_{r\in S_{2K-K}} f_{r_1}(z_1)f_{r_2}(z_1)f_{r_3}(z_2)f_{r_4}(z_2)\ldots f_{r_{2K-1}}(z_K)f_{r_{2K}}(z_K)}{\{\prod_{i=1}^{K}(p_1f_1(z_i)+p_2f_2(z_i))\}^2}. \tag{3.66}$$

PROOF. First note that $U_\ell^{(1)} \xrightarrow{\text{a.s.}} z_\ell, V_\ell^{(1)} \xrightarrow{\text{a.s}} z_\ell, \forall \ell = 1,2,\ldots,K$.

Then

$$\lim_{n\to\infty} \frac{C_{2K}(n_1,n_2)}{B(n_1,n_2)} \overset{\text{a.s.}}{=} \frac{p_1^{2K}\prod_{i=1}^{K} f_1^2(z_i)}{\{\prod_{i=1}^{K}(p_1f_1(z_i)+p_2f_2(z_i))\}^2} = \widehat{\eta}_{2K},$$

as

$$\sum_{i=1}^{K} p_1^{(i)} = \sum_{i=1}^{K}[F_1(v_i) - F_1(u_i)] \overset{\text{a.s.}}{\to} 0$$

$$\sum_{i=1}^{K} p_2^{(i)} = \sum_{i=1}^{K}(F_2(v_i) - F_2(u_i)) \overset{\text{a.s.}}{\to} 0$$

$$(1 - \sum_{i=1}^{K} p_1^{(i)})^K \overset{\text{a.s.}}{\to} 1$$

$$\text{and} \quad (1 - \sum_{i=1}^{K} p_2^{(i)})^K \overset{\text{a.s.}}{\to} 1,$$

and also upon dividing the numerator and the denominator by appropriate factors. Similarly,

$$\lim_{n\to\infty} \frac{C_{2K-1}(n_1, n_2)}{B(n_1, n_2)} = \widehat{\eta}_{2K-1} \quad \text{a.s.}$$

$$\lim_{n\to\infty} \frac{C_{2K-i}(n_1, n_2)}{B(n_1, n_2)} = \widehat{\eta}_{2K-i} \quad \text{a.s.}$$

$$\lim_{n\to\infty} \frac{C_K(n_1, n_2)}{B(n_1, n_2)} = \widehat{\eta}_K \quad \text{a.s.}$$

$$\vdots$$

$$\lim_{n\to\infty} \frac{C_0(n_1, n_2)}{B(n_1, n_2)} = p_2^{2K} \prod_{i=1}^{K} f_2(z_i) = \widehat{\eta}_0 \text{ a.s..}$$

Therefore, by (3.58) to (3.60), when $n \to \infty$,

$$P_{n_1,n_2}^{(2K,\ldots,K+1)}(u_1, v_1, \ldots, u_K, v_K; z_1, \ldots, z_K) \overset{\text{a.s.}}{\to} \widehat{\eta}_{2K} + \widehat{\eta}_{2K-1} + \ldots + \widehat{\eta}_{K+1}, \tag{3.67}$$

$$P_{n_1,n_2}^{(K)}(u_1, v_1, \ldots, u_K, v_K; z_1, \ldots, z_K) \overset{\text{a.s.}}{\to} \widehat{\eta}_K, \tag{3.68}$$

and

$$P_{n_1,n_2}^{(K-1,\ldots,0)}(u_1, v_1, \ldots, u_K, v_K; z_1, \ldots, z_K) \overset{\text{a.s.}}{\to} \widehat{\eta}_{K-1} + \widehat{\eta}_{K-2} + \ldots + \widehat{\eta}_0. \tag{3.69}$$

Using (3.55) to (3.57) we write

$$\Pi^{(1)}(z_1, z_2, \ldots, z_K) = \lim_{n\to\infty} \Pi^{(1)}(z_1, z_2, \ldots, z_K; n_1, n_2)$$

$$= \lim_n EP^{(2K,\ldots,K+1)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, \ldots U_K^{(1)}, V_K^{(1)}; z_1, \ldots, z_K)$$

$$+ \frac{1}{2} \lim_n EP^{(K)}_{n_1,n_2}(U_1^{(1)}, V_1^{(1)}, \ldots, U_K^{(1)}, V_K^{(1)}; z_1, \ldots, z_K)$$

which by (3.67), (3.68) and the (Lebesque) Dominated Convergence theorem, yields

$$\Pi^{(1)}(z_1, z_2, \ldots, z_K) = \sum_{i=0}^{K-1} \widehat{\eta}_{2K-i} + \frac{1}{2}.\widehat{\eta}_K \quad .$$

The proof is complete. □

The limiting PMC's of the 1st-stage RNN rule for K-observations are given below:

$$\alpha_1^{(1)}(K) = \lim_{n\to\infty} P(\text{Decide}(z_1, \ldots, z_K) \in \pi_2/(z_1, \ldots, z_K) \in \pi_1)$$
$$= \int \cdots \int (1 - \Pi^{(1)}(z_1, \ldots, z_K)) \prod_{i=1}^{K} f_1(z_i) dz_1 dz_2 \ldots dz_K$$
$$\alpha_2^{(1)}(K) = \lim_{n\to\infty} P(\text{Decide}(z_1, \ldots, z_K) \in \pi_1/(z_1, \ldots, z_K) \in \pi_2)$$
$$= \int \cdots \int (1 - \Pi^{(1)}(z_1, \ldots, z_K)) \prod_{i=1}^{K} f_2(z_i) dz_1 dz_2 \ldots dz_K \quad .$$

Hence the total limiting PMC of the 1st-stage RNN rule for classifying K observations is given by

$$R_{(2)}^{(1)}(K) = \xi_1 \alpha_1^{(1)}(K) + \xi_2 \alpha_2^{(1)}(K). \tag{3.70}$$

3.3. Limiting PMC's of the M-Stage MRNN rule

Let $\pi^{(M)}(z_1, \ldots z_K; n_1, n_2)$ be the conditional probability given $Z_1 = z_1, \ldots, Z_K = z_K$ that the M-stage MRNN rule classifies $\underset{\sim}{Z} = (Z_1, \ldots Z_K)$ into π_1 .

Recall the definition of $\phi^{(M)}$ in (3.8). It is easily seen that

$$\pi^{(M)}(z_1,\ldots,z_K;n_1,n_2) = P(\phi^{(1)} = 1/\ \underset{\sim}{Z}=\underset{\sim}{z}) + \sum_{i=2}^{M} P(\phi^{(1)} = 1/2, \phi^{(2)} = 1/2, \ldots \phi^{(i)} = 1/\ \underset{\sim}{Z}=\underset{\sim}{z}) + \frac{1}{2}P(\phi^{(1)} = 1/2, \ldots, \phi^{(M)} = 1/2/\ \underset{\sim}{Z}=\underset{\sim}{z}). \tag{3.71}$$

$$\text{Let} \quad \pi^{(M)}(\underset{\sim}{z}) = \lim_{n\to\infty} \pi^{(M)}(z_1,\ldots,z_K;n_1,n_2). \tag{3.72}$$

Now

$$\begin{aligned}
&P(\phi^{(1)} = 1/2, \ldots, \phi^{(i-1)} = 1/2, \phi^{(i)} = 1/\ \underset{\sim}{Z}=\underset{\sim}{z}) \\
&= P(\phi^{(1)} = 1/2/\ \underset{\sim}{Z}=\underset{\sim}{z}) \cdot P(\phi^{(2)} = 1/2/\phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\times P(\phi^{(3)} = 1/2/\phi^{(2)} = 1/2, \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\ldots P(\phi^{(j)} = 1/2/\phi^{(j-1)} = 1/2, \ldots, \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \ldots \\
&\times P(\phi^{(i-1)} = 1/2/\phi^{(i-2)} = 1/2 \ldots \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\times P(\phi^{(i)} = 1/\phi^{(i-1)} = 1/2, \ldots, \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&= P(\phi^{(1)} = 1/2/\ \underset{\sim}{Z}=\underset{\sim}{z}) \cdot \Pi_{j=2}^{i-1} P(\phi^{(j)} = 1/2/\phi^{(j-1)} = 1/2, \ldots, \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\quad \times P(\phi^{(i)} = 1/\phi^{(i-1)} = 1/2, \ldots, \phi^{(1)} = 1/2; \underset{\sim}{Z}=\underset{\sim}{z});
\end{aligned} \tag{3.73}$$

similarly,

$$\begin{aligned}
&P(\phi^{(1)} = 1/2, \ldots, \phi^{(M-1)} = 1/2, \phi^{(M)} = 1/2/\ \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\quad = \Pi_{j=2}^{M} P(\phi^{(j)} = 1/2/\phi^{(j-1)} = 1/2, \ldots, \phi^{(1)} = 1/2, \underset{\sim}{Z}=\underset{\sim}{z}) \\
&\qquad \times P(\phi^{(1)} = 1/2/\ \underset{\sim}{Z}=\underset{\sim}{z}).
\end{aligned} \tag{3.74}$$

Now we state two Lemmas whose proofs are similar to those of Lemma 3.1 and 3.2 of Das Gupta and Lin (1980) and hence omitted.

Suppose $\phi^{(1)} = 1/2$. Delete the observation corresponding to $U_\ell^{(1)}$ and $V_\ell^{(1)}$, $\forall \ell = 1, 2, \ldots, K$ from the pooled training sample. Denote the remaining $n_1 - K$ X-observations and $n_2 - K$ Y-observations by $X_i^{(2)}(i = 1, \ldots, \overline{n_1 - K})$ and $Y_j^{(2)}(j = 1, \ldots, \overline{n_2 - K})$ respectively.

LEMMA 3.3.1. *Given* $Z_1 = z_1, \ldots, Z_K = z_K, \phi^{(1)} = 1/2, U_1^{(1)} = u_1, V_1^{(1)} = v_1, \ldots, U_K^{(1)} = u_k$ *and* $V_K^{(1)} = v_K$, *then*

(i) $X_i^{(2)}$'s and $Y_j^{(2)}$'s are mutually conditionally independent;

(ii) the conditional density of $X_i^{(2)}$ is

$$f_1^{(2)}(x) = \frac{f_1(x)}{[1-(p_1^{(1)}+p_1^{(2)}+\ldots+p_1^{(K)})]},$$

$$\text{on } [(u_1,v_1)\cup\ldots\cup(u_K,v_K)]^c. \qquad (3.75)$$

(iii) the conditional density of $Y_j^{(2)}$ is

$$f_2^{(2)}(y) = \frac{f_2(y)}{[1-(p_2^{(1)}+p_2^{(2)}+\ldots+p_2^{(K)})]}$$

$$\text{on } [(u_1,v_1)\cup\ldots\cup(u_K,v_K)]^c. \qquad (3.76)$$

where

$$p_1^{(i)} = F_1(v_i) - F_1(u_i) \text{ and } p_2^{(i)} = F_2(v_i) - F_2(u_i) \quad \text{for } i = 1, \ldots, K.$$

Lemma 3.3.1 can be extended in similar fashion inductively. Suppose $\phi^{(j)} = 1/2 \quad (j = 1, \ldots, i-1)$. Delete the observation corresponding to $U_\ell^{(i)}$ and $V_\ell^{(i)} (\ell = 1, \ldots, K; j = 1, 2, \ldots, i-1)$ and denote the remaining $n_1 - (i-1)K$ X-observations and $n_2 - (i-1)K$ Y-observations by $X_r^{(i)} (r = 1, \ldots, \overline{n_1 - (i-1)K})$ and $Y_r^{(i)} (r = 1, \ldots, \overline{n_2 - (i-1)K})$ respectively.

LEMMA 3.3.2. *Given* $\underset{\sim}{Z} = \underset{\sim}{z}, \phi^{(j)} = 1/2, U_\ell^{(j)}, V_\ell^{(j)} (\ell = 1, \ldots, K; j = 1, \ldots, i-1)$ *the following are true:*

(i) $X_r^{(i)}$ *and* $Y_r^{(i)}$ *are mutually conditionally independent;*

(ii) the conditional density of $X_r^{(i)}$ is

$$f_1^{(i)}(x) = \frac{f_1^{(i-1)}(x)}{[1-(p_1^{(1)}(i-1)+p_1^{(2)}(i-1)+\ldots+p_1^{(K)}(i-1))]} \tag{3.77}$$

$$\text{on } [(u_1^{(i-1)},v_1^{(i-1)})\cup\ldots\cup(u_K^{(i-1)},v_K^{(i-1)})]^c$$

where

$$p_1^{(\ell)}(i-1) = F_1^{(i-1)}(v^{(i-1)}) - F_1^{(i-1)}(u^{(i-1)}), \quad \ell = 1,2,\ldots,K$$

and $F_1^{(i-1)}$ is the c.d.f. corresponding to $f_1^{(i-1)}$, defined inductively by (3.75) and (3.77).

(iii) the conditional density of $Y_r^{(i)}$ is

$$f_2^{(i)}(y) = \frac{f_2^{(i-1)}(y)}{[1-(p_2^{(1)}(i-1)+\ldots+p_2^{(k)}(i-1))]}$$

$$\text{on } [(u_1^{(i-1)},v^{(i-1)})\cup\ldots\cup(u_k^{(i-1)},v_k^{(i-1)})]^c, \tag{3.78}$$

where

$$p_2^{(\ell)}(i-1) = F_2^{(i-1)}(v^{(i-1)}) - F_2^{(i-1)}(u^{(i-1)}), \quad \ell = 1,\ldots,K$$

and $F_2^{(i-1)}$ is the c.d.f. corresponding to $f_2^{(i-1)}$ defined inductively by (3.76) and (3.78).

First we state a theorem and then derive exact expressions for the limiting PMC's of M-stage MRNN rule when $K=2$.

THEOREM 3.3.1. *Suppose z_1 and z_2 are points of continuity of both f_1 and f_2 then the limiting probability of classifying Z_1 and Z_2 into π_1 using the M-stage MRNN rule, given $Z_1 = z_1, Z_2 = z_2$, is given by*

$$\begin{aligned}\pi^{(M)}(z_1,z_2) &= \lim_{n\to\infty} \pi^{(M)}(z_1,z_2;n_1,n_2) \\ &= (\widehat{\eta}_4+\widehat{\eta}_3)\sum_{i=1}^{M-1}\widehat{\eta}_2^i + \frac{1}{2}\widehat{\eta}_2^M,\end{aligned} \tag{3.79}$$

where $\widehat{\eta}_4, \widehat{\eta}_3$ and $\widehat{\eta}_2$ are given by (3.19), (3.20) and (3.21) respectively.

PROOF. The conditional probabilities $\phi^{(i)} = 1$ and $\phi^{(i)} = 1/2$ given $Z_1 = z_1, Z_2 = z_2, U_\ell^{(j)} = u_\ell^{(j)}, V_\ell^{(j)} = v_\ell^{(j)} (\ell = 1,2)(j = 1,2,\ldots,i)$ and $\phi^{(i)} = 1/2 (j = 1,2,\ldots,i-1)$ are respectively given by $\frac{C_4^{(i)}+C_3^{(i)}}{B^{(i)}}$ and $\frac{C_2^{(i)}}{B^{(i)}}$, where $C_4^{(i)}, C_3^{(i)}, C_2^{(i)}, B^{(i)}$ are obtained from C_4, C_3, C_2, B (given in 3.15-3.16) respectively after replacing $n_1, n_2, u_1, v_1, u_2, v_2, f_1, f_2$ by $n_1 - 2(i-1), n_2 - 2(i-1), u_1^{(i)}, v_1^{(i)}, u_2^{(i)}, v_2^{(i)}, f_1^{(i)}, f_2^{(i)}$ respectively (using (i) of Lemma 3.3.2).

Since f_j's $(j = 1,2)$ are continuous $u_\ell^{(i)}, v_\ell^{(i)} \to z_\ell$ $(\ell = 1,2)$ a.s. imply $f_j^{(i)}(u_\ell^{(i)}) \to f_j(z_\ell)$ and $f_j^{(i)}(v_\ell^{(i)}) \to f_j(z_\ell)$ a.s. (note that from Lemma 3.3.2(ii)

$$f_j^{(i)}(x) = \frac{f_j(x)}{\prod_{K=1}^{i-1}\{1 - (p_j^{(1)}(K) + p_j^{(2)}(K))\}}$$

and $p_j^{(1)}(i) + p_j^{(2)}(i) \to 0$ a.s., for $i = 1,\ldots,M, j = 1,2$ as $n \to \infty$). Thus the limiting values of $\frac{C_4^{(i)}+C_3^{(i)}}{B^{(i)}}$ and $\frac{C_2^{(i)}}{B^{(i)}}$ are $(\widehat{\eta}_4 + \widehat{\eta}_3)$ and $\widehat{\eta}_2$ respectively. Introducing the sets $A_1^{(1)}, A_2^{(1)}$ and arguing as in (3.9)-(3.11), (3.79) follows from (3.71)-(3.74). The proof is completed. □

The limiting PMC's of the M-stage MRNN rule $(K = 2)$ are given below:

$$\alpha_1^{(M)}(2) = \lim_{n\to\infty} P[\text{ M-stage MRNN rule decides } (z_1, z_2) \in \pi_2/(z_1, z_2) \in \pi_1]$$
$$= \int\int [1 - \pi^{(M)}(z_1, z_2)] f_1(z_1) f_1(z_2) dz_1 dz_2$$

$$\alpha_2^{(M)}(2) = \lim_{n\to\infty} P[\text{ M-stage MRNN rule decides } (z_1, z_2) \in \pi_1/(z_1, z_2) \in \pi_2]$$
$$= \int\int \pi^{(M)}(z_1, z_2) f_2(z_1) f_2(z_2) dz_1 dz_2.$$

If we take $p_i = \xi_i (i = 1,2)$ as prior probabilities, then the limiting total PMC for the M-stage MRNN rule is given by

$$R_{(2)}^{(M)}(2) = \xi_1 \alpha_1^{(M)}(2) + \xi_2 \alpha_2^{(M)}(2). \tag{3.80}$$

Note that

$$1-\pi^{(M)}(z_1,z_2)=(\widehat{\eta}_1+\widehat{\eta}_0)\sum_{i=0}^{M-1}\widehat{\eta}_2^i+\frac{1}{2}\widehat{\eta}_2^M, \tag{3.81}$$

where $\widehat{\eta}_1$ and $\widehat{\eta}_0$ are given by

$$\widehat{\eta}_1=\frac{2p_1p_2^3[f_1(z_2)f_2^2(z_1)f_2(z_2)+f_1(z_1)f_2^2(z_2)f_2(z_1)]}{\{\pi_{i=1}^2(p_1f_1(z_i)+p_2f_2(z_i)\}^2} \tag{3.82}$$

$$\widehat{\eta}_0=\frac{(p_2^2f_2(z_1)f_2(z_2))^2}{\{\pi_{i=1}^2(p_1f_1(z_i)+p_2f_2(z_i)\}^2}. \tag{3.83}$$

THEOREM 3.3.2.

$$R_{(2)}^{(M)}(2)\le R_{(2)}^{(M-1)}(2)\ \textit{if}\ \xi_1=\xi_2$$
$$\textit{and}\ R_{(2)}^{*}(2)<R_{(2)}^{(\infty)}(2)$$

where $R_{(2)}^{*}(2)$ is Bayes risk given by (2.9).

PROOF. We write $R_{(2)}^{(M)}(2)$ from (3.80) as

$$R_{(2)}^{(M)}(2)=\xi_1\int\int\left(1-\pi^{(M)}(z_1,z_2)\right)f_1(z_1)f_1(z_2)dz_1dz_2$$
$$+\xi_2\int\int\pi^{(M)}(z_1,z_2)f_2(z_1)f_2(z_2)dz_1dz_2.$$

Thus,

$$R_{(2)}^{(M)}(2)-R_{(2)}^{(M-1)}(2)$$
$$=\xi_1\int\int\left(\pi^{(M-1)}(z_1,z_2)-\pi^{(M)}(z_1,z_2)\right)f_1(z_1)f_1(z_2)dz_1dz_2$$
$$-\xi_2\int\int\left(\pi^{(M-1)}(z_1,z_2)-\pi^{(M)}(z_1,z_2)\right)f_2(z_1)f_2(z_2)dz_1dz_2. \tag{3.84}$$

We know from (3.79)

$$\pi^{(M)}(z_1,z_2)=\left(\widehat{\eta}_4+\widehat{\eta}_3\right)\sum_{i=0}^{M-1}\widehat{\eta}_2^i+\frac{1}{2}\widehat{\eta}_2^M.$$

Therefore,

$$\begin{aligned}
\pi^{(M)}(z_1,z_2) - \pi^{(M-1)}(z_1,z_2) \\
&= \Big(\widehat{\eta}_4 + \widehat{\eta}_3\Big)\widehat{\eta}_2^{M-1} - \frac{1}{2}\widehat{\eta}_2^{M-1}(1-\widehat{\eta}_2) \\
&= \frac{1}{2}\widehat{\eta}_2^{M-1}\Big(2\widehat{\eta}_4 + 2\widehat{\eta}_3 - \widehat{\eta}_4 - \widehat{\eta}_3 - \widehat{\eta}_1 - \widehat{\eta}_0\Big) \\
&= \frac{1}{2}\widehat{\eta}_2^{M-1}\Big(\widehat{\eta}_4 - \widehat{\eta}_0 + \widehat{\eta}_3 - \widehat{\eta}_1\Big).
\end{aligned} \tag{3.85}$$

Using (3.84) and (3.85) we obtain

$$\begin{aligned}
R_{(2)}^{(M-1)}(2) - R_{(2)}^{(M)}(2) \\
= -\frac{1}{2}\int\int(\xi_2 f_2(z_1)f_2(z_2) - \xi_1 f_1(z_1)f_1(z_2))(\widehat{\eta}_0 - \widehat{\eta}_4 + \widehat{\eta}_1 - \widehat{\eta}_3)dz_1dz_2,
\end{aligned} \tag{3.86}$$

where we have put $p_i = \xi_i$ for $i = 1, 2$.

Thus, we have

$$\widehat{\eta}_0 - \widehat{\eta}_4 = \frac{(\xi_2^2 f_2(z_1)f_2(z_2))^2}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2} - \frac{(\xi_1^2 f_1(z_1)f_1(z_2))^2}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}$$

$$= \frac{(\xi_2^2 f_2(z_1)f_2(z_2) - \xi_1^2 f_1(z_1)f_1(z_2))(\xi_2^2 f_2(z_1)f_2(z_2) + \xi_1^2 f_1(z_1)f_1(z_2))}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}, \tag{3.87}$$

$$\widehat{\eta}_1 - \widehat{\eta}_3 = \frac{1}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}[2\xi_1\xi_2^3(f_1(z_2)f_2^2(z_1)f_2(z_2)$$

$$+ f_1(z_1)f_2^2(z_2)f_2(z_1))$$

$$- 2\xi_1\xi_2^3(f_1^2(z_1)f_1(z_2)f_2(z_2) + f_1^2(z_2)f_1(z_1)f_2(z_1)]$$

$$= \frac{2\xi_1\xi_2}{\{\pi_{i=1}^2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i))\}^2}[\xi_2^2 f_1(z_2)f_2^2(z_1)f_2(z_2)$$

$$+ \xi_2^2 f_1(z_1)f_2^2(z_2)f_2(z_1)$$

$$- \xi_1^2 f_1^2(z_1)f_1(z_2)f_2(z_2) - \xi_1^2 f_1^2(z_2)f_1(z_1)f_2(z_1)]$$

$$=\frac{2\xi_1\xi_2}{\{\pi_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\}^2}[f_1(z_2)f_2(z_1)(\xi_2^2 f_2(z_1)f_2(z_2)-\xi_2^2 f_1(z_1)f_1(z_2))$$

$$+f_1(z_1)f_2(z_2)(\xi_2^2 f_2(z_1)f_2(z_2)-\xi_1^2 f_1(z_1)f_1(z_2))]$$

$$=\frac{2\xi_1\xi_2[(\xi_2^2 f_2(z_1)f_2(z_2)-\xi_1^2 f_1(z_1)f_1(z_2))(f_1(z_1)f_2(z_2))+f_1(z_2)f_2(z_1))]}{\{\pi_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\}^2} \quad (3.88)$$

Using (3.87) and (3.88) we write (3.86) as

$$\begin{aligned}
&R_{(2)}^{(M)}(2)-R_{(2)}^{(M-1)}(2)\\
&=-\frac{1}{2}\int\int(\xi_2 f_2(z_1)f_2(z_2)-\xi_1 f_1(z_1)f_1(z_2))(\xi_2^2 f_2(z_1)f_2(z_2)-\xi_1^2 f_1(z_1)f_1(z_2))\\
&\times\frac{[\xi_2^2 f_2(z_1)f_2(z_2)+\xi_1^2 f_1(z_1)f_1(z_2)+2\xi_1\xi_2(f_1(z_1)f_2(z_2)+f_1(z_2)f_2(z_1))]}{\{\pi_{i=1}^2(\xi_1 f_1(z_i)+\xi_2 f_2(z_i))\}^2}\\
&\qquad\times dz_1dz_2
\end{aligned} \quad (3.89)$$

If $\xi_1=\xi_2=\xi=\frac{1}{2}$ we get from (3.89),

$$\begin{aligned}
R_{(2)}^{(M)}-R_{(2)}^{(M-1)}&=-\frac{1}{4}\int\int(f_2(z_1)f_2(z_2)-f_1(z_1)f_1(z_2))^2\\
&\times(\eta_2(z_1)\eta_2(z_2)+\eta_1(z_1)\eta_1(z_2)\\
&+2(\eta_1(z_1)\eta_2(z_2)+\eta_1(z_2)\eta_2(z_1))dz_1dz_2.
\end{aligned}$$

The above expressions implies

$$R_{(2)}^{(M)}\le R_{(2)}^{(M-1)}. \quad (3.90)$$

Now

$$\begin{aligned}
R_{(2)}^{(\infty)}(2)&=\lim_M R_{(2)}^{(M)}(2)\\
&=\int\int\left[\left(\widehat{\eta}_1+\widehat{\eta}_0\right)\xi_1 f_1(z_1)f_1(z_2)+\left(\widehat{\eta}_4+\widehat{\eta}_3\right)\xi_2 f_2(z_1)f_2(z_2)\right]\\
&\qquad\times(1-\widehat{\eta}_2)^{-1}dz_1dz_2\\
&=\int\int\left[\frac{\widehat{\eta}_1+\widehat{\eta}_0}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_1 f_1(z_1)f_1(z_2)\right.\\
&\qquad\left.+\frac{\widehat{\eta}_4+\widehat{\eta}_3}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_2 f_2(z_1)f_2(z_2)\right]dz_1dz_2
\end{aligned} \quad (3.91)$$

Note that

$$\begin{aligned}&\min(\xi_1 f_1(z_1)f_1(z_2), \quad \xi_2 f_2(z_1)f_2(z_2))\\ &\leq \frac{\widehat{\eta}_1+\widehat{\eta}_0}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_1 f_1(z_1)f_1(z_2)\\ &+\frac{\widehat{\eta}_4+\widehat{\eta}_3}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_2 f_2(z_1)f_2(z_2)\end{aligned}$$

Integrating both sides w.r.t. z_1 and z_2, we get

$$\begin{aligned}&\int\int \min(\xi_1 f_1(z_1)f_1(z_2), \quad \xi_2 f_2(z_1)f_2(z_2))dz_1dz_2\\ &\leq \int\int\left[\frac{\widehat{\eta}_1+\widehat{\eta}_0}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_1 f_1(z_1)f_1(z_2)\right.\\ &\left.+\frac{\widehat{\eta}_4+\widehat{\eta}_3}{\widehat{\eta}_4+\widehat{\eta}_3+\widehat{\eta}_1+\widehat{\eta}_0}\xi_2 f_2(z_1)f_2(z_2)\right]dz_1dz_2\end{aligned}$$

Which by (2.9) and (3.91) implies

$$R^*_{(2)}(2) \leq R^{(\infty)}_{(2)}(2). \tag{3.92}$$

The proof is complete. □

3.4 Estimation of PMC'S of the first stage RNN rule

We shall propose estimates of PMC'S of the first stage MRNN rule for $K = 2$ using a deleted counting method and show that these estimates are asymptotically unbiased and consistent. Let

$$\psi_x^{(i,i+1)} = 1 - \phi^{(1)}(X_i, X_{i+1}; X_\alpha\text{'s}, Y_\beta \text{ 's } i \neq \alpha \neq i+1) \tag{3.93}$$

$$\psi_x^{(k,k+1)} = \phi^{(1)}(Y_k, Y_{k+1}; X_\alpha\text{'s}, Y_\beta \text{ 's } k \neq \beta \neq k+1), \tag{3.94}$$

where $\phi^{(1)}$ is defined in (3.8).

Let

$$S_x = \{1, 3, 5, \ldots, N_1; N_1 = n_1 - 1 \text{ if } n_1 \text{ even, } N_1 = n_1 - 2 \text{ if } n_1 \text{ odd }\},$$

$$S_y = \{1, 3, 5, \ldots, N_2; N_2 = n_2 - 1 \text{ if } n_2 \text{ even, } N_2 = n_2 - 2 \text{ if } n_2 \text{ odd }\},$$

$$p_x = \frac{1}{[\frac{n_1}{2}]}\sum_{i\in S_x}\psi_x^{(i,i+1)} \tag{3.95}$$

$$p_y = \frac{1}{[\frac{n_2}{2}]} \sum_{i \in S_y} \psi_y^{(i,i+1)} \tag{3.96}$$

$$U_x = \frac{1}{[\frac{n_1}{2}]} \sum_{i \in S_x} E(\psi_x^{(i,i+1)}/X_i, X_{i+1}) \tag{3.97}$$

$$U_y = \frac{1}{[\frac{n_2}{2}]} \sum_{i \in S_y} E(\psi_y^{(k,k+1)}/Y_k, Y_{k+1}) \tag{3.98}$$

and

$$p_{xy} = \frac{n_1 p_x + n_2 p_y}{n_1 + n_2}. \tag{3.99}$$

Then p_x and p_y can be used as estimates of PMC's and p_{xy} can be used as an estimate of total PMC.

Note that

$$E(p_x) = \int\int (1 - \pi^{(1)}(z_1, z_2; n_1 - 2, n_2) f_1(z_1) f_2(z_2) dz_1 dz_2,$$

therefore by

$$\begin{aligned} \lim E(p_x) &= \int\int (1 - \pi^{(1)}(z_1, z_2)) f_1(z_1) f_1(z_2) dz_1 dz_2 \\ &= \alpha_1^{(1)}(2). \end{aligned} \tag{3.100}$$

Similarly,

$$\lim E(p_y) = \alpha_2^{(1)}(2). \tag{3.101}$$

Also note that U_x and U_y are U-statistics so that (see Serfling (1980, p. 190)) as $n \to \infty$, we have

$$U_x \xrightarrow{\text{a.s.}} \alpha_1^{(1)}(2), \quad U_y \xrightarrow{\text{a.s.}} \alpha_1^{(1)}(2). \tag{3.102}$$

Therefore, to establishthe consistency of p_x as an estimator for $\alpha_1^{(1)}(2)$, it is enough to show that $p_x - U_x \xrightarrow{\text{P}} 0$ which is implied by $E(p_x - U_x)^2 \to 0$.

LEMMA 3.4.1. *$E(p_x - U_x)^2 = O(\frac{1}{n_1})$ and $E(p_y - U_y)^2 = O(\frac{1}{n_2})$ where p_x, p_y, U_x and U_y are defined in (3.95)-(3.98) respectively.*

PROOF. Using (3.95) and (3.97), we have

$$p_x - U_x = \frac{1}{[\frac{n_1}{2}]} \sum_{i \in S_x} (\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1}),$$

and

$$E(p_x - U_x)^2 = \frac{1}{[\frac{n_1}{2}]^2}\Big[\sum_{i\in s_x} E(\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))^2$$

$$+ \sum_{i\neq j}\sum E(\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))$$

$$\cdot (\psi_x^{(j,j+1)} - E(\psi_x^{(j,j+1)}/X_j, X_{j+1}))\Big]. \tag{3.103}$$

Now

$$E\{\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1})\}^2 = E \text{ Var } E(\psi_x^{(i,i+1)}/X_i, X_{i+1})$$

$$= E\{E(\psi_x^{(i,i+1)^2}/X_i, X_{i+1}) - (E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))^2\}$$

$$\leq E\{E(\psi_x^{(i,i+1)}/X_i, X_{i+1}) - (E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))^2\}$$

$$= E\{1 - \pi^{(1)}(X_i, X_{i+1}; n_1 - 2, n_2) - (1 - \pi^{(1)}(X_i, X_{i+1}; n_1 - 2, n_2))^2\}$$
$$= E\{(1 - \pi^{(1)}(\underset{\sim}{X}; n_1 - 2, n_2))\pi^{(1)}(\underset{\sim}{X}; n_1 - 2, n_2)\}$$
$$\leq \frac{1}{4}. \tag{3.104}$$

and

$$E(\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))(\psi^{(j,j+1)} - E(\psi_x^{(j,j+1)}/X_i, X_{i+1}))$$

$$= E\{E(\psi_x^{(i,i+1)} - E(\psi_x^{(i,i+1)}/X_i, X_{i+1}))$$

$$\cdot (\psi^{(j,j+1)} - E(\psi_x^{(j,j+1)}/X_i, X_{i+1}))/X_i, X_{i+1}; X_j, X_{j+1}\}$$

$$= E\{E(\psi_x^{(i,i+1)}/X_i, X_{i+1}) \cdot E(\psi_x^{(j,j+1)}/X_j, X_{j+1})$$

$$- E(\psi_x^{(i,i+1)}/X_i, X_{i+1}) \cdot E(\psi^{(j,j+1)}/X_j, X_{j+1}))$$
$$- E(\psi^{(i,i+1)}/X_i, X_{i+1}) \cdot E(\psi^{(j,j+1)}/X_j, X_{j+1})$$

$$+ E(\psi^{(i,i+1)}/X_i, X_{i+1}) \cdot E(\psi^{(j,j+1)}/X_j, X_{j+1})\}$$
$$= 0. \tag{3.105}$$

Therefore using (3.103)-(3.105) we get

$$E(p_x - U_x)^2 = O\left(\frac{1}{n_1}\right);$$

similarly, one gets

$$E(p_y - U_y)^2 = O\left(\frac{1}{n_2}\right).$$

This completes the proof. □

Note that Lemma 3.4.1 shows that $p_x - U_x \xrightarrow{P} 0$ and also $U_x \xrightarrow{P} \alpha_1^{(1)}(2)$ so that by (3.102) we have, as $n \to \infty$

$$p_x \xrightarrow{P} \alpha_1^{(1)}(2) \quad \text{and} \quad p_y \xrightarrow{P} \alpha_1^{(1)}(2). \tag{3.106}$$

By letting $\xi_i = \lim \frac{n_i}{n_1+n_2}$ $i = 1, 2$ we obtain by (3.99) and (3.106) we get

$$p_{xy} \xrightarrow{P} \xi_1 \alpha_1^{(1)}(2) + \xi_2 \alpha_2^{(1)}(2) = R_{(2)}^{(1)}(2), \text{ as } n \to \infty.$$

Similarly, one could get an asymptotically unbiased and consistent estimator for $R_{(2)}^{(1)}(K)$.

CHAPTER 4

CLASSIFICATION OF MULTIPLE OBSERVATIONS INTO ONE OF s POPULATIONS.

4.0 **Introduction.** In this Chapter, we are concerned with the classification of K ($K \geq 1$) observations into one of s populations. First in Section 4.1, we consider the problem of classifying a single univariate observation into one among s univariate populations using left and right hand Rank Nearest Neighbor rule. This extends the results of Das Gupta and Lin (1980) to more than two populations. In Section 4.2, we propose a rule which generalizes the rule (3.1) for classifying K univariate observations to one of s univariate population using first-stage RNN rule and indicate how to get the asymptotic total probability of misclassification of the proposed rule. Finally, in Section 4.3, we consider the problem of classifying K ($K \geq 2$) multivariate observations to one of s multivariate populations in a Bayesian framework. For $K = 2$ and general s, we obtain the limiting risk of the proposed rule which turns out to be the exactly same as the limiting risk of 1-NN rule obtained by Cover and Hart (1967) for s populations and we also indicate bounds on it.

4.1 Classification of a single univariate observation into one among s populations using left and right Rank Nearest Neighbor rule.

Suppose we have s populations $\pi_1, \pi_2, \ldots, \pi_s$ and an observation Z to be classified into one among given s-populations $\pi_1, \pi_2, \ldots, \pi_s$. Let $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ be the training sample from i-th population π_i $i = 1, 2, \ldots, s$. The classification RNN rule may be described as follows:

Pool observatons X_{ij}'s, $i = 1, 2, \ldots, s, j = 1, 2, \ldots, n_i$ and Z and rank them; then

(i) if Z is either the smallest or the largest, classify Z into the population of its nearest rank neighbors;

(ii) if both the left and right neighbors of Z belong to one population, classify Z to that population;

(iii) If both the immediate left and right neighbors of Z belong to different populations, classify Z into either of the two populations with probabilities 1/2 and 1/2.

In the sub section 4.1.1 the asymptotic probabilities of misclassification are derived from the RNN rule. It turns out that the total asymptotic probability of misclassification is same as the asymptotic risk of the 1-NN rule with d as the metric distance, obtained by Cover and Hart (1967) for s populations. It may be pointed out that the Cover and Hart analysis is based on Bayesian set up whereas our analysis is based on a nonparametric frame work and depends on ranks.

To reduce the chance of randomization one may consider a multi-stage version as follows:

If the first stage leads to a tie delete these two tied observations and apply the first-stage rule to the remaining. Proceed this way and move to the next stage whenever a tie occurs. The M-stage RNN rule is defined to be the one which terminates at the M-th stage (and allows for a tie in the final stage). In the sub section 4.1.2 the asymptotic PMC's of the M-stage RNN rule are obtained. It is shown that total PMC decreases as the stage M increases. Therefore our result agrees with the intuition.

The estimates of the PMC's of first stage RNN rule based on deletion-counting method are given in the sub section 4.1.3. It is shown that the estimates are asymptotically unbiased and consistent.

Let X_{ij} have c.d.f F_i and corresponding p.d.f $f_i, i = 1, 2, \ldots, s$.

4.1.1 Limiting PMC's of the First-stage RNN rule. Let the right-hand and the left-hand RNN of Z at the M-th stage be $V^{(M)}$ and $U^{(M)}$ respectively. Let $n = \min(n_1, n_2, \ldots, n_s)$.

LEMMA 4.1.1. *If $M/n \to 0$ as $n \to \infty$ the probability that (under any $Z \sim f_i, i = 1, 2, \ldots, s$) there are at least M observations to the left of Z and at least M observations to the right of Z in the pooled sample as $n \to \infty$ is one.*

LEMMA 4.1.2. *Given that $Z \sim F_1$ and $\frac{M}{n_1} \to 0$ as $n_1 \to \infty$ then $V^{(M)}$ and $U^{(M)} \to Z$ a.s. as $n \to \infty$.*

The proofs of the above two Lemmas are similar to those of Lemma 2.1 and 2.2 of Das Gupta and Lin (1980) hence they are omitted.

Let $U^{(M)}$ and $V^{(M)}$ be the left-hand and right-hand RNN of Z at the M-stage respectively. Define $\phi_\ell^{(M)}(Z; X_{ij}\text{'s} \quad i = 1, \ldots s;\ j = 1, 2, \ldots, n_i)$ as

$$\phi_\ell^{(M)} = \begin{cases} 1 & \text{if both } U^{(M)} \text{ and } V^{(M)} \text{ are from } \ell\text{-th population or } Z \text{ is} \\ & \text{an extreme observation and its RNN is an observation} \\ & \text{from } \ell\text{-th population.} \\ \frac{1}{2} & \text{if either } U^{(M)} \text{ or } V^{(M)} \text{ belongs to } \ell\text{-th population} \\ & \text{(i.e. one of } U^{(M)} \text{ and } V^{(M)} \text{ belongs to } \ell\text{-th population)} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Let A_1 be the event that both $U^{(1)}$ and $V^{(1)}$ are well defined.

The conditional probability that Z comes from π_ℓ using the one-stage RNN rule given $Z = z$ is

$$\begin{aligned} \pi_\ell^{(1)}(z; n_1, \ldots, n_s) &= E(\phi_\ell^{(1)}/Z = z) \\ &= E(\phi_\ell^{(1)} I_{A_1}/Z = z) + E(\phi_\ell^{(1)} I_{A_1^c}/Z = z). \end{aligned} \tag{4.2}$$

However, the lemma 4.1.1 implies

$$E(\phi_\ell^{(1)} I_{A_1^c}/Z = z) \le P(A_1^c/Z = z) \to 0. \tag{4.3}$$

Also

$$\begin{aligned} &E(\phi_\ell^{(1)} I_{A_1}/Z = z) \\ &= P(\phi_\ell^{(1)} = 1 \cap A_1/Z = z) + \frac{1}{2}P(\phi_\ell^{(1)} = \frac{1}{2} \cap A_1/Z = z) \\ &= EP^{(11)}_{n_1,\ldots,n_s}(U^{(1)}, V^{(1)}, z) \\ &+ \frac{1}{2}EP^{(10)}_{n_1,\ldots,n_s}(U^{(1)}, V^{(1)}, z), \end{aligned} \tag{4.4}$$

where

$$P^{(11)}_{n_1,\dots,n_s}(U^{(1)}, V^{(1)}, z) = P(\phi^{(1)}_\ell = 1 / U^{(1)} = u, V^{(1)} = v, A_1)$$

$$P^{(10)}_{n_1,\dots,n_s}(U^{(1)}, V^{(1)}, z) = P(\phi^{(1)}_\ell = 1/2 \; / U^{(1)} = u, V^{(1)} = v, A_1).$$

Now note that

$$P^{(11)}_{n_1,\dots,n_s}(u, v, z) = \frac{C_\ell(n_1, \dots, n_s)}{B(n_1, \dots, n_s)} \tag{4.5}$$

$$P^{(10)}_{n_1,\dots,n_s}(u, v, z) = \frac{C_{\ell 0}(n_1, \dots, n_s)}{B(n_1, \dots, n_s)}, \tag{4.6}$$

Where $\forall \ell = 1, 2, \dots, s$,

$$\begin{aligned} C_\ell(n_1, \dots, n_s) &= n_\ell(n_\ell - 1)(1 - (F_\ell(v) - F_\ell(u)))^{n_\ell - 2} \\ &\times \prod_{\substack{i=1 \\ i \neq \ell}}^{s} (1 - (F_i(v) - F_i(u)))^{n_i} \cdot f_\ell(u) f_\ell(v), \end{aligned} \tag{4.7}$$

$$\begin{aligned} C_{\ell 0}(n_1, \dots, n_s) = \sum_{\substack{m=1 \\ m \neq \ell}}^{s} & n_\ell n_m [1 - (F_\ell(v) - F_\ell(u))]^{n_\ell - 1} [1 - (F_m(v) - F_m(u))]^{n_m - 1} \\ \times \prod_{\substack{j=1 \\ j \neq m, j \neq \ell}}^{s} & [1 - (F_j(v) - F_j(u))]^{n_j} [f_\ell(u) f_m(v) + f_m(u) f_\ell(v)] \end{aligned} \tag{4.8}$$

and

$$B(n_1, \dots, n_s) = \sum_{\ell=1}^{s} C_\ell(n_1, \dots, n_s) + \sum_{\ell=1}^{s} C_{\ell 0}(n_1, \dots, n_s). \tag{4.9}$$

$C_\ell(n_1, \dots, n_s)$ is proportional to the conditional probability of classifying Z into π_ℓ, given $Z = z$ and $C_{\ell 0}(n_1, \dots, n_s)$ is the proportional to the conditional probability that only one of $V^{(1)}$ and $U^{(1)}$ is from π_ℓ given $Z = z$. Let

$$p_i = \lim_{n \to \infty} \frac{n_i}{n_1 + n_2 + \dots + n_s} \quad \text{and} \tag{4.10}$$

assume $0 < p_i < 1$.

THEOREM 4.1.1. *Suppose z is a continuity point of all $f_1, f_2, \ldots, f_s$. Then the asymptotic conditional probability that z comes from π_ℓ, given $Z = z$ is given by*

$$\pi_\ell^{(1)}(z) = \lim_{n\to\infty} \pi_\ell^{(1)}(z; n_1, \ldots, n_s)$$

$$= \widehat{\eta}_\ell + \frac{1}{2}\widehat{\eta}_{\ell 0},$$

where

$$\widehat{\eta}_\ell = \frac{p_\ell^2 f_\ell^2(z)}{(\sum_{\ell=1}^s p_\ell f_\ell(z))^2} \tag{4.11}$$

and

$$\widehat{\eta}_{\ell 0} = \frac{2p_\ell f_\ell(z)(\sum_{\substack{m=1\\ m\neq\ell}}^s p_m f_m(z))}{(\sum_{\ell=1}^s p_\ell f_\ell(z))^2}. \tag{4.12}$$

PROOF. Write $G_i(n) = F_i(v) - F_i(u)$. Since $U^{(1)} \to z$ a.s. and $V^{(1)} \to z$ a.s., as $n \to \infty$, and F_i is continuous, $G_i(n) \to 0$ as $n \to \infty$. Now by (4.5) and dividing the numerator and denominator of $\frac{C_\ell(n_1,\ldots,n_s)}{B(n_1,\ldots,n_s)}$ by $\sum_{i=1}^s n_i$ and $\sum_{\substack{i=1\\ i\neq\ell}}^s (1-(F_i(v) - F_i(u))^{n_i}$, we get

$$P^{(11)}_{n_1,\ldots,n_s}(u,v,z) = \frac{C_\ell(n_1,\ldots,n_s)}{B(n_1,\ldots,n_s)}$$

$$= \frac{\left(\frac{n_\ell}{\sum_{i=1}^s n_i}\right)\left(\frac{n_\ell - 1}{\sum_{i=1}^s n_i}\right) f_\ell(u) f_\ell(v)}{[Denominator]}, \tag{4.13}$$

where,

$$[Denominator] = \sum_{\ell=1}^s \left(\frac{n_\ell}{\sum n_i}\right)\left(\frac{n_\ell - 1}{\sum n_i}\right) f_\ell(u) f_\ell(v)$$

$$+ \sum_{\substack{\ell=1\\ \ell<m}}^s \sum_{m=1}^s \left(\frac{n_\ell}{\sum n_i}\right)\left(\frac{n_m}{\sum n_i}\right)\left(\frac{1-G_\ell(n)}{1-G_\ell(n)}\right)[f_\ell(u) f_m(v)$$

$$+ f_m(u) f_\ell(v)]. \tag{4.14}$$

By (4.10) and (4.14) and as $n \to \infty$,

$$[Denominator] \to \left(\sum_{i=1}^{s} p_i f_i(z)\right)^2 \text{ a.s.} \tag{4.15}$$

so that by (4.13) and (4.15) we get, as $n \to \infty$,

$$P^{(11)}_{n_1,\ldots,n_s}(u,v,z) \to \frac{p_\ell^2 f_\ell^2(z)}{\left(\sum_{i=1}^{2} p_i f_i(z)\right)^2} \text{ a.s. .}$$

Similarly, we write (4.6) as

$$P^{(10)}_{n_1,\ldots,n_s}(u,v,z) = \frac{C_{\ell 0}(n_1,\ldots,n_s)}{B(n_1,\ldots,n_s)}$$

$$= \frac{\sum_{m=1,m\neq\ell}^{s} \left(\frac{n_\ell - 1}{\sum n_i}\right)\left(\frac{n_m}{\sum n_i}\right)\frac{(1-G_\ell)}{(1-G_m)}(f_\ell(u)f_m(v) + f_m(u)f_\ell(v))}{[Denominator]}, \tag{4.16}$$

where [Denominator] is given by (4.14). Thus by (4.10), (4.15) and (4.16), we have

$$P^{(10)}_{n_1,\ldots,n_s}(u,v,z) \longrightarrow \frac{2p_\ell f_\ell(z)\left(\sum_{\substack{i=1\\ m\neq\ell}}^{s} p_m f_i(z)\right)}{\left(\sum_{i=1}^{s} p_i f_i(z)\right)^2} \text{ a.s. .}$$

Hence, using Dominated Convergence Theorem, we get

$$EP^{(11)}_{n_1,\ldots,n_s}(u,v,z) \longrightarrow \frac{p_\ell^2 f_\ell^2(z)}{(\sum p_i f_i(z))^2} \text{ as } n \to \infty \tag{4.17}$$

and

$$EP^{(10)}_{n_1,\ldots,n_s}(u,v,z) \to \frac{2p_\ell f_\ell(z)\sum_{\substack{m=1\\ m\neq\ell}}^{s} p_m f_m(z)}{\left(\sum_{i=1}^{s} p_i f_i(z)\right)^2}. \tag{4.18}$$

Thus from (4.2), (4.3), (4.4), (4.17) and (4.18), we have

$$\begin{aligned}
\pi_\ell^{(1)}(z) &= \lim_{n\to\infty} \pi_\ell^{(1)}(z; n_1,\ldots,n_s) \\
&= \frac{p_\ell^2 f_\ell^2(z)}{\left(\sum_{i=1}^{s} p_\ell f_\ell(z)\right)^2} + \frac{1}{2}\cdot\frac{2p_\ell f_\ell(z)\sum_{\substack{m=1\\ m\neq\ell}}^{s} p_m f_m(z)}{\left(\sum_{i=1}^{s} p_\ell f_\ell(z)\right)^2} \\
&= \widehat{\eta}_\ell + \frac{1}{2}\widehat{\eta}_{\ell 0}.
\end{aligned}$$

The proof is complete. □

Now we derive the limiting TPMC of the RNN rule. First-stage limiting PMC's of the RNN rule are given as follows:

$$\begin{aligned}
\alpha_{\ell j}^{(1)} &= \lim_{n\to\infty} P(\text{ Decide } z \in \pi_\ell / z \in \pi_j) \\
&= \int \pi_\ell^{(1)}(z) f_j(z) dz \\
&= \int \left[\frac{p_\ell^2 f_\ell^2(z)}{\left(\sum_{i=1}^s p_i f_i(z)\right)^2} + \frac{p_\ell f_\ell(z) \sum_{\substack{m=1 \\ m\neq \ell}}^s p_m f_m(z)}{\left(\sum_{i=1}^s p_i f_i(z)\right)^2} \right] f_j(z) dz \\
&= \int \left[\frac{p_\ell f_\ell(z) \left[p_\ell f_\ell(z) + \sum_{\substack{m=1 \\ m\neq \ell}}^s p_m f_m(z) \right]}{\left(\sum_{i=1}^s p_i f_i(z)\right)^2} \right] f_j(z) dz \\
&= \int \frac{p_\ell f_\ell(z) f_j(z)}{\left(\sum_{i=1}^s p_i f_i(z)\right)} dz. \qquad (4.19)
\end{aligned}$$

We may take $p_\ell = \xi_\ell$ as prior probabilities for $\forall\ \ell = 1, 2, \ldots, s$. Therefore the total limiting PMC is given by

$$\begin{aligned}
R_{(s)}^{(1)}(1) &= \sum_{\substack{\ell=1 \\ \ell\neq j}}^s \sum_{j=1}^s \xi_j \alpha_{\ell,j}^{(1)} \\
&= \int \frac{\sum_{\substack{\ell=1 \\ \ell\neq j}}^s \sum_{j=1}^s \xi_\ell f_\ell(z) \xi_j f_j(z)}{\sum_{i=1}^s \xi_\ell f_i(z)} dz \\
&= \int \frac{\sum_{\substack{\ell=1 \\ \ell\neq j}}^s \sum_{j=1}^s \xi_\ell f_\ell(z) \xi_j f_j(z)}{\left(\sum_{i=1}^s \xi_i f_i(z)\right)^2} \sum_{i=1}^s \xi_i f_i(z) dz. \qquad (4.20)
\end{aligned}$$

The Bayes risk for $K = 1$ in (2.9) is given by

$$\begin{aligned}
R_{(s)}^*(1) &= \int \min\left(\sum_{i\neq 1} \xi_i f_i(z), \sum_{i\neq 2} \xi_i f_i(z), \ldots, \sum_{i\neq s} \xi_i f_i(z) \right) dz \\
&= \int \min(1 - \eta_1(z), 1 - \eta_2(z), \ldots, 1 - \eta_s(z)) \sum_{i=1}^s \xi_i f_i(z) dz, \qquad (4.21)
\end{aligned}$$

where

$$\eta_i(z) = \frac{\xi_i f_i(z)}{\sum_{j=1}^{s} \xi_j f_j(z)}$$

THEOREM 4.1.2. *If $f_1, f_2, \ldots, f_s$ are continuous, then $R_{(s)}^{(1)}(1)$ has the following bounds:*

$$R_{(s)}^{*}(1) \le R_{(s)}^{(1)}(1) \le R_{(s)}^{*}(1)\left(2 - \frac{s}{s-1} R_{(s)}^{*}(1)\right)$$

where $R_s^*(1)$ is the Bayes risk given in (2.9).

PROOF. Note that

$$\min\left(\sum_{j\neq 1} \xi_i f_i(z), \sum_{i\neq 2} \xi_i f_i(z), \ldots, \sum_{i\neq s} \xi_i f_i(z)\right)$$
$$\le \sum_{\ell=1}^{s} \frac{\xi_\ell f_\ell(z)}{\sum_{i=1}^{s} \xi_i f_i(z)} \sum_{\substack{j=1 \\ j\neq \ell}}^{s} \xi_i f_j(z).$$

Integrating both sides with respect to z in above, we get

$$R_{(s)}^{*}(1) \le R_{(s)}^{(1)}(1). \tag{4.22}$$

To prove the upper limit, let

$$\begin{aligned} r(z) &= \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \frac{\xi_\ell f_\ell(z) \xi_j f_j(z)}{\left(\sum_{i=1}^{s} \xi_i f_i(z)\right)^2} \\ &= \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \eta_\ell(z)\eta_j(z) \\ &= 1 - \sum_{j=1}^{s} \eta_j^2(z), \end{aligned} \tag{4.23}$$

which in view of (4.20) gives

$$R_{(s)}^{(1)}(1) = Er(z). \tag{4.24}$$

Suppose $\max_j\{\eta_j(z)\} = \eta_\ell(z)$ so that the conditional Bayes risk $r^*(z)$ using (2.8) is given by

$$r^*(z) = 1 - \max_j\{\eta_j(z)\} = 1 - \eta_\ell(z). \tag{4.25}$$

Using Cauchy-Schwartz inequality we have from (4.23) and (4.25)

$$(s-1)\sum_{\substack{\ell=1 \\ \ell\neq j}}^{s}\sum_{j=1}^{s}\eta_j^2(z) \geq \Big[\sum_{\substack{\ell=1 \\ \ell\neq j}}^{s}\sum_{j=1}^{s}\eta_j(z)\Big]^2$$

which implies

$$(s-1)\sum_{j=1}^{s}\eta_j^2(z) \geq (r^*(z))^2 + (s-1)(1-r^*(z))^2$$

i.e.,

$$r(z) \leq 2r^*(z) - \frac{s}{s-1}(r^*(z))^2. \tag{4.26}$$

Therefore, by (4.24), (4.26), (4.21) and Jensen's inequality, we have

$$\begin{aligned} R_{(s)}^{(1)}(1) = Er(z) &\leq 2Er^*(z) - \frac{s}{s-1}E(r^*(z))^2 \\ &\leq 2R_{(s)}^*(1) - \frac{s}{s-1}R_{(s)}^{*2}(1) \\ &= R_{(s)}^*(1)(2 - \frac{s}{s-1}R_{(s)}^*(1)). \end{aligned} \tag{4.27}$$

Thus by (4.22) and (4.27), we get

$$R_{(s)}^*(1) \leq R_{(s)}^{(1)}(1) \leq R_{(s)}^*(2 - \frac{s}{s-1}R_{(s)}^*(1)). \tag{4.28}$$

The proof is complete. □

Note that the inequality obtained in (4.28) is exactly the same as the inequality obtained by Cover and Hart (1967) for s-population case. Now we move to another sub-section where we will obtain the limiting TPMC for the M-stage RNN rule.

4.1.2 Limiting PMC's of the M-stage RNN Rule. Let $\pi_\ell^{(M)}(z; n_1, n_2, \ldots, n_s)$ be the conditional probability that the M-stage RNN rule classifies Z into π_ℓ given $Z = z$. Let

$$\pi_\ell^{(M)}(z) = \lim_{n \to \infty} \pi_\ell^{(M)}(z; n_1, n_2, \ldots, n_s). \tag{4.29}$$

By the definition of $\phi_\ell^{(M)}$ (see (4.1)), we have

$$\begin{aligned}\pi_\ell^{(M)}(z; n_1, \ldots, n_s) &= P\Big[\phi_\ell^{(1)} = 1/Z = z\Big] \\ &+ \sum_{i=1}^{M} P\Big[\phi_\ell^{(1)} = 1/2, \ldots, \phi_\ell^{(i-1)} = 1/2, \phi_\ell^{(i)} = 1/Z = z\Big] \\ &+ \frac{1}{2} \cdot P\Big[\phi_\ell^{(1)} = 1/2, \ldots, \phi_\ell^{(M)} = 1/2/Z = z\Big].\end{aligned} \tag{4.30}$$

Now

$$\begin{aligned}&P\Big[\phi_\ell^{(1)} = 1/2, \ldots, \phi_\ell^{(i-1)} = 1/2, \phi_\ell^{(i)} = 1/Z = z\Big] \\ &= P\Big(\phi_\ell^{(1)} = 1/2/Z = z\Big) \cdot P\Big(\phi_\ell^{(2)} = 1/2, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &\times P\Big(\phi_\ell^{(3)} = 1/2/\phi_\ell^{(2)} = 1/2, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &\quad \ldots\ldots P\Big(\phi_\ell^{(j)} = 1/2/\phi_\ell^{(j-1)} = 1/2, \ldots, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &\ldots\ldots P\Big(\phi_\ell^{(i-1)} = 1/2/\phi_\ell^{(i-2)} = 1/2, \ldots, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &\times P\Big(\phi_\ell^{(i)} = 1/\phi_\ell^{(i-1)} = 1/2, \ldots, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &= P\Big(\phi_\ell^{(1)} = 1/2/Z = z\Big) \\ &\times \prod_{j=2}^{i-1} P\Big(\phi_\ell^{(j)} = 1/2/\phi_\ell^{(j-1)} = 1/2, \ldots, \phi_\ell^{(1)} = 1/2, Z = z\Big) \\ &\times P\Big(\phi_\ell^{(i)} = 1/\phi_\ell^{(i-1)} = 1/2, \ldots, \phi_\ell^{(1)} = 1/2, Z = z\Big).\end{aligned} \tag{4.31}$$

Similarly, we write

$$P\Big[\phi_\ell^{(1)} = 1/2, \dots, \phi_\ell^{(M)} = 1/2, Z = z\Big]$$
$$= P\Big(\phi_\ell^{(1)} = 1/2/Z = z\Big)$$
$$\times \prod_{j=2}^{M} P\Big(\phi_\ell^{(j)} = 1/2/\phi_\ell^{(j-1)} = 1/2, \dots, \phi_\ell^{(1)} = 1/2, Z = z\Big). \tag{4.32}$$

Under certain conditions we shall show that

$$\lim_{n\to\infty} P\Big(\phi_\ell^{(i)} = 1/\phi_\ell^{(i-1)} = 1/2, \dots, \phi_\ell^{(1)} = 1/2, Z = z\Big)$$
$$= \lim_{n\to\infty} P\Big(\phi_\ell^{(1)} = 1/Z = z\Big) = \widehat{\eta}_\ell, \tag{4.33}$$

and

$$\lim_{n\to\infty} P\Big(\phi_\ell^{(i)} = 1/2/\phi_\ell^{(i-1)} = 1/2, \dots, \phi_\ell^{(1)} = 1/2, Z = z\Big)$$
$$= \lim_{n\to\infty} P\Big(\phi_\ell^{(1)} = 1/2/Z = z\Big) = \widehat{\eta}_{\ell 0}. \tag{4.34}$$

Where $\widehat{\eta}_\ell$ and $\widehat{\eta}_{\ell 0}$ are given in (4.11) and (4.12).
Therefore, using (4.29) to (4.34), we have

$$\pi_\ell^{(M)}(z) = \widehat{\eta}_\ell + \widehat{\eta}_\ell \sum_{i=1}^{(M-1)} \widehat{\eta}_{\ell 0}^i + \frac{1}{2}\widehat{\eta}_{\ell 0}^M$$
$$= \widehat{\eta}_\ell\Big(1 + \sum_{i=1}^{M-1} \widehat{\eta}_{\ell 0}^i\Big) + \frac{1}{2}\widehat{\eta}_{\ell 0}^M$$
$$= \widehat{\eta}_\ell \cdot \sum_{i=1}^{M-1} \widehat{\eta}_{\ell 0}^i + \frac{1}{2}\widehat{\eta}_{\ell 0}^M. \tag{4.35}$$

Suppose $\phi_\ell^{(1)} = 1/2$. Delete the observations corresponding to $U^{(1)}$ and $V^{(1)}$ from the pooled ordered training sample. Denote the remaining $(n-2)$ observations by $X_{ij}^{(2)}$.

LEMMA 4.1.3. *Given* $Z = z, \phi_\ell^{(1)} = 1/2, U^{(1)} = u^{(1)}, V^{(1)} = v^{(1)}$, *the conditional distributions of* $X_{\ell j}^{(2)} (j = 1, 2, \ldots, n_\ell - 1)$ *and* $X_{kj}^{(2)} (k \neq \ell)$, *are given as follows:*

(i) $X_{ij}^{(2)}$*'s are mutally independent,*

(ii) the density of $X_{\ell j}^{(2)}$ *is*

$$f_\ell^{(2)}(x) = \frac{f_\ell(x)}{[1 - (F_\ell(v^{(1)}) - F_\ell(u^{(1)}))]} \text{ on } [u_1, v_1]^c \tag{4.36}$$

and

(iii) the density of $X_{kj}^{(2)}$ *is*

$$f_k^{(2)}(x) = \frac{f_k(x)}{[1 - (F_k(v^{(1)}) - F_k(u^{(1)}))]} \text{ on } [u_1, v_1]^c. \tag{4.37}$$

Lemma 4.1.3 can be extended in a similar manner as follows: Let $\phi_\ell^{(m)} = 1/2 (m = 1, 2, \ldots, i-1)$. Delete the observations corresponding to $U^{(m)}$ and $V^{(m)} (m = 1, 2, \ldots, i - 1)$ and denote the remaining $n - 2(i - 1)$ observations by $X_{\alpha\beta}^{(i)}$.

LEMMA 4.1.4. *Given* $Z = z, U^{(m)} = u^{(m)}, V^{(m)} = v^{(m)}$ *and* $\phi_\ell^{(m)} = 1/2$ $(m = 1, 2, \ldots, i - 1)$ *the conditional distributions of* $X_{\ell\beta}^{(i)} (\beta = 1, 2, \ldots, n_\ell - i + 1)$ *and* $X_{k\beta}^{(i)} (k \neq \ell)$ *'s are given as follows:*

(i) $X_{\alpha\beta}^{(i)}$ *'s are mutually independent,*

(ii) the density of $X_{\ell\beta}^{(i)}$ *is*

$$f_\ell^{(i)}(x) = f_\ell^{(i-1)}(x)/[1 - (F_\ell^{(i-1)}(v^{(i-1)}) - F_\ell^{(i-1)}(u^{(i-1)}))] \tag{4.38}$$

on $[u^{(i-1)}, v^{(i-1)}]^c$, *where* $F_\ell^{(i-1)}$ *is c.d.f corresponding to* $f_\ell^{(i-1)}$ *defined inductively*

(iii) and the density of $X_{k\beta}^{(i)}$ *is*

$$f_k^{(i)}(x) = f_k^{(i-1)}(x)/[1 - (F_k^{(i-1)}(v^{(i-1)}) - F_k^{(i-1)}(u^{(i-1)}))] \tag{4.39}$$

on $[u^{(i-1)}, v^{(i-1)}]^c$, *where* $F_\ell^{(i-1)}$ *is c.d.f corresponding to* $f_k^{(i-1)}$, *defined inductively.*

The proofs of the above two lemmas are similar to those of Lemma 3.1 and 3.2 of Das Gupta and Lin (1980), and hence they are omitted.

THEOREM 4.1.3. *If* z *is a continuity point of all* $f_1, \ldots, f_s$ *then the limiting conditional probability of classifying* Z *into* π_ℓ *using M-stage RNN rule, given* $Z = z$, *is given by*

$$\pi_\ell^{(M)}(z) = \lim_{n \to \infty} \pi_\ell^{(M)}(z; n_1, \ldots, n_s)$$

$$= \widehat{\eta}_\ell \cdot \sum_{i=0}^{M-1} \widehat{\eta}_\ell^i + \frac{1}{2} \widehat{\eta}_{\ell 0}^M$$

PROOF. As in (4.5) and (4.6) the conditional probabilities $\phi_\ell^{(i)} = 1$ and $\phi_\ell^{(j)} = 1/2$ given $Z = z, U^{(j)} = u^{(j)}, V^{(j)} = v^{(j)} (j = 1, \ldots, i)$ and $\phi_\ell^{(j)} = 1/2 (j = 1, \ldots, i-1)$, are respectively given by $C_\ell^{(i)} / B^{(i)}(n_1, \ldots, n_s)$ and $C_{\ell 0}^{(i)} / B^{(i)}(n_1, \ldots, n_s)$, where $C_\ell^{(i)}, C_{\ell 0}^{(i)}$ and $B^{(i)}(n_1, \ldots, n_s)$ are obtained from $C_\ell, C_{\ell 0}$ and $B(n_1, \ldots, n_s)$ respectively (see (4.7), (4.8) and (4.9)) after replacing $n_\ell, n_k (k \neq \ell) u, v, f_\ell, f_k$ by $n_\ell - i + 1, n_k - i_k + 1, u^{(i)}, U^{(i)}, f_\ell^{(i)}, f_k^{(i)}$ respectively i_k's are such that $\sum_k i_k = i$. Let

$$p_\ell(i-1) = F_\ell^{(i-1)}(v^{(i-1)}) - F_\ell^{(i-1)}(u^{(i-1)}). \tag{4.40}$$

We note that $p_\ell(i-1) \xrightarrow{\text{a.s}} 0$ as $n \to \infty$ since $v^{(i-1)} \to z$ and $u^{(i-1)} \to z$ a.s. So that from (4.36), (4.38) and (4.40), (Lemma 4.1.3 and 4.1.4) we get

$$f_\ell^{(i)}(x) = \frac{f_\ell(x)}{(1 - p_\ell(1))(1 - p_\ell(2)) \ldots (1 - p_\ell(i-1))}. \tag{4.41}$$

Therefore using (4.41), (Lemma 4.1.2) and continuity of $f_j, (j = 1, \ldots, s)$ at z, we get

$$f_\ell^{(i)}(u^{(i)}) \longrightarrow f_\ell(z) \text{ a.s.}$$

and

$$f_\ell^{(i)}(v^{(i)}) \longrightarrow f_\ell(z) \text{ a.s.}$$

similarly,

$$f_k^{(i)}(u^{(i)}) \longrightarrow f_k(z) \text{ a.s.}$$

and

$$f_k^{(i)}(v^{(i)}) \longrightarrow f_k(z) \text{ a.s. as } n \to \infty.$$

Thus the limiting values of $\frac{C_\ell^{(i)}}{B^{(i)}}$ and $\frac{C_{\ell 0}^{(i)}}{B^{(i)}}$ are $\widehat{\eta}_\ell$ and $\widehat{\eta}_{\ell 0}$ respectively and hence (4.33) and (4.34) hold. Now introducting the set A_1 as in Theorem 4.1.1 and arguing as for (4.2) to (4.4) we get the result (4.35) in view of (4.30) to (4.34). □

Now we obtain limiting TPMC for the M-stage RNN rule.

The limiting PMC's of the M-stage RNN rule are given as follows:

$$\begin{aligned} \alpha_{\ell,j}^{(M)} &= \lim_{n\to\infty} P[\text{M-stage RNN rule decides } z \in \pi_\ell / z \in \pi_j] \\ &= \int \pi_\ell^{(M)}(z) f_j(z) dz. \end{aligned} \tag{4.42}$$

In the limiting case we may take $p_i = \xi_i (i = 1, 2, \ldots s)$ so that the limiting total PMC for the M-stage RNN rule is given by

$$R_{(s)}^{(M)}(1) = \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \xi_j \alpha_{\ell,j}^{(M)}. \tag{4.43}$$

Now we prove a theorem regarding a property of $R_{(s)}^{(M)}(1)$.

THEOREM 4.1.4. $R_{(s)}^{(M)}(1)$ *defined in (4.43) has the following property:*

$$R_{(s)}^{(M)}(1) \leq R_{(s)}^{(M-1)}(1).$$

PROOF. Using (4.42) and (4.43), we write

$$\begin{aligned} R_{(s)}^{(M)}(1) - R_{(s)}^{(M-1)}(1) &= \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \xi_j \left(\alpha_{\ell,j}^{(M)} - \alpha_{\ell,j}^{(M-1)} \right) \\ &= \int \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \xi_j f_j(z) (\pi_\ell^{(M)}(z) - \pi_\ell^{(M-1)}(z)) dz. \end{aligned} \tag{4.44}$$

We also know from (4.35)

$$\pi_\ell^{(M)}(z) = \widehat{\eta}_\ell \sum_{i=0}^{M-1} \widehat{\eta}_{\ell 0}^i + \frac{1}{2}\widehat{\eta}_{\ell 0}^M$$

so that by applying (4.11), (4.12) and (2.1), we have

$$\begin{aligned}\pi_\ell^{(M)} - \pi_\ell^{(M-1)} &= \widehat{\eta}_\ell \cdot \widehat{\eta}_{\ell 0}^{(M-1)} + \frac{1}{2}\widehat{\eta}_{\ell 0}^{(M-1)}(\widehat{\eta}_{\ell 0} - 1) \\ &= \widehat{\eta}_{\ell 0}^{(M-1)}[\widehat{\eta}_\ell + \frac{1}{2}\widehat{\eta}_{\ell 0} - \frac{1}{2}] \\ &= \widehat{\eta}_{\ell 0}^{(M-1)}\left[\frac{\xi_\ell^2 f_\ell^2(z)}{(\sum \xi_i f_i(z))^2} + \frac{\xi_\ell f_\ell(z)\sum_{j\neq\ell}\xi_j f_j(z)}{(\sum \xi_i f_i(z))^2} - \frac{1}{2}\right] \\ &= \widehat{\eta}_{\ell 0}^{(M-1)}\left[\frac{\xi_\ell f_\ell(z)}{\sum_{i=1}^s \xi_i f_i(z)} - \frac{1}{2}\right] = \widehat{\eta}_{\ell 0}^{(M-1)}(\eta_\ell - \frac{1}{2}). \end{aligned} \tag{4.45}$$

Now writing (4.45) as

$$\pi_\ell^{(M)}(z) - \pi_\ell^{(M-1)}(z) = -\frac{1}{2}\widehat{\eta}_{\ell 0}^{(M-1)}(1 - 2\eta_\ell) \tag{4.46}$$

we get from (4.44) and (4.46),

$$\begin{aligned} R_{(s)}^{(M)}(1) - R_{(s)}^{(M-1)}(1) &= \int \frac{\sum_{\ell=1}^s \sum_{\substack{j=1 \\ j\neq\ell}}^s \xi_i f_j(z)}{\sum_{i=1}^s \xi_i f_i(z)}(\pi_\ell^{(M)}(z) - \pi_\ell^{(M-1)}(z))\sum_{i=1}^s \xi_i f_i(z)dz \\ &= -\frac{1}{2}\int \sum_{\ell=1}^s (1-\eta_\ell)(1-2\eta_\ell)\widehat{\eta}_{\ell 0}^{(M-1)} \sum_{i=1}^s \xi_i f_i(z)dz. \end{aligned} \tag{4.47}$$

Again we write $\widehat{\eta}_{\ell 0}$ in (4.12) using (2.1), as

$$\begin{aligned} \widehat{\eta}_{\ell 0} &= \frac{2\xi_\ell f_\ell(z)\sum_{j\neq\ell}\xi_j f_j(z)}{\{\sum \xi_i f_i(z)\}^2} \\ &= 2\eta_\ell(1-\eta_\ell). \end{aligned} \tag{4.48}$$

Thus, by (4.47) and (4.48), we have

$$\begin{aligned} R_{(s)}^{(M)}(1) - R_{(s)}^{(M-1)}(1) = -2^{M-2}\int \sum_{\ell=1}^s \eta_\ell^{M-1}(1-\eta_\ell)^{M-2}(1-\eta_\ell)^2(1-2\eta_\ell) \\ \times \sum_{i=1}^s \xi_i f_i(z)dz. \end{aligned}$$

Now we apply the fact that $-(1-\eta_\ell)^2 \le -(1-2\eta_\ell)$ in the above to obtain

$$
\begin{aligned}
&R_{(s)}^{(M)}(1) - R_{(s)}^{(M-1)}(1) \\
&\quad \le -2^{M-2} \int \sum_{\ell=1}^{s} \eta_\ell^{M-1}(1-\eta_\ell)^{M-2}(1-2\eta_\ell)^2 \sum_{i=1}^{s} \xi_i f_i(z)dz \\
&\quad = -2^{M-2} \sum_{i=1}^{s} E[((1-\eta_\ell)\eta_\ell)^{M-2}\eta_\ell(1-2\eta_\ell)^2] \qquad (4.49) \\
&\quad \Rightarrow R_{(s)}^{(M)}(1) \le R_{(s)}^{(M-1)}(1). \qquad (4.50)
\end{aligned}
$$

The proof is complete. □

4.1.3. **Estimation of PMC's of the one-stage RNN rule.** We shall estimate the PMC's of the one-stage RNN rule by the deleted counting method which we describe as follows. Let

$$\psi_\ell^{(k)} = 1 - \phi_\ell^{(1)}(X_{\ell k}; X_{ij}, i = 1, \ldots, s,\ j = 1, \ldots, n_i,\ j \ne k) \qquad (4.51)$$

where $\phi_\ell^{(1)}$ is defined in (4.1).

Let

$$p_\ell(n_1, \ldots, n_s) = \sum_{k=1}^{n_\ell} \psi_\ell^{(k)}/n_\ell \quad \ell = 1, 2, \ldots, s \qquad (4.52)$$

and

$$p(n_1, \ldots, n_s) = \frac{n_1 p_1(n_1, \ldots n_s) + \ldots + n_s p_s(n_1, \ldots, n_s)}{n_1 + n_2 + \ldots + n_s}. \qquad (4.53)$$

Then p_ℓ's $\ell = 1, 2, \ldots, s$ are estimates of PMC's. From (4.2), (4.51) and (4.52), we note that

$$Ep_1(n_1, \ldots, n_s) = \int (1 - \pi_1^{(1)}(z; n_1 - 1, n_2, \ldots, n_s)) f_1(z)dz,$$

$$\vdots$$

$$Ep_\ell(n_1, \ldots, n_s) = \int (1 - \pi_\ell^{(1)}(z; n_1, n_2, \ldots, n_\ell - 1, \ldots, n_s)) f_\ell(z)dz,$$

$$\vdots$$

$$Ep_s(n_1, \ldots, n_s) = \int (1 - \pi_s^{(1)}(z; n_1, n_2, \ldots, n_s - 1)) f_s(z)dz.$$

Therefore

$$E\sum_{\ell=1}^{s} p_\ell(n-1,\ldots,n_s) = \int \sum_{\ell\neq 1} \pi_\ell^{(1)}(z;n_1-1,n_2,\ldots,n_s)f_1(z)dz$$
$$+\int \sum_{\ell\neq 2} \pi_\ell^{(1)}(z;n_1,n_2-1,\ldots n_s)f_2(z)dz$$
$$\vdots$$
$$+\int \sum_{\ell\neq s} \pi_\ell^{(1)}(z;n_1,n_2,\ldots n_s-1)f_s(z)dz$$

so that from above we have

$$E\sum_{\ell=1}^{s}\left(\frac{n_\ell p_\ell(n-1,\ldots,n_s)}{\sum n_i}\right) = \int \frac{n_1}{\sum n_i}\sum_{\ell\neq 1} \pi_\ell^{(1)}(z;n_1-1,n_2,\ldots,n_s)f_1(z)dz$$
$$+\int \frac{n_2}{\sum n_i}\sum_{\ell\neq 2} \pi_\ell^{(1)}(z;n_1,n_2-1,\ldots n_s)f_2(z)dz$$
$$\vdots$$
$$+\int \frac{n_s}{\sum n_i}\sum_{\ell\neq s}^{s} \pi_\ell^{(1)}(z;n_1,\ldots n_s-1)f_s(z)dz. \tag{4.54}$$

Let

$$\xi_\ell = \lim_n \frac{n_\ell}{\sum n_i}. \tag{4.55}$$

By (4.53) to (4.55) and the (Lebesque) Dominated Convergence Theorem, we get

$$\lim_n E[p(n_1,\ldots,n_s)] = \lim_n E\Bigg[\sum_{\substack{\ell=1\\ \ell\neq j}}^{s}\sum_{j=1}^{s} \frac{n_j}{\sum n_i}\pi_\ell^{(1)}(z;n_1,\ldots n_\ell-1,\ldots,n_s)\Bigg]$$
$$= E\sum_{\substack{\ell=1\\ \ell\neq j}}^{s}\sum_{j=1}^{s} \xi_j\pi_\ell^{(1)}(z)$$
$$= \sum_{\substack{\ell=1\\ \ell\neq j}}^{s}\sum_{j=1}^{s} \xi_j\int \pi_\ell^{(1)}(z)f_j(z)dz$$
$$= \sum_{\substack{\ell=1\\ \ell\neq j}}^{s}\sum_{j=1}^{s} \xi_j\alpha_{\ell,j}^{(1)} = R_{(s)}^{(1)}(1). \tag{4.56}$$

From (4.56) it follows that $p(n_1, \ldots, n_s)$ is asymptocally unbiased for $R_{(s)}^{(1)}(1)$. Now order the combined observations and let $r_\ell(\ell = 1, 2, \ldots, s)$ be the number of runs for the observations from ℓ-th population. We note that

$$n_\ell p_\ell(n_1, \ldots, n_2) = r_\ell + \delta_\ell, \tag{4.57}$$

where $|\delta_\ell| \leq 1$, δ_ℓ's $(\ell = 1, 2, \ldots, s)$ are the contributions arising from the extreme observations. Now, by (4.52) and (4.57), we write

$$\begin{aligned} \lim_n E(r_\ell / n_\ell) &= \lim_n E p_\ell(n_1, \ldots, n_s) \\ &= \int (1 - \pi_\ell^{(1)}(z)) f_\ell(z) dz \\ &= \int \sum_{\substack{j=1 \\ j \neq \ell}}^{s} \pi_j^{(1)}(z) f_\ell(z) dz \end{aligned} \tag{4.58}$$

and let r be the total number of runs; then from (4.56) and (4.58) we have

$$\begin{aligned} \lim_n E\left(\frac{r}{\sum n_i}\right) &= \lim_{n \to \infty} E\left(\frac{\sum n_\ell p_\ell(n_1, \ldots, n_s)}{\sum n_i}\right) \\ &= \sum_{\substack{\ell=1 \\ \ell \neq j}}^{s} \sum_{j=1}^{s} \xi_j \int \pi_\ell^{(1)}(z) f_j(z) dz \\ &= \sum_{\substack{\ell=1 \\ \ell \neq j}}^{s} \sum_{j=1}^{s} \xi_j \alpha_{\ell, j}^{(1)} = R_{(s)}^{(1)}(1). \end{aligned} \tag{4.59}$$

Let U_ℓ be a U-statistic defined by

$$U_\ell = \frac{1}{n_\ell} \sum_{k=1}^{n_\ell} E(\psi_\ell^{(k)} / X_{\ell k}). \tag{4.60}$$

It follows from Theorem A of Serfling (1980, p. 190) we get, as $n \to \infty$,

$$U_\ell \xrightarrow{\text{a.s.}} \sum_{j \neq \ell} \alpha_{j, \ell}^{(1)} \tag{4.61}$$

We will now show that $E(p_\ell - U_\ell)^2 = O(n^{-1})$ which will establish that

$$p_\ell - U_\ell \xrightarrow{Pr} 0, \tag{4.62}$$

so that from (4.61) and (4.62), we may conclude that

$$p_\ell \xrightarrow{p_r} \sum_{j\neq\ell} \alpha_{j,\ell}^{(1)}. \tag{4.63}$$

LEMMA 4.1.5. *$E(p_\ell - U_\ell)^2 = O(1/n)$, where p_ℓ and U_ℓ are as defined in (4.52) and (4.60) repsecitvely.*

PROOF. Using (4.52) and (4.60), we write

$$(p_\ell - U_\ell) = \frac{1}{n_\ell}\sum_{k=1}^{n_\ell}(\psi_\ell^{(k)} - E(\psi_\ell^{(k)}/X_{\ell k}))$$

from which we get

$$(p_\ell - U_\ell)^2 = \frac{1}{n_\ell^2}\Big\{\sum_{k=1}^{n_\ell}(\psi_\ell^{(k)} - E(\psi_\ell^{(k)}/X_{\ell k}))^2$$
$$+ 2\sum\sum_{k<m}(\psi_\ell^{(k)} - E(\psi_\ell^{(k)}/X_{\ell k}))(\psi_\ell^{(m)} - E(\psi_\ell^{(m)}/X_{\ell m}))\Big\}. \tag{4.64}$$

Now

$$\begin{aligned}
E(\psi_\ell^{(k)} - E(\psi_\ell^{(k)}/X_{\ell m}))^2 &= EVar(E(\psi_\ell^{(k)}/X_{\ell m})) \\
&= E\{E(\psi_\ell^{(k)2}/X_{\ell m}) - (E(\psi_\ell^{(k)}/X_{\ell m}))^2\} \\
&\leq E\{E(\psi_\ell^{(k)}/X_{\ell m}) - (E(\psi_\ell^{(k)}/X_{\ell m}))^2\} \\
&= E\{(1 - \pi_\ell^{(1)}(X_{\ell k};\, n_1, \ldots, n_{\ell_1}, \ldots, n_2) \\
&- (1 - \pi_\ell^{(1)}(X_{\ell k};\, n_1, \ldots, n_{\ell_1}, \ldots, n_2))^2\} \\
&= E\pi_\ell^{(1)}(X_{\ell k}; n_1, \ldots, n_\ell - 1, \ldots, n_s) \\
&\times (1 - \pi_\ell^{(1)}(X_{\ell k}; n_1, \ldots, n_\ell - 1, \ldots, n_s)) \\
&\leq \frac{1}{4}
\end{aligned} \tag{4.65}$$

and

$$
\begin{aligned}
&E(\psi_\ell^{(k)} - E(\psi_\ell^{(k)}/X_\ell))(\psi_\ell^{(m)} - E(\psi_\ell^{(m)}/X_{\ell m})) \\
&= E\{E(\psi_\ell^{(k)}/X_\ell) - E(\psi_\ell^{(k)}/X_{\ell m}))(\psi_\ell^{(m)} - E(\psi_\ell^{(m)}/X_{\ell m}))/X_{\ell k}, X_{\ell m}\} \\
&= E[E\{\psi_\ell^{(k)}\psi_\ell^{(m)} - \psi_\ell^{(k)}E(\psi_\ell^{(m)}/X_{\ell m}) - \psi_\ell^{(m)}E(\psi_\ell^{(k)}/X_{\ell k}) \\
&\qquad + E(\psi_\ell^{(k)}/X_{\ell k})\cdot E(\psi_\ell^{(m)}/X_{\ell m}\}/X_{\ell k}, X_{\ell m}\}] \\
&= E[E(\psi_\ell^{(k)}\psi_\ell^{(m)}/X_{\ell k}X_{\ell m}) - E(\psi_\ell^{(k)}\cdot E(\psi_\ell^{(m)}/X_{\ell m})/X_{\ell k}, X_{\ell m}) \\
&- E(\psi_\ell^{(m)}\cdot E(\psi_\ell^{(k)}/X_{\ell k})/X_{\ell k}, X_{\ell m}) + E(\psi_\ell^{(k)}/X_{\ell k})\cdot E(\psi_\ell^{(m)}/X_{\ell m})] \\
&= E[E(\psi_\ell^{(k)}/X_{\ell k})\cdot E(\psi_\ell^{(m)}/X_{\ell m}) - E(\psi_\ell^{(m)}/X_{\ell m})\cdot E(\psi_\ell^{(k)}/X_{\ell k}) \\
&- E(\psi_\ell^{(k)}/X_{\ell k})\cdot E(\psi_\ell^{(m)}/X_{\ell m}) + E(\psi_\ell^{(k)}/X_{\ell k})\cdot E(\psi_\ell^{(m)}/X_{\ell m})] \\
&= 0. \qquad (4.66)
\end{aligned}
$$

Hence, from (4.64) to (4.66), we get

$$E(p_\ell - U_\ell)^2 = O\left(\frac{1}{n_\ell}\right).$$

The proof is complete. □

Thus from (4.53), (4.55) and (4.63) we note that

$$
\begin{aligned}
p(n_1,\ldots,n_s) &\xrightarrow{P} \xi_1 \sum_{j\neq 1} \alpha_{j,1}^{(1)} + \xi_2 \sum_{j\neq 2} \alpha_{j,2}^{(1)} + \ldots + \xi_s \sum_{j\neq s} \alpha_{j,s}^{(1)} \\
&= \sum_{\substack{\ell=1 \\ \ell\neq j}}^{s} \sum_{j=1}^{s} \xi_\ell \alpha_{j\ell} = R_s^{(1)}(1)
\end{aligned}
$$

Similar estimates for limiting PMC's of the M-stage RNN rule can be obtained.

4.2 Classification of multiple univariate observations to one of s populations using left and right Rank Nearest Neighbors.

Here we consider a problem of classifing K univariate observations to one among s univariate populations using first-stage RNN rule. This would be an extension of the first-stage RNN rule given in (3.1) for $s = 2$ to the general s populations. Let $X_{ij}, j = 1, 2, \ldots, n_i$ be the random training sample from the i-th population $\pi_i, i = 1, 2, \ldots, s$. Suppose we have a random sample $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$

to be classified to one of s populations. We propose the classification rule $\delta_{KD}^{(s)}$ as follows:

Combine X_{ij}'s and Z_ℓ's $i = 1, 2, \ldots, s, j = 1, 2, \ldots, n_i, \ell = 1, 2, \ldots, K$ and arrange them in ascending order , then identify the first left and right hand Rank Nearest Neighbors of Z_ℓ for all $\ell = 1, 2, \ldots, K$. Let m_i be the number of Rank Nearest Neighbors identified as from being the population $\pi_i, i = 1, 2, \ldots, s$. Then classify $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ to the ℓ-th population π_ℓ

$$\begin{aligned}
&\text{if } m_\ell = m^*, \text{ where } m^* \text{ is unique maximum of } \{m_i,\ 1 \le i \le s\};\\
&\text{if } \ m^* = m_{i_1} = m_{i_2} = \ldots = m_{i_J}, \text{ then classify}\\
&\underset{\sim}{Z} \ \text{ to } \pi_\ell \text{ with probability } 1/J \text{ for } \ell = i_1, i_2, \ldots i_J.
\end{aligned} \tag{4.67}$$

Now to obtain limiting total probability of misclassification for the rule $\delta_{KD}^{(s)}$ described in (4.67), one should follow similar type of argument given in Section 3.1 of Chapter 3. Because of its natural cumbersome expressions we do not include a discussion on them.

4.3 Classification of multiple ($K \ge 2$) multivariate observations to one of s populations using 1-NN rule in Bayesian model.

Suppose we have a random identified training sample $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ taking values in $\mathbb{R}^d \times \{1, 2, \ldots, s\}$, where $\{\theta_i\}_{i=1,2,\ldots,n}$ are i.i.d discrete random variable with $P(\theta_i = j) = \xi_j, i = 1, \ldots, n, j = 1, \ldots, s$ and $\sum_{j=1}^{s} \xi_j = 1$. The object is to classify $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ to one of s populations $\pi_1, \pi_2, \ldots, \pi_s$. Let X'_{nj} be the NN of Z_j (see (2.15)) for all $j = 1, 2, \ldots, K$. Let m_i be the number of Nearest Neighbor observations identified as being from the population $\pi_i, i = 1, 2, \ldots, s$. Now we define a classification rule $\delta_{KB}^{(s)}$ as follows:

Classify $\underset{\sim}{Z}$ to population π_ℓ

$$\begin{aligned}
&\text{if } m_\ell = m^*, \text{ where } m^* \text{ is unique maximum of}\\
&\{m_i,\ 1 \le i \le s\}; \text{ if } m^* = m_{i_1} = m_{i_2} = \ldots = m_{i_J}, \text{ then classify}\\
&\underset{\sim}{Z} \ \text{ to } \pi_\ell \text{ with probability } \frac{1}{J} \text{ for } \ell = i_1, i_2, \ldots, i_J.
\end{aligned} \tag{4.68}$$

Now we will obtain the limiting risk of the rule $\delta^{(s)}_{KB}$, in the case when $K = 2$ observations are to be classified. Suppose the real category of $\underset{\sim}{Z}$ is θ and its estimate θ'_n. We will use the $(0-1)$ loss defined in (2.7). so that the conditional risk given $\underset{\sim}{Z} = (z_1, z_2)$ and $\underset{\sim}{X}'_n = (x'_{n1}, x'_{n2})$, is given by

$$\begin{aligned} r(\underset{\sim}{z}; x'_n) &= E\{L(\theta, \theta'_n) / \underset{\sim}{z}, x'_n\} \\ &= P(\theta \neq \theta'_n / \underset{\sim}{z}, \underset{\sim}{x}'_n); \end{aligned}$$

now using the conditional independence of θ and θ'_n in above, we get

$$\begin{aligned} r(\underset{\sim}{z}; \underset{\sim}{x}'_n) &= \sum_{\substack{i=1 \\ i \neq \ell}}^{s} \sum_{\ell=1}^{s} P(\theta = i / \underset{\sim}{z}) \cdot P(\theta'_n = \ell / \underset{\sim}{x}'_n) \\ &= \sum_{\substack{i=1 \\ i \neq \ell}}^{s} \sum_{\ell=1}^{s} \eta_i^*(\underset{\sim}{z}) \cdot \pi_\ell(\underset{\sim}{x}'_n), \end{aligned} \tag{4.69}$$

where $\eta_i^*(\underset{\sim}{z})$ for $K = 2$ is given as in (2.2) and

$$\begin{aligned} \pi_\ell(\underset{\sim}{x}'_n) &= P(\theta'_n = \ell / \underset{\sim}{x}'_n) \\ &= \frac{\xi_\ell^2 f_\ell(\underset{\sim}{x}'_n) f_\ell(\underset{\sim}{x}'_n)}{B_n} \\ &+ \frac{1}{2} \frac{\xi_\ell \sum_{m \neq \ell}^{s} \xi_m (f_\ell(\underset{\sim}{x}'_{n1}) f_m(\underset{\sim}{x}'_{n2}) + f_\ell(\underset{\sim}{x}'_{n2}) f_m(\underset{\sim}{x}'_{n1}))}{B_n}, \end{aligned} \tag{4.70}$$

$\ell = 1, 2, \ldots, s$, and

$$\begin{aligned} B_n &= \sum_{\ell=1}^{s} \xi_\ell^2 f_\ell(\underset{\sim}{x}'_{n1}) f_\ell(\underset{\sim}{x}'_{n2}) + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{s} \sum_{m=1}^{s} \xi_\ell \xi_m (f_\ell(\underset{\sim}{x}'_{n1}) f_m(\underset{\sim}{x}'_{n2}) + f_\ell(\underset{\sim}{x}'_{n2}) f_m(\underset{\sim}{x}'_{n1})) \\ &= (\sum_{\ell=1}^{s} \xi_\ell f_\ell(\underset{\sim}{x}'_{n1}))(\sum_{\ell=1}^{s} \xi_\ell f_\ell(\underset{\sim}{x}'_{n2})). \end{aligned} \tag{4.71}$$

Now we prove the theorem on limiting risk.

THEOREM 4.3.1. *Suppose z_1 and z_2 are continuity points of all $f_i, i = 1, 2, \ldots, s$. Then the limiting conditional risk given $\underset{\sim}{z} = (z_1, z_2)$ and the limiting unconditional risk for the rule $\delta_{2B}^{(s)}$, are given respectively by*

$$r(\underset{\sim}{z}) = \lim_{n\to\infty} r(\underset{\sim}{z}; \underset{\sim n}{x'}) = \sum_{\substack{i=1 \\ i\neq\ell}}^{s} \sum_{\ell=1}^{s} \eta_i^*(\underset{\sim}{z}) \cdot \pi_\ell(\underset{\sim}{z}) \quad a.s.$$

and

$$R_{(s)}(2) = \lim_{n\to\infty} Er(\underset{\sim}{z}; \underset{\sim n}{x'}) = E\{r(\underset{\sim}{z})\},$$

where

$$\begin{aligned}\pi_\ell(\underset{\sim}{z}) = & \frac{\xi_\ell^2 f_\ell(z_1) f_\ell(z_2)}{(\sum_{j=1}^{s} \xi_j f_j(z_1))(\sum_{j=1}^{s} \xi_j f_j(z_2))} \\ & + \frac{1}{2} \frac{\xi_\ell \sum_{m\neq\ell} \xi_m (f_\ell(z_1) f_m(z_2) + f_\ell(z_2) f_m(z_1))}{(\sum_{j=1}^{s} \xi_j f_j(z_1))(\sum_{j=1}^{s} \xi_j f_j(z_2))}, \\ & \ell = 1, 2, \ldots, s. \end{aligned} \tag{4.72}$$

PROOF. By the continuity of f_i, at $\underset{\sim}{z}, i = 1, 2, \ldots, s$, Lemma 2.1.1 and (4.70), we have

$$\pi_\ell(\underset{\sim n}{x'}) \xrightarrow{\text{a.s.}} \pi_\ell(\underset{\sim}{z}) \tag{4.73}$$

which by (4.69) and (4.73), yields

$$\begin{aligned} r(\underset{\sim}{z}) &= \lim_n r(\underset{\sim}{z}; \underset{\sim n}{x'}) \\ &= \sum_{\substack{i=1 \\ i\neq\ell}}^{s} \sum_{\ell=1}^{s} \eta_i^*(\underset{\sim}{z}) \cdot \pi_\ell(\underset{\sim}{z}). \end{aligned} \tag{4.74}$$

Now using the Dominated Convergence thoerem we obtain

$$R_{(s)}(2) = \lim_{n\to\infty} E\{r(\underset{\sim}{z}; \underset{\sim n}{x'})\} = E\{r(\underset{\sim}{z})\}. \tag{4.75}$$

The proof is complete. □

Now from (4.72), (4.74) and (4.75), $R_{(s)}(2)$ can be expressed as:

$$R_{(s)}(2) = \frac{1}{2}\int \sum_{\substack{i=1 \\ i\neq\ell}}^{s}\sum_{\ell=1}^{s} \eta_i^*(\underset{\sim}{z})$$

$$\times \frac{2\xi_\ell^2 f_\ell(z_1)f_\ell(z_2) + \xi_\ell f_\ell(z_1)\sum_{m\neq\ell}^{s}\xi_m f_m(z_2) + \xi_\ell f_\ell(z_2)\sum_{m\neq\ell}^{s}\xi_m f_m(z_1)}{(\sum_{m=1}^{s}\xi_m f_m(z_1))(\sum_{m=1}^{s}\xi_m f_m(z_2))}$$

$$\times \sum_{m=1}^{s}\xi_m f_m(z_1)f_m(z_2)dz_1dz_2$$

$$= \frac{1}{2}\int \sum_{\substack{i=1 \\ i\neq\ell}}^{s}\sum_{\ell=1}^{s} \eta_i^*(\underset{\sim}{z})\frac{\xi_\ell f_\ell(z_1)\sum_{m=1}^{s}\xi_m f_m(z_2) + \xi_\ell f_\ell(z_2)\sum_{m=1}^{s}\xi_m f_m(z_1)}{(\sum_{m=1}^{s}\xi_m f_m(z_2))(\sum_{m=1}^{s}\xi_m f_m(z_1))}$$

$$\times \sum_{m=1}^{s}\xi_m f_m(z_1)f_m(z_2)dz_1dz_2$$

$$= \int \sum_{\substack{i=1 \\ i\neq\ell}}^{s}\sum_{\ell=1}^{s} \frac{1}{2}(\eta_\ell(z_1) + \eta_\ell(z_2))\cdot\xi_i f_i(z_1)f_i(z_2)dz_1dz_2,$$

where $\eta_\ell(z_i), i = 1, 2$, in the above is defined in (2.1).

$$= \sum_{\substack{i=1 \\ i\neq\ell}}^{s}\sum_{\ell=1}^{s} \frac{1}{2}[\int \eta_\ell(z_1)\xi_i f_i(z_1)dz_1 + \int \eta_\ell(z_2)\xi_i f_i(z_2)dz_2]$$

$$= \int \sum_{\substack{i=1 \\ i\neq\ell}}^{s}\sum_{\ell=1}^{s} \eta_\ell(z_1)\eta_i(z_1)\cdot\sum_{i=1}^{s}\xi_i f_i(z_1)dz_1$$

$$= \int r(z_1)\cdot\sum_{i=1}^{s}\xi_i f_i(z_1)dz_1. \text{ (say)} \tag{4.76}$$

Now we obtain bounds on $R_{(s)}(2)$ in the following

THEOREM 4.3.2. *The limiting risk $R_{(s)}(2)$ of the rule $\delta_{2H}^{(s)}$ has the following bounds*

$$R_{(s)}^*(1) \leq R_{(s)}(2) \leq 2R_{(s)}^*(1)(1 - R_{(s)}^*(1)). \tag{4.77}$$

where $R^*_{(s)}(1)$ is the Bayes risk (see (2.9)).

PROOF . From (4.76), write

$$r(z_1) = \sum_{\substack{i=1 \\ i\neq \ell}}^{s} \sum_{\ell=1}^{s} \eta_\ell(z_1)\eta_1(z_1)$$

$$= 1 - \sum_{j=1}^{s} \eta_j^2(z_1). \tag{4.78}$$

Note that if $\eta_\ell(z_1) = \max_j\{\eta_j(z_1)\}$, the conditional Bayes risk $r^*(z_1)$ in (2.8), can be written as

$$r^*(z_1) = 1 - \max_j\{\eta_j(z_1)\}$$

$$= 1 - \eta_\ell(z_1). \tag{4.79}$$

By the Cauchy-Schwartz inequality and (4.79), we have

$$(s-1)\sum_{j\neq\ell} \eta_j^2(z_1) \geq [\sum_{j\neq\ell}^{s} \eta_j(z_1)]^2$$

$$= [1-\eta_\ell(z_1)]^2 = (r^*(z_1))^2.$$

Thus, adding $(s-1)\eta_\ell^2(z_1)$ both sides in above, we get

$$\sum_{j=1}^{s} \eta_j^2(z_1) \geq \frac{(r^*(z_1))^2}{(s-1)} + (1 - r^*(z_1))^2. \tag{4.80}$$

From (4.78) and (4.80), we have

$$r(z_1) \leq 2r^*(z_1) - \frac{s}{s-1}(r^*(z_1))^2 \tag{4.81}$$

so that by (4.76), (4.81) and Jensen's inequality, we obtain

$$R_{(s)}(2) \leq 2Er^*(z_1) - \frac{s}{s-1}E(r^*(z_1))^2,$$

$$\leq 2R^*_{(s)}(1) - \frac{s}{s-1}[Er^*(z_1)]^2,$$

$$= 2R^*_{(s)}(1) - \frac{s}{s-1}R^{*2}_{(s)}(1)$$

$$= R^*_{(s)}(1)(2 - \frac{s}{s-1}R^*_{(s)}(1)). \tag{4.82}$$

Also, note that

$$\min(\sum_{i\neq 1}\xi_i f_i(z_1), \sum_{i\neq 2}\xi_i f_i(z_1) \dots, \sum_{i\neq s}\xi_i f_i(z_1))$$
$$\leq \sum_{j=1}^{s} \frac{\xi_j f_j(z_1)}{\sum_{i=1}^{s}\xi_i f_i(z_1)} \sum_{\substack{i=1 \\ i\neq j}}^{s} \xi_i f_i(z_1)$$

so that by integrating both sides w.r.t. z_1, we get

$$R^*_{(s)}(1) \leq R_{(s)}(2). \tag{4.83}$$

The proof is complete in view of (4.82) and (4.83). □

Also, similar results can be obtained by using the above arguments for general $K > 2$; the details are lengthy and hence are omitted. We also note that in the nonparametric model, similar a limiting risk for $K \geq 2$ can also be obtained.

CHAPTER 5.

CLASSIFICATION OF MULTIPLE OBSERVATIONS USING SUB-SAMPLE APPROACH

5.0 **Introduction** . In this chapter, we propose a 1-Nearest Neighbor discriminant rule for the classification of multiple observations $\{Z_1, \ldots, Z_K\}$ among two or more populations based on sub-groupings of size K from the training samples. This rule generalizes the Cover and Hart's (1967) 1-Nearest Neighbor rule to K observations. Here we prove the almost sure convergence of Nearest Neighbor and derive its asymptotic risk. We show that asymptotic risk has a bound that is parallel to the one obtained by Cover and Hart (1967). We also propose an estimator of the risk function and show that it is unbiased and consistent.

Let $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ be i.i.d random vectors from $\mathbb{R}^d \times \{1, 2, \ldots, s\}$ with $P(\theta_i = j) = \xi_j$, $j = 1, 2, \ldots, s$ and $\sum_{j=1}^{s} \xi_j = 1$ where ξ_j's are called the prior probabilities. If θ_i takes value $j, (i.e.\ \theta_i = j)$ then we say X_i is from population π_j, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, s$. Let $(Z_1, \theta), (Z_2, \theta), \ldots, (Z_K, \theta)$ be a random sample from one of the distinct populations among $\pi_1, \pi_2, \ldots, \pi_s$. Assume that π_j has density f_j (w.r. to Lebesque measure μ on R^d) $j = 1, 2, \ldots, s$, and use the notation $((Z_1, \theta), (Z_2, \theta) \ldots (Z_K, \theta)) = (\underset{\sim}{Z}, \theta)$ noting that only $\underset{\sim}{Z}$ is observable. It is desired to estimate θ by making use of the information contained in the training sample described above. Assume that the training sample has been identified. We shall now describe the sub-sample approach as follows:

Let n_j be the numbers of X-observations in the training sample identified as being from the population $\pi_j, j = 1, \ldots, s$. Thus $\sum_{j=1}^{s} n_j = n$ where we also assume $K \leq \min(n_1, n_2, \ldots, n_s)$ without loss of generality.

Consider the set of all possible sub-samples from each of the identified parts of the training samples corresponding to the populations $\pi_1, \pi_2, \ldots, \pi_s$ *viz.*,

$$\mathcal{S} = \{\underset{\sim}{Y} = (X_{i_1}, X_{i_2}, \ldots, X_{i_K})^T,\ \theta_{i_1} = \theta_{i_2} = \ldots = \theta_{i_K} = \tau_{(i_1 \ldots i_K)}, i_1 \neq i_2 \neq \ldots \neq i_K\}. \tag{5.1}$$

Suppose $\{\tau_m\}$ is a sequence of i.i.d random variables with $P(\tau_m = j) = \xi_j \; \forall \; j = 1, \ldots, s$ and $\sum_{j=1}^{s} \xi_j = 1$. Thus we treat the sequence $\{\underset{\sim m}{Y}, \tau_m\}_{m=1,2,\ldots,N}$ as an identically distributed training sample with Range space $\mathbb{R}^{Kd} \times \{1, 2, \ldots, s\}$, Where $N = n_{1_{P_K}} + n_{2_{P_K}} + \ldots + n_{s_{P_K}}$. The symbol n_{P_r} denote the number of permutations of n things taken r at a time. We shall now define the "sub-sample" classification procedure δ_s^* : Denote by $\underset{\sim N}{Y'} \in \{\underset{\sim 1}{Y}, \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}\}$ the Nearest Neighbor to $\underset{\sim}{Z}$, that is, if

$$\min_m \| \underset{\sim m}{Y} - \underset{\sim}{Z} \| = \| \underset{\sim N}{Y'} - \underset{\sim}{Z} \|; \tag{5.2}$$

suppose $\underset{\sim N}{Y'}$ is identified as from the population corresponding to τ'_N, say, π_j, then classify $\underset{\sim}{Z}$ to the population $\pi_j, j = 1, \ldots, s$. A mistake is made in classification if $\tau \neq \tau'_N$, where we have set $\theta = \tau$, the true category of $\underset{\sim}{Z}$.

5.1 Asymptotic NN risk when $K = 2$ observations are to be classified between two populations

In this section we shall treat the simpler case $s = 2$ and $K = 2$ for convenience of presentation – the general case is treated in the next section. Let $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ be an identified training sample from $\mathbb{R}^d \times \{1, 2\}$ and let $\{\theta_i\}_{i=1,2,\ldots,n}$ be i.i.d random variables with $P(\theta_i = 1) = \xi_1$ and $P(\theta_i = 2) = \xi_2 \; \forall \; i = 1, \ldots, n$ where $\xi_1 + \xi_2 = 1$.

If $\theta_i = 1$ we say X_i is from π_1 with p.d.f f_1 and if $\theta_i = 2$ we say X_i is from π_2 with p.d.f f_2. Let n_i be the number of X's that have been identified as coming from π_i for $i = 1, 2$. Note that $n_1 + n_2 = n$.

According to the sub-sample procedure described above we consider the set

$$\mathcal{S} = \{\underset{\sim}{Y} = (X_i, X_j)^T; \; \theta_i = \theta_j = \tau_{ij}, i \neq j\}. \tag{5.3}$$

The set defined in (5.3) has $(n_{1_{P_2}} + n_{2_{P_2}})$ elements. Set $N = (n_{1_{P_2}} + n_{2_{P_2}})$. We note that there are $\ell_2(N) = [\frac{n_1}{2}] + [\frac{n_2}{2}]$ mutually independent pairs of random variables in the set defined by (5.3).

Let $\{\tau_m\}$ be i.i.d random variables with $P(\tau_m = 1) = \xi_1$ and $P(\tau_m = 2) = \xi_2$ and

$\xi_1 + \xi_2 = 1$ so that we can consider $\{\underset{\sim m}{Y}, \tau_m\}_{m=1,2,\ldots,N}$ as identically distributed training samples of which $\ell_2(N)$ of them are mutually independent.

Let (Z_1, τ) and (Z_2, τ) be the new observations to be classified. Denote $\{(Z_1, \tau), (Z_2, \tau)\} = (\underset{\sim}{Z}, \tau)$. Suppose $\underset{\sim N}{Y'}$ is the Nearest Neighbor of $\underset{\sim}{Z}$ and $\underset{\sim N}{Y'}$ belongs to the category τ'_N. Then we classify $\underset{\sim}{Z}$ into the category of τ'_N.

Now we give, the proof of almost sure convergence of Nearest Neighbor $\underset{\sim n}{Y'}$ to $\underset{\sim}{Z}$ in the following lemma.

LEMMA 5.1.1. *Suppose $\underset{\sim}{Z}$ and $\underset{\sim 1}{Y}, \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}$ are identically distributed random vectors. Also suppose that there are $\ell_2(N)$ mutually independent random vectors present among $\underset{\sim 1}{Y}, \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}$, where $\ell_2(N) \to \infty$ as $N \to \infty$. If $\underset{\sim N}{Y'}$ be the Nearest Neighbor of $\underset{\sim}{Z}$ from the set $\{\underset{\sim 1}{Y} \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}\}$, then*

$$\underset{\sim N}{Y'} \longrightarrow \underset{\sim}{Z} \text{ a.s.} \quad \text{as } n \to \infty. \tag{5.4}$$

PROOF. For any $\varepsilon > 0$

$$\begin{aligned}
&P\{\| \underset{\sim N}{Y'} - \underset{\sim}{Z} \| > \varepsilon\} = P\{\min_m \| \underset{\sim m}{Y} - \underset{\sim}{Z} \| > \varepsilon\} \\
&= P\{\| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| > \varepsilon, \| \underset{\sim 2}{Y} - \underset{\sim}{Z} \| > \varepsilon, \ldots, \| \underset{\sim N}{Y} - \underset{\sim}{Z} \| > \varepsilon\} \\
&\leq \{P \| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| > \varepsilon\}^{\ell_2(N)} \\
&= \{1 - P(\| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| \leq \varepsilon)\}^{\ell_2(N)} \\
&\longrightarrow 0 \text{ as } n \to \infty.
\end{aligned} \tag{5.5}$$

Since $\| \underset{\sim N}{Y'} - \underset{\sim}{Z} \|$ is monotonically decreasing in N, similar arguments as for (2.18) together with (5.5) imply $\underset{\sim N}{Y'} \to \underset{\sim}{Z}$ a.s. as $n \to \infty$. □

Now since τ is the true category of $\underset{\sim}{Z}$, taking the loss due to misclassification using the NN rule to be of the $(0, 1)$ type, i.e.,

$$\begin{aligned}
L(\tau, \tau'_N) &= 0 \text{ if } \tau = \tau'_N \\
&= 1 \text{ if } \tau \neq \tau_N,
\end{aligned} \tag{5.6}$$

the risk $R_{(2)}(2)$ (in general $R_{(s)}(K)$) of δ_s^* is given by

$$R_{(2)}(2) = \lim_{n\to\infty} EL(\tau, \tau'_N). \tag{5.7}$$

Denote by $R^*_{(s)}(K)$ = the Bayes risk for classifying $\underset{\sim}{Z} = (Z_1, \dots, Z_K)$ into one of s populations. Then $R^*_{(s)}(K) = EL(\tau, \tau'_N) = E[r^*(\underset{\sim}{z})]$ with $r^*(\underset{\sim}{z}) = \min_{1\le j\le s}(1 - \eta_j^*(\underset{\sim}{z}))$ and

$$\begin{aligned} \eta_j^*(\underset{\sim}{z}) &= P(\tau = j / \underset{\sim}{Z} = \underset{\sim}{z}) \\ &= \frac{\xi_j f_j(z_1)\dots f_j(z_K)}{\sum_{j=1}^{s} \xi_j f_j(z_1)\dots f_j(z_K)} \quad \text{for } j = 1, 2, \dots, s; \end{aligned} \tag{5.8}$$

$r^*(\underset{\sim}{z})$ is called the conditional Bayes risk given $\underset{\sim}{Z} = \underset{\sim}{z}$. We can now state in the special case $K = 2, s = 2$ the following theorem.

THEOREM 5.1.1. *Let z_1 and z_2 be the continuity points of both f_1 and f_2 and $f_i(z_j) > 0$ for $i = 1, 2; j = 1, 2$. Then the NN risk $R_{(2)}(2)$ satisfies the inequalities*

$$R^*_{(2)}(2) \le R_{(2)}(2) \le 2R^*_{(2)}(2)(1 - R^*_{(2)}(2)),$$

the bound on the right being the best possible.

PROOF. The conditional NN risk given $\underset{\sim}{Z} = \underset{\sim}{z}$ and $\underset{\sim}{Y}'_N$ is given by

$$\begin{aligned} r(\underset{\sim}{z}, \underset{\sim}{Y}'_N) &= E(L(\tau, \tau'_N) / \underset{\sim}{z}, \underset{\sim}{Y}'_N) \\ &= P(\tau \ne \tau'_N / \underset{\sim}{z}, \underset{\sim}{Y}'_N) \\ &= P(\tau = 2, \tau'_N = 1 / \underset{\sim}{z}, \underset{\sim}{Y}'_N) + P(\tau = 1, \tau'_N = 2 / \underset{\sim}{z}, \underset{\sim}{Y}'_N) \\ &= P(\tau = 2 / \underset{\sim}{z}) \cdot P(\tau'_N = 1 / \underset{\sim}{Y}'_N) + P(\tau = 1 / \underset{\sim}{z}) \cdot P(\tau'_N = 2 / \underset{\sim}{Y}'_N), \end{aligned}$$

where in the last equality we have used the conditional independence of τ and τ'_N. By (5.8), we thus obtain

$$r(\underset{\sim}{z}, \underset{\sim}{Y}'_N) = \eta_2^*(\underset{\sim}{z}) \cdot \eta_1^*(\underset{\sim}{Y}'_N) + \eta_1^*(\underset{\sim}{z}) \cdot \eta_2^*(\underset{\sim}{Y}'_N). \tag{5.9}$$

Since z_1 and z_2 are continuity points of both f_1 and f_2, $\underset{\sim}{z} = (z_1, z_2)$ is a continuity point of η^*. Now since $\underset{\sim}{Y}'_N \to \underset{\sim}{z}$ a.s. as $n \to \infty$ by lemma 5.1.1., we have

$$\eta_1^*(\underset{\sim}{Y}'_N) \to \eta_1^*(\underset{\sim}{z}) \text{ a.s.} \tag{5.10}$$

$$\eta_2^*(\underset{\sim}{Y}'_N) \to \eta_2^*(\underset{\sim}{z}) \text{ a.s. .} \tag{5.11}$$

By (5.9), (5.10) and (5.11) we have

$$r(\underset{\sim}{z}, \underset{\sim}{Y}'_N) \to r(\underset{\sim}{z}) = 2\eta_1^*(\underset{\sim}{z})\eta_2^*(\underset{\sim}{z}) \text{ a.s. ;} \tag{5.12}$$

in fact, we can express the conditional, given $\underset{\sim}{Z} = \underset{\sim}{z}$, risk of δ_s^* as

$$\begin{aligned} r(\underset{\sim}{z}) &= 2\eta_1^*(\underset{\sim}{z})\eta_2^*(\underset{\sim}{z}) = 2\eta_1^*(\underset{\sim}{z})(1 - \eta_1^*(\underset{\sim}{z})) \\ &= 2r^*(\underset{\sim}{z})(1 - r^*(\underset{\sim}{z})), \end{aligned} \tag{5.13}$$

the last equality following in view of the fact that $r^*(\underset{\sim}{z})$ is symmetric in $\eta_1^*(\underset{\sim}{z})$'s. Further, since $r(\underset{\sim}{z}, \underset{\sim}{Y}'_N)$ is bounded and by (5.12) $r(\underset{\sim}{z}, \underset{\sim}{Y}'_N) \to r(\underset{\sim}{z})$ a.s., using dominated convergence theorem we have

$$\begin{aligned} R_{(2)}(2) &= \lim_n E\{r(\underset{\sim}{z}, \underset{\sim}{Y}'_N)\} \\ &= E\{\lim r(\underset{\sim}{z}, \underset{\sim}{Y}'_N)\} \\ &= E(r(\underset{\sim}{z})). \end{aligned} \tag{5.14}$$

From (5.13) and (5.14), we thus have using Jensen's inequality

$$\begin{aligned} R_{(2)}(2) &= Er(\underset{\sim}{z}) = 2E(r^*(\underset{\sim}{z})(1 - r^*(\underset{\sim}{z})) \\ &= 2\{E(r^*(\underset{\sim}{z})) - E(r^{*2}(\underset{\sim}{z}))\}, \\ &\leq 2\{E(r^*(\underset{\sim}{z})) - (Er^*(\underset{\sim}{z}))^2\}, \\ &= 2(R^*_{(2)}(2) - R^{*2}_{(2)}(2)), \\ &= 2R^*_{(2)}(2)(1 - R^*_{(2)}(2)). \end{aligned} \tag{5.15}$$

Again, note that

$$
\begin{aligned}
r^*(\underset{\sim}{z}) &= \min(\eta_1^*(\underset{\sim}{z}), \eta_2^*(\underset{\sim}{z})) \\
&\leq 2\eta_1^*(\underset{\sim}{z})\eta_2^*(\underset{\sim}{z}) = r(\underset{\sim}{z}).
\end{aligned}
\tag{5.16}
$$

Now taking expectations of both sides of (5.16) we get

$$
R^*_{(2)}(2) \leq R_{(2)}(2). \tag{5.17}
$$

Thus, (5.15) and (5.17) yield

$$
R^*_{(2)}(2) \leq R_{(2)}(2) \leq 2R^*_{(2)}(2)(1 - R^*_{(2)}(2)). \tag{5.18}
$$

The proof is complete. □

5.2 Asymptotic NN risk when K multiple observations are to be classified into one among s-populations.

Let $(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n)$ be the identified training sample from $\mathbb{R}^d \times \{1, 2, \ldots, s\}$. If $\theta_i = j$, we say X_i is from the population π_j with p.d.f. f_j, $\forall\, i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, s$.

Let n_j be the number of observations in the training sample identified as being from π_j $j = 1, 2, \ldots, s$ with $\sum_{j=1}^{s} n_j = n$.

Consider again (from (5.1))

$$
\mathcal{S} = \{\underset{\sim}{Y} = (X_{i_1}, X_{i_2}, \ldots, X_{i_n});\ \theta_{i_1} = \theta_{i_2} = \ldots = \theta_{i_K} = \tau_{(i_1, \ldots, i_K)},\ i_1 \neq i_2 \neq \ldots \neq i_K\}
$$

Note that the set $\mathcal{S}$ has $N = \left({}^{n_1}P_K + {}^{n_2}P_K + \ldots + {}^{n_s}P_K\right)$ elements. Let $\{\tau_m\}$ be i.i.d. random variables with $P(\tau_m = i) = \xi_i$ and $\sum_{i=1}^{s} \xi_i = 1$; then $\{\underset{\sim}{Y}_m, \tau_m\}_{m=1,2,\ldots,N}$ may be taken as an identically distributed (not all mutually independent) training samples from $\mathbb{R}^{Kd} \times \{1, 2, \ldots, s\}$ with $\ell_K(N) = [\frac{n_1}{2}] + \ldots + [\frac{n_K}{2}]$ mutually independent variables. Let (Z_1, τ), $(Z_2, \tau), \ldots, (Z_K, \tau)$ be K-observations to be classified. For notational convenience we denote $(\underset{\sim}{Z}, \tau) = (Z_1, Z_2, \ldots, Z_K, \tau)$. Now τ is to be estimated in order to classify $\underset{\sim}{Z}$.

Suppose $\underset{\sim N}{Y'}$ is the Nearest Neighbor of $\underset{\sim}{Z}$ and $\underset{\sim N}{Y'}$ is identified as being from τ'_N. So an estimate of τ is τ'_N. Thus, we classify $\underset{\sim}{Z}$ as from τ'_N. The following Lemma gives almost sure convergence of the Nearest neighbor Y'_N to $\underset{\sim}{Z}$.

LEMMA 5.2.1. *Let $\underset{\sim}{Z}$ and $\underset{\sim 1}{Y}, \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}$ be identically distributed random vectors. We assume $\ell_K(N)$ of the $\underset{\sim m}{Y}$'s are mutually independent where $\ell_K(N) \to \infty$, as $n \to \infty$. Let $\underset{\sim N}{Y'}$ denote the Nearest Neighbor of $\underset{\sim}{Z}$ from the set $\{\underset{\sim 1}{Y}, \underset{\sim 2}{Y}, \ldots, \underset{\sim N}{Y}\}$. Then, as $n \to \infty$,*

$$\underset{\sim N}{Y'} \to \underset{\sim}{Z} \quad \text{a.s.} \; . \tag{5.19}$$

PROOF. For any $\delta > 0$

$$\begin{aligned}
P\{\| \underset{\sim N}{Y'} - \underset{\sim}{Z} \| > \delta\} &= P\{\min_m \| \underset{\sim m}{Y} - \underset{\sim}{Z} \| > \delta\} \\
&= P\{\| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| > \delta, \| \underset{\sim 2}{Y} - \underset{\sim}{Z} \| > \delta, \ldots, \| \underset{\sim N}{Y} - \underset{\sim}{Z} \| > \delta\} \\
&\leq \{P(\| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| > \delta)\}^{\ell_K(N)} \\
&= \{1 - P(\| \underset{\sim 1}{Y} - \underset{\sim}{Z} \| \leq \delta\}^{\ell_K(N)} \\
&\to 0 \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{5.20}$$

where K is finite. The result now follows as in Lemma 5.1.1. The proof is complete. □

Taking our loss function again as:

$$\begin{aligned}
L(i,j) &= 0 \quad \text{if} \quad i = j \\
&= 1 \quad \text{if} \quad i \neq j,
\end{aligned}$$

the asymptotic NN risk, $R_{(s)}(K)$ is by definition (see (5.7) and (2.9))

$$R_{(s)}(K) = \lim_n EL(\tau, \tau'_N) \tag{5.21}$$

and overall Bayes risk $R^*_{(s)}(K)$ is given by

$$R^*_{(s)}(K) = Er^*(\underset{\sim}{z})$$
$$= \int \left[\min(1-\eta_1^*(\underset{\sim}{z}), 1-\eta_2^*(\underset{\sim}{z}), \ldots, 1-\eta_s^*(\underset{\sim}{z}))\sum_{j=1}^{s}\xi_j f(\underset{\sim}{z})\right] d\,\underset{\sim}{z} \quad (5.22)$$

THEOREM 5.2.1. *Let $z_1, z_2, \ldots, z_K$ be the continuity points of each of the densities $f_1, f_2, \ldots,$ and f_s and suppose that $f_i(z_j) > 0$ for $i = 1, \ldots, s$; $j = 1, \ldots, K$. Then the limiting NN risk $R_{(s)}(K)$ has the following bounds:*

$$R^*_{(s)}(K) \le R_{(s)}(K) \le R^*_{(s)}(K)\left(2 - \frac{s}{s-1}\right)R^*_{(s)}(K). \quad (5.23)$$

PROOF. The conditional NN risk given $\underset{\sim}{Z} = \underset{\sim}{z}$ and $\underset{\sim N}{Y'} = \underset{\sim N}{y'}$ is given by

$$r(\underset{\sim}{z}; \underset{\sim N}{y'}) = E\{L(\tau, \tau'_N)/\, \underset{\sim}{z}, \underset{\sim N}{y'}\}$$
$$= P(\tau \ne \tau'_N/\, \underset{\sim}{z}, \underset{\sim N}{y'}). \quad (5.24)$$

Using the conditional independence of τ and τ'_N, we have

$$r(\underset{\sim}{z}, \underset{\sim N}{y'}) = \sum_{i \ne j} P(\tau = i, \tau'_N = j/\, \underset{\sim}{z}, \underset{\sim N}{y'})$$
$$= \sum_{i \ne j} P(\tau = i/\, \underset{\sim}{z}) P(\tau'_N = j/\, \underset{\sim N}{y'}).$$

so that by (5.8)

$$r(\underset{\sim}{z}; \underset{\sim N}{y'}) = \sum_{i \ne j} \eta_i^*(\underset{\sim}{z})\eta_j^*(\underset{\sim N}{y'}). \quad (5.25)$$

Since $\underset{\sim N}{y'} \to \underset{\sim}{z}$ a.s and f_i's $i = 1, \ldots, s$ are continuous, we have

$$\eta_i(\underset{\sim N}{y'}) \to \eta_i(\underset{\sim}{z}) \ a.s. \quad \forall i = 1, 2, \ldots, s. \quad (5.26)$$

Thus

$$r(\underset{\sim}{z}, \underset{\sim N}{y'}) \to r(\underset{\sim}{z}) \quad a.s. \quad (5.27)$$

where for fixed $\underset{\sim}{z}$,

$$r(\underset{\sim}{z}) = \sum_{i \neq j} \eta_i^*(\underset{\sim}{z})\eta_j^*(\underset{\sim}{z})$$
$$= 1 - \sum_{j=1}^{s} \eta_j^{*2}(\underset{\sim}{z}). \tag{5.28}$$

By letting $\eta_\ell^*(\underset{\sim}{z}) = \max_j\{\eta_j^*(\underset{\sim}{z})\}$, the conditional Bayes risk $r^*(\underset{\sim}{z})$ can be written as

$$r^*(\underset{\sim}{z}) = 1 - \max_j\{\eta_j^*(\underset{\sim}{z})\}$$
$$= 1 - \eta_\ell^*(\underset{\sim}{z}), \tag{5.29}$$

so that by the Cauchy-Schwartz inequality

$$(s-1)\sum_{j \neq \ell} \eta_j^{*2}(\underset{\sim}{z}) \geq \left[\sum_{j \neq \ell}^{s} \eta_j^*(\underset{\sim}{z})\right]^2$$
$$= (1 - \eta_\ell^*(\underset{\sim}{z}))^2$$
$$= (r^*(\underset{\sim}{z}))^2. \tag{5.30}$$

Adding $(s-1)\,\eta_\ell^{*2}(\underset{\sim}{z})$ both sides in (5.30), we have

$$\sum_{j=1}^{s} \eta_j^{*2}(\underset{\sim}{z}) \geq \frac{(r^*(\underset{\sim}{z}))^2}{(s-1)} + (1 - r^*(\underset{\sim}{z}))^2, \tag{5.31}$$

so that from (5.28) and (5.31), we get

$$r(\underset{\sim}{z}) \leq 2r^*(\underset{\sim}{z}) - \frac{s}{s-1}(r^*(\underset{\sim}{z}))^2. \tag{5.32}$$

Now by the (Lebesque) Dominated Converge Theorem, (5.32) and Jensen's inequality we have

$$R_{(s)}(K) = \lim_n E(r(\underset{\sim}{z}; \underset{\sim N}{y'}))$$
$$= Er(\underset{\sim}{z})$$
$$\leq 2R_{(s)}^*(K) - \frac{s}{s-1}E(r^*(\underset{\sim}{z}))^2$$
$$\leq 2R_{(s)}^*(K) - \frac{s}{s-1}(Er^*(\underset{\sim}{z}))^2$$
$$= 2R_{(s)}^*(K) - \frac{s}{s-1}R_{(s)}^{*2}(K)$$
$$= R_{(s)}^*(K)(2 - \frac{s}{s-1}R_{(s)}^*(K)). \tag{5.33}$$

Also note that

$$\min\left(\sum_{i\neq 1}\xi_i\prod_{\ell=1}^{K}f_i(z_\ell),\sum_{i\neq 2}\xi_i\prod_{\ell=1}^{K}f_i(z_\ell),\ldots,\sum_{i\neq s}\xi_i\prod_{\ell=1}^{K}f_i(z_\ell)\right)$$
$$\leq\sum_{j=1}^{s}\frac{\xi_i\prod_{\ell=1}^{K}f_i(z_\ell)}{\sum_{i=1}^{s}\xi_i\prod_{\ell=1}^{K}f_i(z_\ell)}\sum_{\substack{i=1\\ i\neq j}}^{s}\xi_i\prod_{\ell=1}^{K}f_i(z_\ell) \tag{5.34}$$

which on integrating both sides w.r.t. to $\underset{\sim}{z}$, and adjusting the integral, yields

$$Er^*(\underset{\sim}{z})\leq E\sum_{i\neq j}\eta_i^*(z)\eta_j^*(\underset{\sim}{z}),$$

$$R^*_{(s)}(K)\leq R_{(s)}(K). \tag{5.35}$$

Thus from (5.33) and (5.35) we get

$$R^*_{(s)}(K)\leq R_{(s)}(K)\leq R^*_{(s)}(K)(2-\frac{s}{s-1}R^*_{(s)}(K)).$$

This completes the proof. □

5.3 Estimation of asymptotic NN risk $R_{(2)}(K)$.

Let $(\underset{\sim 1}{Y},\tau_1),(\underset{\sim 2}{Y},\tau_2),\ldots,(\underset{\sim N}{Y},\tau_N)$ be identically distributed random vectors from $\mathbb{R}^{Kd}\times\{1,2\}$ where $\{\tau_m\}_{m=1,\ldots,N}$ is sequence of i.i.d. random variables with $P(\tau_m=1)=\xi_1, P(\tau_m=2)=\xi_2$, $\xi_1+\xi_2=1$ and $N=n_{1p_K}+n_{2p_K}$. Now delete $(\underset{\sim j}{Y},\tau_j)$ from $(\underset{\sim 1}{Y},\tau_1),(\underset{\sim 2}{Y},\tau_2),\ldots,(\underset{\sim N}{Y},\tau_N)$, and let τ'_{jN} be the NN estimate of τ_j for $1\leq j\leq N$.

Let

$$\widehat{p}_N=\frac{1}{N}\sum_{j=1}^{N}I_{(\tau'_{jN}\neq\tau_j)} \tag{5.36}$$

and

$$U_N=\frac{1}{N}\sum_{j=1}^{N}E\{I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim j}{Y},\tau_j)\} \tag{5.37}$$

where

$$I_{(\tau'_{jN}\neq\tau_j)}=\begin{cases}1 & \text{if } \tau'_{jN}\neq\tau_j\\ 0 & \text{if } \tau_{jN}=\tau_j.\end{cases} \tag{5.38}$$

Note that

$$E(\widehat{p}_N) = E(U_N) \tag{5.39}$$

and by (5.7)

$$E(\widehat{p}_N) \to R_{(2)}(K) \text{ as } n \to \infty. \tag{5.40}$$

So $\widehat{p}_N$ is asymptotically unbiased for $R_{(2)}(K)$. We also note that U_N defined by (5.37) is a U-statistic and hence by Theorem A of Serfling (1980, p. 190)

$$U_N \xrightarrow{a.s.} R_{(2)}(K) \text{ as } \quad n \to \infty. \tag{5.41}$$

Now we will show $E(\widehat{p}_N - U_N)^2 \to 0$ as $n \to \infty$ which, in view of (5.41) will imply the consistency of $\widehat{p}_N$.

LEMMA 5.3.1. *$E(\widehat{p}_N - U_N)^2 = O(\frac{1}{N})$ where $\widehat{p}_N$ and U_N are defined in (5.36) and (5.37) respectively.*

PROOF. Using (5.36) and (5.37), we write

$$\widehat{p}_N - U_N = \frac{1}{N}\sum_{j=1}^{N}\left\{I_{(\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\right\}. \tag{5.42}$$

Thus

$$\begin{aligned}E(\widehat{p}_N - U_N)^2 &= \frac{1}{N^2}\sum_{j=1}^{N} E\{I_{\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}^2 \\ &+ \frac{1}{N^2}\sum_{i\neq j}\sum E\{I_{(\tau'_{iN}\neq\tau_i)} - E(I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i))\} \\ &\times \{I_{(\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}.\end{aligned} \tag{5.43}$$

Now

$$\begin{aligned}&E\{I_{(\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}^2 \\ &\quad = E var(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j)) \\ &\quad \leq 1/4\,,\end{aligned} \tag{5.44}$$

and

$$E\{I_{(\tau'_{iN}\neq\tau_i)} - E(I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i))\}$$
$$\times\{I_{(\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}$$
$$= E[E\{I_{(\tau'_{iN}\neq\tau_i)} - E(I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i))\}$$
$$\times\{I_{(\tau'_{jN}\neq\tau_j)} - E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}/(\underset{\sim}{Y}_i,\tau_i),(\underset{\sim}{Y}_j,\tau_j)]$$
$$= E\{E(I_{(\tau'_{iN}\neq\tau_i)}I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_i,\tau_i),(\underset{\sim}{Y}_j,\tau_j))$$
$$- E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))E(I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_j))$$
$$- E(I_{\tau'_{iN}\neq\tau_i})/(\underset{\sim}{Y}_i,\tau_j))E(I_{\tau'_{jN}\neq\tau_j})/(\underset{\sim}{Y}_j,\tau_j))$$
$$+ E(I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i))E(I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\},$$

using conditional independence of $I_{(\tau'_{in}\neq\tau_i)}$ and $I_{(\tau'_{jn}\neq\tau_j)}$, the last expression equals

$$E\{E\{I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i)I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))$$
$$- E\{I_{(\tau'_{iN}\neq\tau_i)}/(\underset{\sim}{Y}_i,\tau_i)I_{(\tau'_{jN}\neq\tau_j)}/(\underset{\sim}{Y}_j,\tau_j))\}$$
$$= 0. \tag{5.45}$$

Thus, using (5.43) - (5.45) we have, as $n \to \infty$,

$$E(\widehat{p}_N - U_N)^2 = O(\frac{1}{N}). \tag{5.46}$$

The proof is complete. □

COROLLARY 5.3.1. *Under the conditions of Theorem 5.2.1 the estimate $\widehat{p}_N$ defined by (5.36) is a consistent estimate of $R_{(2)}(K)$.*

REMARKS 5.3.1. The estimate of (5.36) is known as cross-validation estimate (see Breiman et al (1984), p. 11-12). The estimate $\widehat{p}_N$ and the consistency results of this section can be generalized to the case of more than two populations.

CHAPTER 6
SOME EMPIRICAL RESULTS USING MONTE CARLO SIMULATION

6.0 **Introduction.** As we have noted in earlier chapters the theoretical properties of the classification rules tend to be quite complex. We have studied some asymptotic properties of proposed classification rules, but these properties are not necessarily true for small samples. The present chapter is concerned with the performance of some of the proposed procedures when the sample sizes are small. In any Monte Carlo simulation study, it is only possible to deal with special cases, but one may still cover a wide range of interesting situations. In these simulation studies we consider classification of one or more observations among two populations.

6.1 Notation.

$N(\mu, \sigma^2) =$ Normal distribution with mean μ and variance σ^2;

the p.d.f. is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad \sigma > 0, -\infty < \mu < \infty.$$

$LA(a, b) =$ Laplace distribution with location and scale parameters

a and b respectively;

the p.d.f. is given by

$$f(x) = \frac{1}{2b} e^{-|x-a|/b}, \quad b > 0.$$

$LG(a, b) =$ Logistic distribution with location and scale parameters

a and b respectively;

the p.d.f. is given by

$$f(x) = \frac{e^{-(x-a)/b}}{b[1 + e^{-(x-a)/b}]^2}, \quad b > 0, -\infty < a < \infty.$$

$EX(a,b) =$ Exponential distribution with location and scale parameters a and b respectively;

the p.d.f is given by

$$f(x) = \frac{1}{b}e^{-(x-a)/b}, \quad x > a,\ b > 0.$$

$GA(a,b) =$ Gamma distribution with parameters a and b;

the p.d.f. is given by

$$f(x) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}, \quad x > 0,\ a > 0,\ b > 0.$$

$WE(a,b) =$ Weibull distribution with parameters a and b;

the p.d.f is given by

$$f(x) = \frac{ax^{a-1}e^{-(x/b)^a}}{b^a}, \quad x > 0,\ a > 0,\ b > 0.$$

6.2 A small sample comparison between the Cover and Hart Procedure and the Das Gupta and Lin procedure.

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ be a random sample from the population π_1 and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be a random sample from the population π_2. Let Z be an observation known to be either from π_1 or from π_2, the problem is to decide which.

According to Cover and Hart NN rule, using distance function d, X_{1i}'s and X_{2j}'s are ranked as per their distance from Z; then classify Z to π_1 if it's Nearest Neighbor is from π_1, otherwise classify Z to π_2.

Das Gupta and Lin's RNN rule proceeds as follows: X_{1i}'s, X_{2j}'s and Z are ranked in the increasing order of magnitude;

(i) if Z is either smallest or largest then classify Z into it's Nearest Rank Neighbor (RNN);

(ii) if both left and right RNN of Z belong to the same class, classify Z to that class;

(iii) if both left and right RNN belong to different classes, classify Z into either of the two classes with probability 1/2 and 1/2.

For the comparison of the above mentioned rules, random samples of equal size ($n = n_1 = n_2$) were generated from a pair of univariate distributions, namely, normal vs normal, Laplace vs Laplace, normal vs Laplace distributoins etc. For given X_{1i}'s and X_{2j}'s, each procedure was based on samples of sizes $n = 10, 20, 30, 40$ and 50. A random sample of size 100 from π_1 and another random sample of the same size from π_2 are classified by using the two classification rules. The proportion among 200 Z's that were misclassified given X_{1i}'s and X_{2j}'s are given in TABLE 6.1, TABLE 6.2 and TABLE 6.3; thus these proportions are the conditional probabilities of misclassification given the training samples X_{1i}'s and X_{2j}'s.

Table 6.1

Comparison of Cover & Hart (I) vs Das Gupta & Lin (II) procedures in the case of normal distributions, given training samples.

	$N(0,1)$ vs $N(1,1)$		$N(0,1)$ vs $N(2,1)$		$N(0,1)$ vs $N(3,1)$	
n	I	II	I	II	I	II
10	0.400	0.4350	0.2450	0.2625	0.2500	0.2500
20	0.3500	0.3500	0.2100	0.2350	0.1900	0.2025
30	0.3000	0.3275	0.2500	0.2600	0.2700	0.2475
40	0.3050	0.3375	0.2400	0.2300	0.2200	0.2425
50	0.3400	0.3350	0.2300	0.2300	0.2300	0.2250

Table 6.2

Comparison of Cover & Hart vs Das Gupta & Lin procedures in the case of Laplace distributions, given training samples

	$LA(0,1)$ vs $LA(1,1)$		$LA(0,1)$ vs $LA(2,1)$		$LA(0,1)$ vs $LA(3,1)$	
n	I	II	I	II	I	II
10	0.3650	0.4000	0.2750	0.2900	0.2150	0.2200
20	0.3775	0.3700	0.2100	0.2325	0.1900	0.2000
30	0.3950	0.3625	0.2850	0.2950	0.1550	0.1575
40	0.3750	0.3675	0.2950	0.2925	0.1600	0.1600
50	0.3750	0.3825	0.2800	0.2725	0.1900	0.1750

Table 6.3

Comparison of Cover & Hart vs Das Gupta & Lin procedures for the pair Normal vs Laplace and Normal vs Logistic distributions, given training samples

	$N(0,1)$ vs $LA(2,1)$		$N(0,1)$ vs $LG(2,1)$		$LA(0,1)$ vs $LA(2,1)$	
n	I	II	I	II	I	II
10	0.2800	0.2800	0.2250	0.2700	0.2750	0.2900
20	0.2350	0.2325	0.3300	0.3200	0.2100	0.2325
30	0.2350	0.2600	0.2600	0.2475	0.2850	0.2950
40	0.2750	0.2625	0.2900	0.2750	0.2950	0.2925
50	0.2100	0.2175	0.3050	0.3075	0.2800	0.2725

n = Sample size taken from each population.

I = Average proportion of Misclassification for Cover & Hart procedure, given X_{1i}'s and X_{2j}'s.

II = Average proportion of Misclassification for Das Gupta & Lin procedure, given X_{1i}'s and X_{2j}'s.

From Table 6.1, Table 6.2 and Table 6.3 it is seen that, the Cover and Hart rule performed uniformly better than Das Gupta and Lin procedure in the case of small samples $n = 10, 20$. The two procedures tend to perform equally or fairly close to each other for $n = 50$. Only in few cases for the pair Normal vs

Laplace, Normal vs Logistic, Laplace vs Laplace Das Gupta and Lin procedure performed better than Cover and Hart procedure in case of moderate sample sizes $n = 30, 40$. However as has been established, both the rules are asymptoticallly equivalent (have the same asymptotic probability of misclassification).

Now we employ another method to compare the performance of the Cover and Hart and Das Gupta and Lin rules. For given Z (the observations to be classified) 100 samples from π_1 and 100 samples from π_2 of sizes $n = 10, 20, 30, 40$ and 50 are generated in each case. A given Z (one from π_1 and another from π_2) is classified each time. The proportion of times that Z is misclassified among 200 samples is noted. The results are given in TABLE 6.4, TABLE 6.5 and TABLE 6.6.

Table 6.4

Comparison of Cover & Hart vs Das Gupta & Lin procedures in the case of Normal distributions, given Z.

	$N(0,1)$ vs $N(1,1)$		$N(0,1)$ vs $N(2,1)$		$N(0,1)$ vs $N(3,1)$	
n	I	II	I	II	I	II
10	0.3450	0.3375	0.2450	0.2350	0.1100	0.1225
20	0.3200	0.2775	0.3000	0.2825	0.0800	0.0950
30	0.3600	0.3350	0.2400	0.2475	0.0700	0.0800
40	0.3050	0.2725	0.2750	0.2925	0.1000	0.0950
50	0.1550	0.1500	0.2750	0.2900	0.0850	0.0925

Table 6.5

Comparison of Cover & Hart vs Das Gupta & Lin procedures in the case of Laplace distributions, given Z.

	$LA(0,1)$ vs $LA(1,1)$		$LA(0,1)$ vs $LA(2,1)$		$LA(0,1)$ vs $LA(3,1)$	
n	I	II	I	II	I	II
10	0.3150	0.2900	0.1100	0.1775	0.0400	0.0575
20	0.2200	0.2475	0.0950	0.1325	0.0300	0.0550
30	0.2800	0.2775	0.1300	0.1250	0.0650	0.0650
40	0.2250	0.2525	0.1125	0.1100	0.0450	0.0425
50	0.2750	0.2700	0.1000	0.1150	0.0700	0.0775

Table 6.6

Comparison of Cover & Hart vs Das Gupta & Lin procedures in the case of Normal vs Laplace & Logistic distributions, given Z.

	$N(0,1)$ vs $LA(2,1)$		$N(0,1)$ vs $LA(3,1)$		$N(0,1)$ vs $LG(0,1)$	
n	I	II	I	II	I	II
10	0.0800	0.0600	0.0400	0.0600	0.4600	0.4300
20	0.0850	0.0875	0.0300	0.0425	0.4450	0.4650
30	0.0750	0.0600	0.0450	0.0500	0.4000	0.4175
40	0.0750	0.0775	0.0200	0.0350	0.4250	0.4475
50	0.0450	0.0525	0.0600	0.0600	0.4600	0.4575

Here we notice two different trends in the average proportion of misclassification. The Das Gupta and Lin procedure performs better when the two distributions are close to each other; but when they are apart Cover and Hart's procedure seemed better.

6.3 Performance of first stage RNN rule in small samples.

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be independent random samples from two populations π_1 and π_2. Another random sample $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ is taken from one of these two populations and the problem is to classify $\underset{\sim}{Z}$ into π_1 or π_2.

We may describe our 1st stage RNN rule as follows:

Combine X_{1i}'s , X_{2j}'s and $\underset{\sim}{Z}$'s and arrange them in increasing order; the left and right hand neighbors of $Z_\ell, \forall\ \ell = 1, 2, \ldots, K$ are identified. These are called Rank Nearest Neighbors (RNN's). Then, classify $\underset{\sim}{Z}$ to the population π_1

$$\text{if } \# \text{ RNN from } \pi_1 > \# \text{ RNN from } \pi_2;$$

classify $\underset{\sim}{Z}$ to the population π_2 with probability 1/2

$$\text{if } \# \text{ RNN from } \pi_1 = \# \text{ RNN from } \pi_2.$$

The object of this section is to study the behavior of average proportion of misclassification $R^{(1)}_{(2)}(K)$ as K varies. In our simulation study we vary K from 1 to 5. Given X_{1i}'s and X_{2j}'s each procedure is based on equal sample sizes of 50. Now 1st stage RNN rule is used to classify 200 random sample $\underset{\sim}{Z} = (Z_1, \ldots, Z_K)$, 100 form each population. The number of these random samples $\underset{\sim}{Z}$ that are misclassified is recorded. The average proportion of misclassification given X_{1i}'s and X_{2j}'s for different pair of distributions, are given in TABLE 6.7, TABLE 6.8 and TABLE 6.9.

Table 6.7

The values of $R^{(1)}_{(2)}(K)$ for the different pair of normal distributions and the pair Normal vs Laplace, given training samples.

	$N(0,1)$ vs $N(1,1)$	$N(0,1)$ vs $N(2,1)$	$N(0,1)$ vs $LA(2,1)$
K	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$
1	0.3875	0.2325	0.2600
2	0.3325	0.1555	0.2150
3	0.3025	0.1225	0.1725
4	0.2175	0.0850	0.1550
5	0.2175	0.0375	0.1375

Table 6.8

The values of $R^{(1)}_{(2)}(K)$ for the pair of Laplace vs Laplace, Logistic vs Logistic and Normal vs Logistic, given training samples.

	$LA(0,1)$ vs $LA(1,1)$	$LG(0,1)$ vs $LG(2,1)$	$N(0,1)$ vs $LG(2,1)$
K	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$
1	0.3075	0.3750	0.4800
2	0.2750	0.3375	0.4075
3	0.1925	0.2875	0.3825
4	0.1825	0.2675	0.3500
5	0.1550	0.2075	0.3425

Table 6.9

The values of $R^{(1)}_{(2)}(K)$ for the pair of Exponential vs Gamma, Exponential vs Weibull and,Weibull vs Gamma, given training samples.

	$EX(0,1)$ vs $GA(2,1)$	$EX(0,1)$ vs $WE(2,1)$	$WE(2,1)$ vs $GA(3,1)$
K	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$
1	0.2950	0.4800	0.2025
2	0.2050	0.4425	0.1675
3	0.1575	0.4375	0.0850
4	0.1050	0.4100	0.0575
5	0.0900	0.3975	0.0500

We notice that for all the distributions the average proportion of misclassification $R^{(1)}_{(2)}(K)$ decreases as the value of K increases; this is a desirable property of $R^{(1)}_{(2)}(K)$.

Now we want to see whether the same property holds for $R^{(1)}_{(2)}(K)$ given $\underset{\sim}{Z} = (Z_1, \ldots, Z_K)$. Here we start with $K = 1$ then increase K one by one keeping those old Z's fixed. We have simulated 100 random samples from π_1 and 100 random samples from π_2 to classify a given pair of random sample $\underset{\sim}{Z}$ (one from each population) for each value of K. First stage RNN rule is applied to decide the class of $\underset{\sim}{Z}$ and the number of misclassifications are recorded. The average proportion of misclassification $R^{(1)}_{(2)}(K)$'s are given in the TABLE 6.10, TABLE 6.11 and TABLE 6.12 for various pairs of distributions.

Table 6.10

The values of $R^{(1)}_{(2)}(K)$ for the different pair of normal distributions and, the pair Normal vs Laplace, given $\underset{\sim}{Z}$

	$N(0,1)$ vs $N(1,1)$	$N(0,1)$ vs $N(2,1)$	$N(0,1)$ vs $LA(2,1)$
K	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$
1	0.4400	0.3175	0.1225
2	0.2800	0.1925	0.0950
3	0.2175	0.1925	0.0900
4	0.1675	0.1300	0.0775
5	0.1675	0.1200	0.0400

Table 6.11

The values of $R^{(1)}_{(2)}(K)$ for the pair of Laplace vs Laplace, Logistic vs Logistic and Normal vs Logistic, given $\underset{\sim}{Z}$.

	$LA(0,1)$ vs $LA(2,1)$	$LG(0,1)$ vs $LG(2,1)$	$N(0,1)$ vs $LG(0,1)$
K	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$	$R^{(1)}_{(2)}(K)$
1	0.3425	0.4275	0.4200
2	0.2050	0.3950	0.4200
3	0.1900	0.2250	0.3675
4	0.1250	0.1475	0.3400
5	0.1250	0.1450	0.2800

Table 6.12

The values of $R^{(1)}_{(2)}(K)$ for the pair of Exponential vs Gamma, Exponential vs Weibull and Weibull vs Gamma, given $\underset{\sim}{Z}$

K	$EX(0,1)$ vs $GA(3,1)$ $R^{(1)}_{(2)}(K)$	$EX(0,1)$ vs $WE(2,1)$ $R^{(1)}_{(2)}(K)$	$WE(2,1)$ vs $GA(3,1)$ $R^{(1)}_{(2)}(K)$
1	0.4900	0.1925	0.0875
2	0.3350	0.1850	0.0725
3	0.1275	0.2425	0.0200
4	0.0800	0.1600	0.0150
5	0.0525	0.0100	0.0000

Similar property seems to hold for $R^{(1)}_{(2)}(K)$ (i.e. $R^{(1)}_{(2)}(K)$ decreases as K increases) here also as expected.

6.4 Sub-sample procedure and comparison with first stage RNN rule in the small samples

Let $X_{11}, X_{12} \ldots X_{1n_1}$ and $X_{21}, X_{22}, \ldots X_{2n_2}$ be random samples from π_1 and π_2 respectively. A random sample $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_K)$ is to be classified into π_1 or π_2.

The sub-sample procedure may be described as follows:

First find all possible permutations taking K observations at a time from the 1st-sample and 2nd-samples. The set of subsamples of size K in the 1st-sample contains ${}^{n_1}P_K$ elements and 2nd-sample contains ${}^{n_2}P_K$ elements. Using the usual Euclidean distance function d, rank the distances of all the $N = ({}^{n_1}P_K + {}^{n_2}P_K)$ elements in the combined sub-sample from $\underset{\sim}{Z}$. Classify $\underset{\sim}{Z}$ into the class to which the nearest neighbor of $\underset{\sim}{Z}$ belongs.

First we study the behavior of average proportion of misclassification $R_{(2)}(K)$ when K increases. In our study we took $n = n_1 = n_2 = 10$ and $K = 1, 2,$ or 3. Suppose X_{1i}'s and X_{2j}'s are given. We simulate a random sample $\underset{\sim}{Z} = (Z_1, \ldots, Z_K)$ ($K = 1, 2, 3$), 100 from each population, and classify them using

sub-sample procedure. The number of times the observed value of random sample vector $\underset{\sim}{Z}$ is misclassified is recorded. Given X_{1i}'s and X_{2j}'s the average proportion of misclassification for different pairs of distributions and for different values of K are given in TABLE 6.13, TABLE 6.14 and TABLE 6.15.

Table 6.13

The values of $R_{(2)}(K)$ for the different pairs of normal distributions for the sub-sample procedure, given training samples.

	$N(0,1)$ vs $N(1,1)$	$N(0,1)$ vs $N(2,1)$	$N(0,1)$ vs $N(3,1)$
K	$R_{(2)}(K)$	$R_{(2)}(K)$	$R_{(2)}(K)$
1	0.3700	0.1750	0.1000
2	0.3400	0.1250	0.0250
3	0.2250	0.0700	0.0200

Table 6.14

The values of $R_{(2)}(K)$ for the different pair of Laplace distributions and a pair of Logistic distributions for the sub-sample procedure, given training samples.

	$LA(0,1)$ vs $LA(1,1)$	$LA(0,1)$ vs $LA(2,1)$	$LG(0,1)$ vs $LG(2,1)$
K	$R_{(2)}(K)$	$R_{(2)}(K)$	$R_{(2)}(K)$
1	0.4550	0.3100	0.4250
2	0.4200	0.2900	0.2750
3	0.4100	0.2450	0.2150

Table 6.15

The values of $R_{(2)}(K)$ for the pair Normal vs Laplace , Normal vs Logistic for the sub-sample procedure, given training samples.

	$N(0,1)$ vs $LA(1,1)$	$N(0,1)$ vs $LA(2,1)$	$N(0,1)$ vs $LG(1,1)$
K	$R_{(2)}(K)$	$R_{(2)}(K)$	$R_{(2)}(K)$
1	0.4400	0.3900	0.4200
2	0.4050	0.3250	0.3650
3	0.3150	0.2300	0.3500

From the above table we observe that average proportion of misclassification $R_{(2)}(K)$ decreases as the value of K increases to 3. This is an expected result.

Finally, we want to compare sub-sample procedure with 1st stage RNN rule which has been described earlier (section 6.3). For simplicity, we take $n = n_1 = n_2 = 10$ and vary K from 1 to 3. For given X_{1i}'s and X_{2j}' s we classify 200 random sample $\underset{\sim}{Z} = (Z_1, \ldots, Z_K)$, 100 from each population using sub-sample procedure and 1st stage RNN rule. The average proportion of misclassifications $R_{(2)}(K)$ were calculated for both the sub-sample procedure and the 1st stage RNN rule. The results are given in TABLE 6.16, TABLE 6.17 and TABLE 6.18.

Table 6.16

Comparison of sub-sample procedure and 1st stage RNN rule for the different pair of Normal distributions, given training samples.

	$N(0,1)$ vs $N(1,1)$		$N(0,1)$ vs $N(2,1)$		$N(0,1)$ vs $N(3,1)$	
K	I	II	I	II	I	II
1	0.3300	0.3500	0.3500	0.3600	0.1000	0.1175
2	0.2600	0.3375	0.2450	0.2975	0.0250	0.0525
3	0.2650	0.3000	0.1500	0.2600	0.0200	0.0400

Table 6.17

Comparison of sub-sample procedure and 1st stage RNN rule for the different pairs of Laplace distributions, and a pair of Loigistic distributions, given training samples.

	$LA(0,1)$ vs $LA(1,1)$		$LA(0,1)$ vs $LA(2,1)$		$LG(0,1)$ vs $LG(2,1)$	
K	I	II	I	II	I	II
1	0.3600	0.3900	0.2350	0.2550	0.3600	0.3625
2	0.2700	0.2800	0.1700	<u>0.1500</u>	0.3500	<u>0.3300</u>
3	0.2600	0.2950	0.1750	<u>0.1350</u>	0.2600	0.3125

Table 6.18

Comparison of sub-sample procedure and 1st stage RNN rule for the pair of Normal vs Laplace and Normal vs Logistic distributions, given training samples.

	$N(0,1)$ vs $LA(1,1)$		$N(0,1)$ vs $LA(2,1)$		$N(0,1)$ vs $LG(2,1)$	
K	I	II	I	II	I	II
1	0.4150	0.4450	0.1700	0.1950	0.3450	0.3500
2	0.2850	0.3425	0.1750	0.1875	0.2400	<u>0.2375</u>
3	0.3350	0.3600	0.0100	0.1025	0.1600	0.2225

I = Average proportion of misclassification of sub-sample procedure

II = Average proportion of misclassification of 1st stage RNN rule.

From the TABLE 6.16 and TABLE 6.18 we Notice that sub-sample procedure perform uniformly better than 1st stage RNN rule with few exceptions (underlined in the tables).

6.5 Concluding Remarks.

It is to see that in all cases the average proportion of misclassifications decreases as the distance between the distributions increases. Diagrams are shown for each pair of distributions at the end of this section to visualize the intersection part of each pair of distributions and accordingly compare the average proportion of misclassifications. In section 6.2, there seems to be no consistent tendency in average proportion of misclassifications for either procedure as the sample size increases. In section 6.3, we notice the remarkable pattern in $R^{(1)}_{(2)}(K)$ which decreases as the value of K increases. We observe a similar pattern for $R_{(2)}(K)$ in section 6.4 where the sub-sample procedure performed better than the 1st stage RNN rule in most of the situations. If one has traning samples of small sizes and a very small sample to be classified, then the sub-sample procedure would be advisable, as its performance seems to be better. In moderately large or large sample cases, the subsample approach involves lot of computations and is quite tedious; the 1st stage RNN rule should be used in this case as it is simple to employ and performs very well in moderately large or large samples.

The results of our simulation study are, of course, not conclusive but at least they throw some light on the relative performance of different classification rules (those existing already in the literature and those proposed in the thesis).

FIG.1 : N(0,1) AND N(1,1)

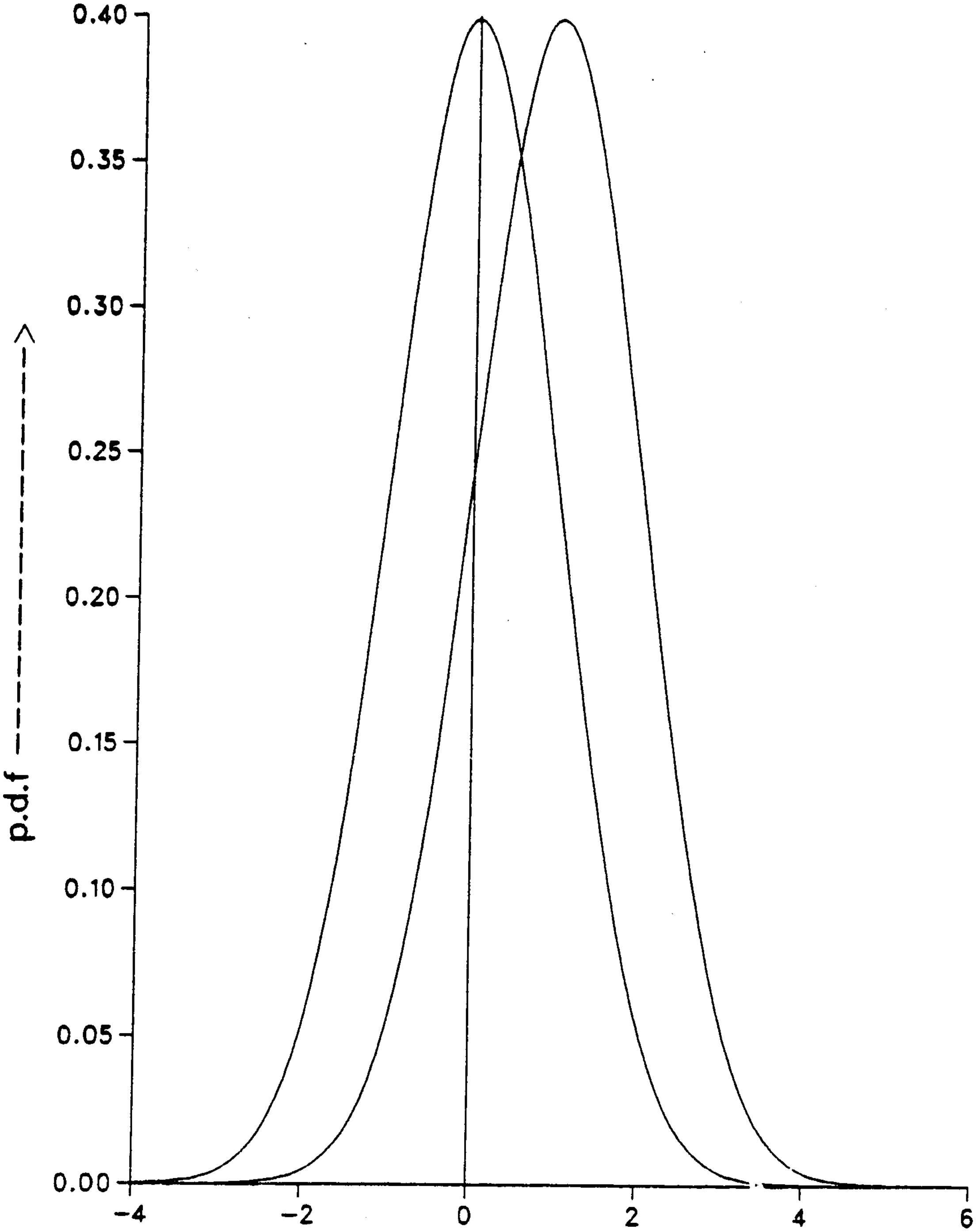


FIG.2 : N(0,1) AND N(2,1)

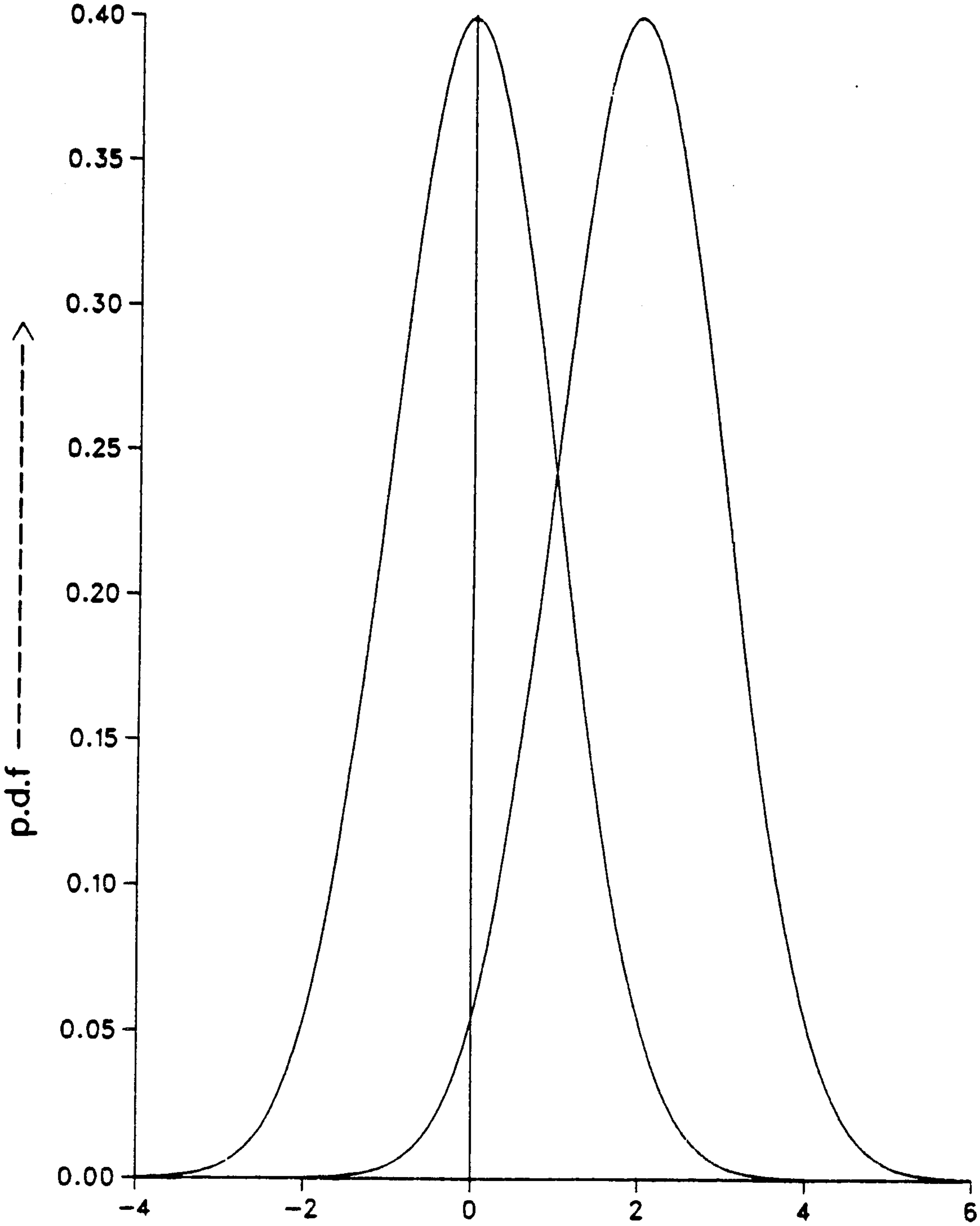


FIG.3 : N(0,1) AND N(3,1)

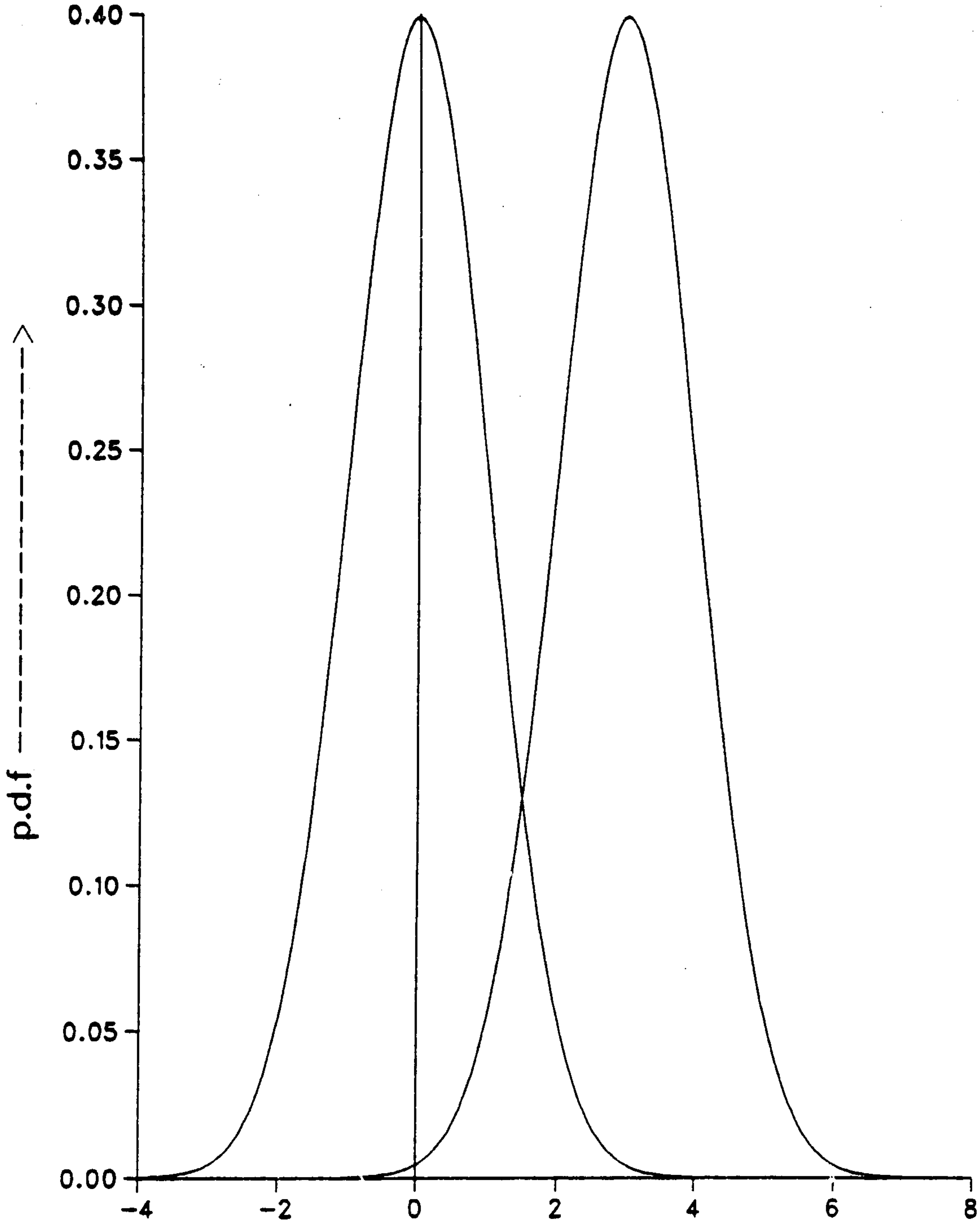


FIG.4 : N(0,1) AND LG(0,1)

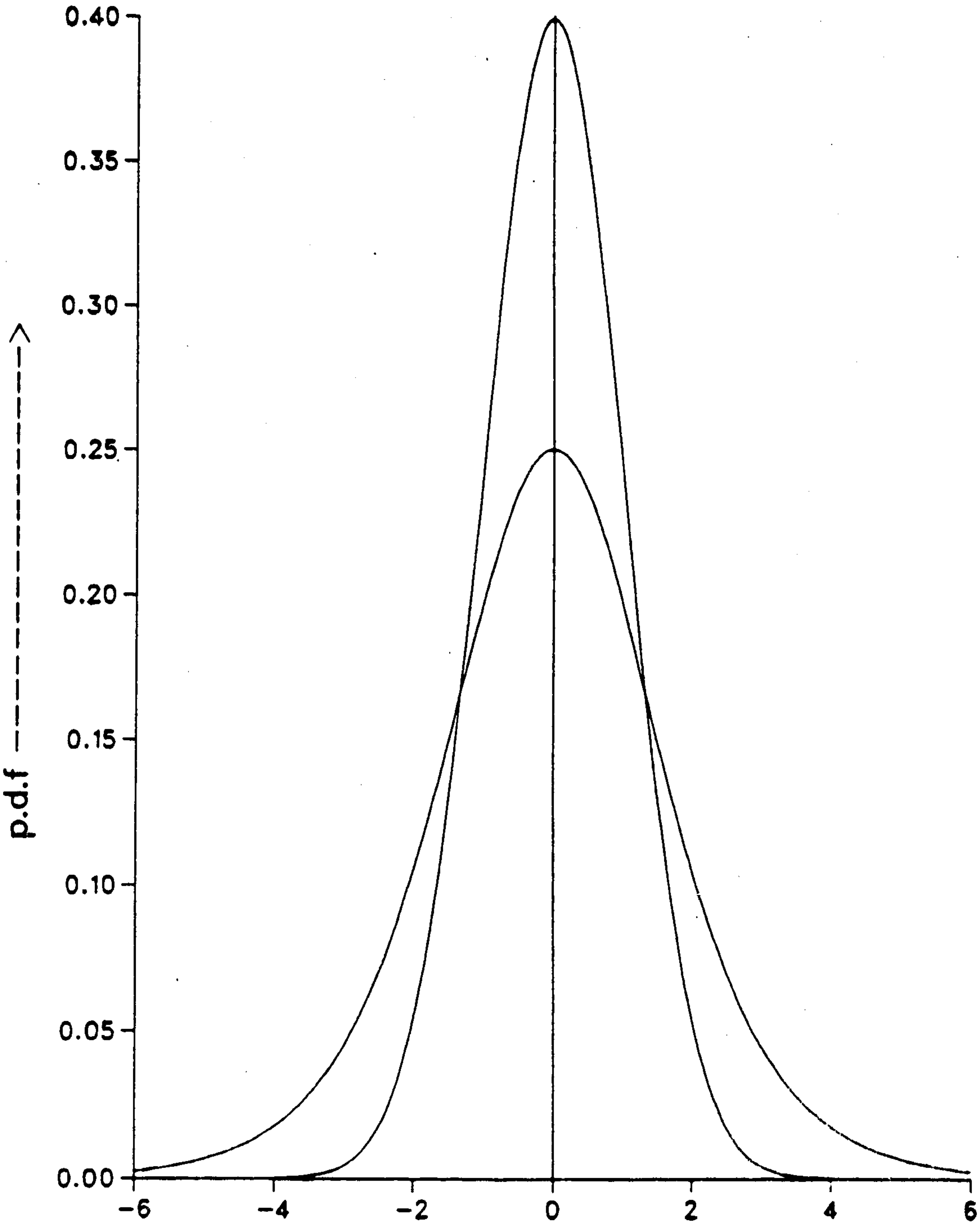


FIG.5 : N(0,1) AND LG(1,1)

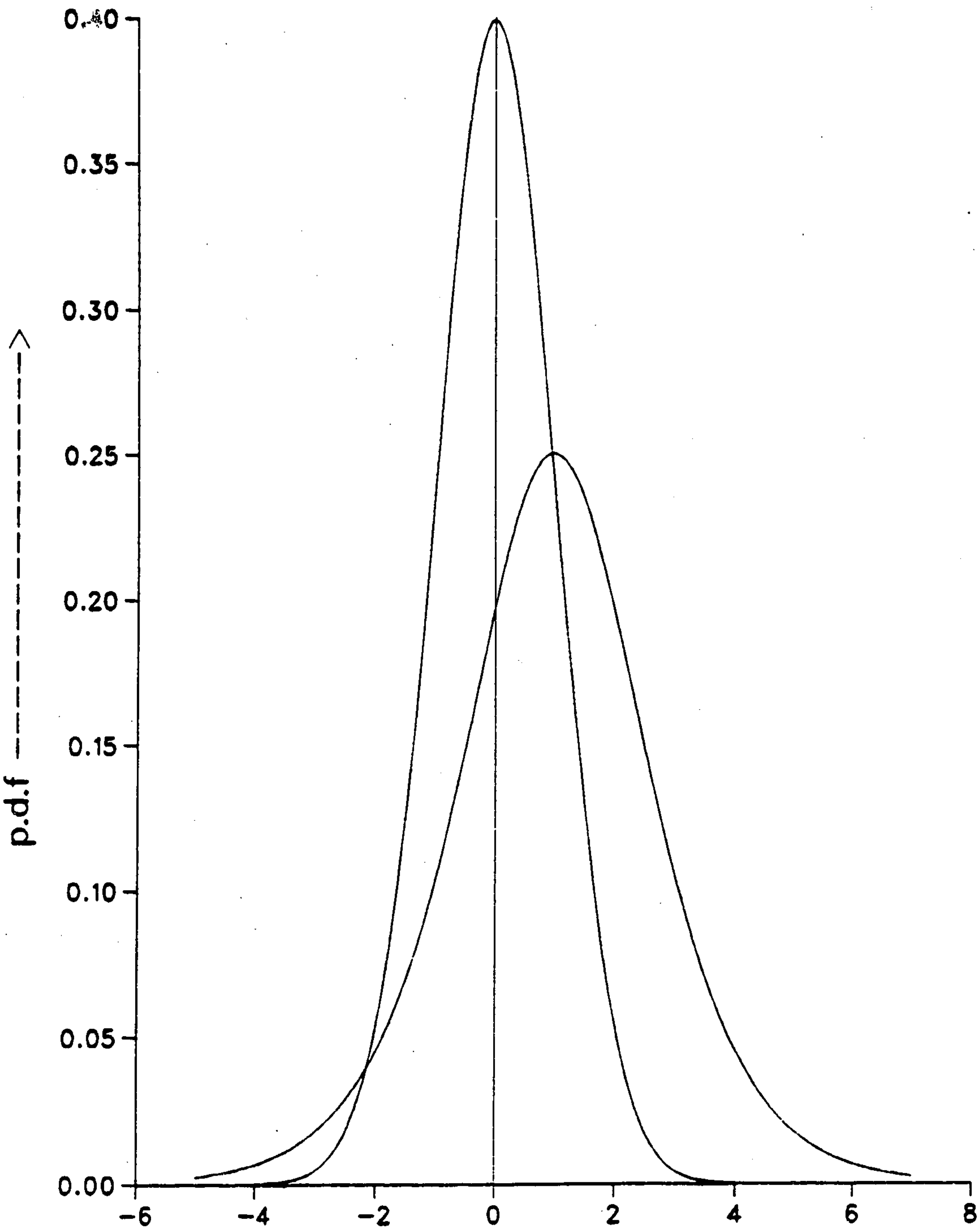


FIG.6 : N(0,1) AND LG(2,1)

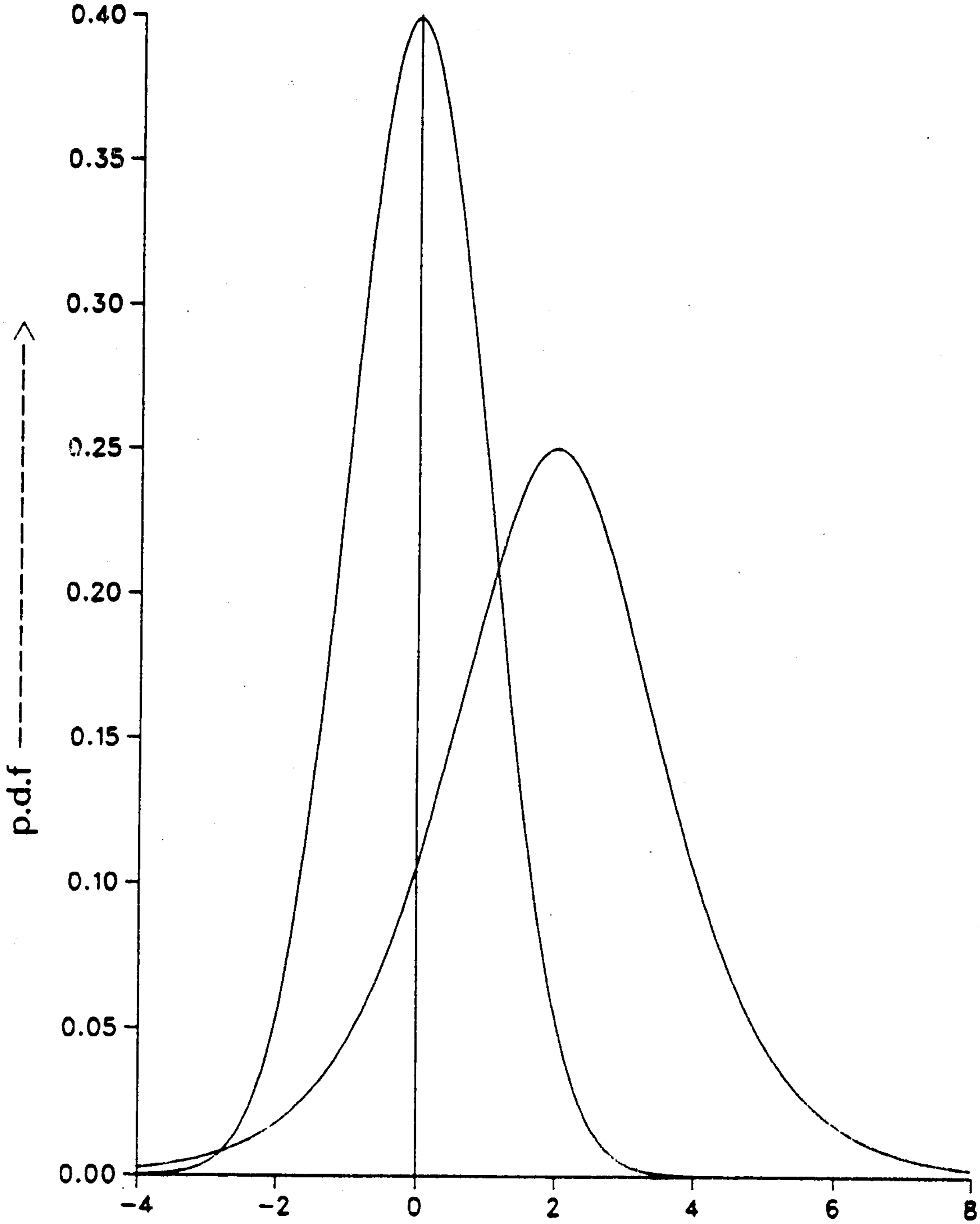


FIG.7 : N(0,1) AND LA(1,1)

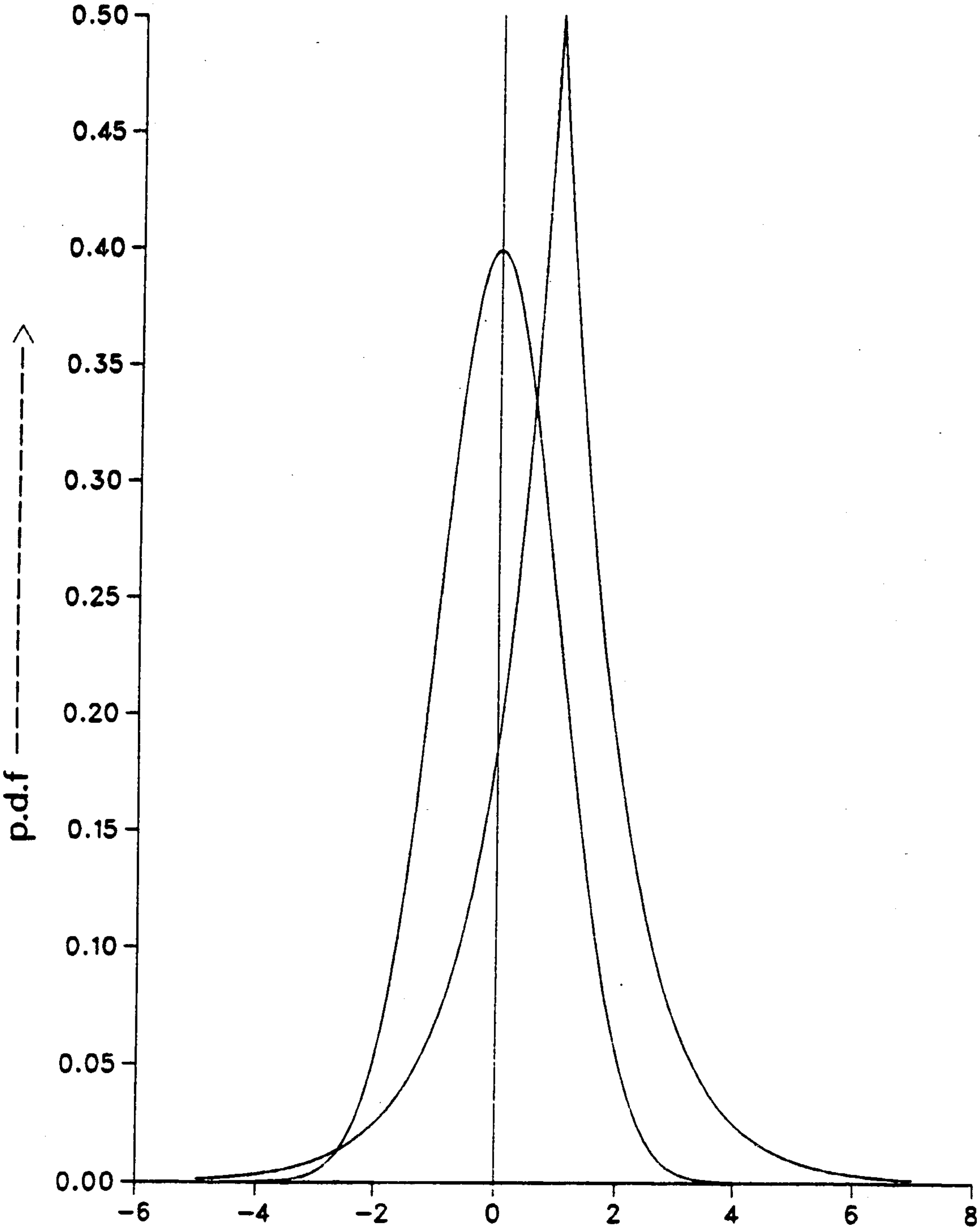


FIG.8 : N(0,1) AND LA(2,1)

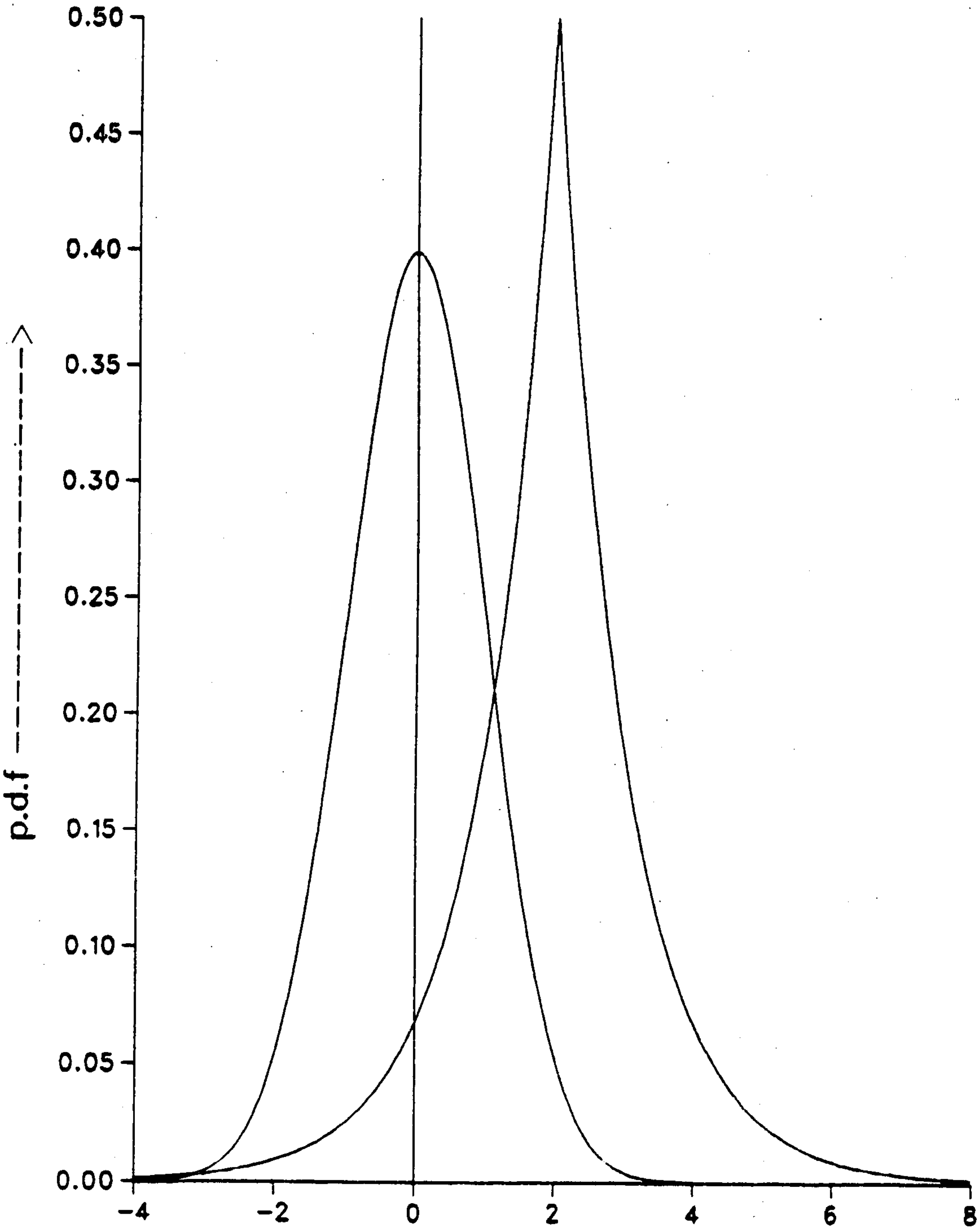


FIG.9 : LA(0,1) AND LA(1,1)

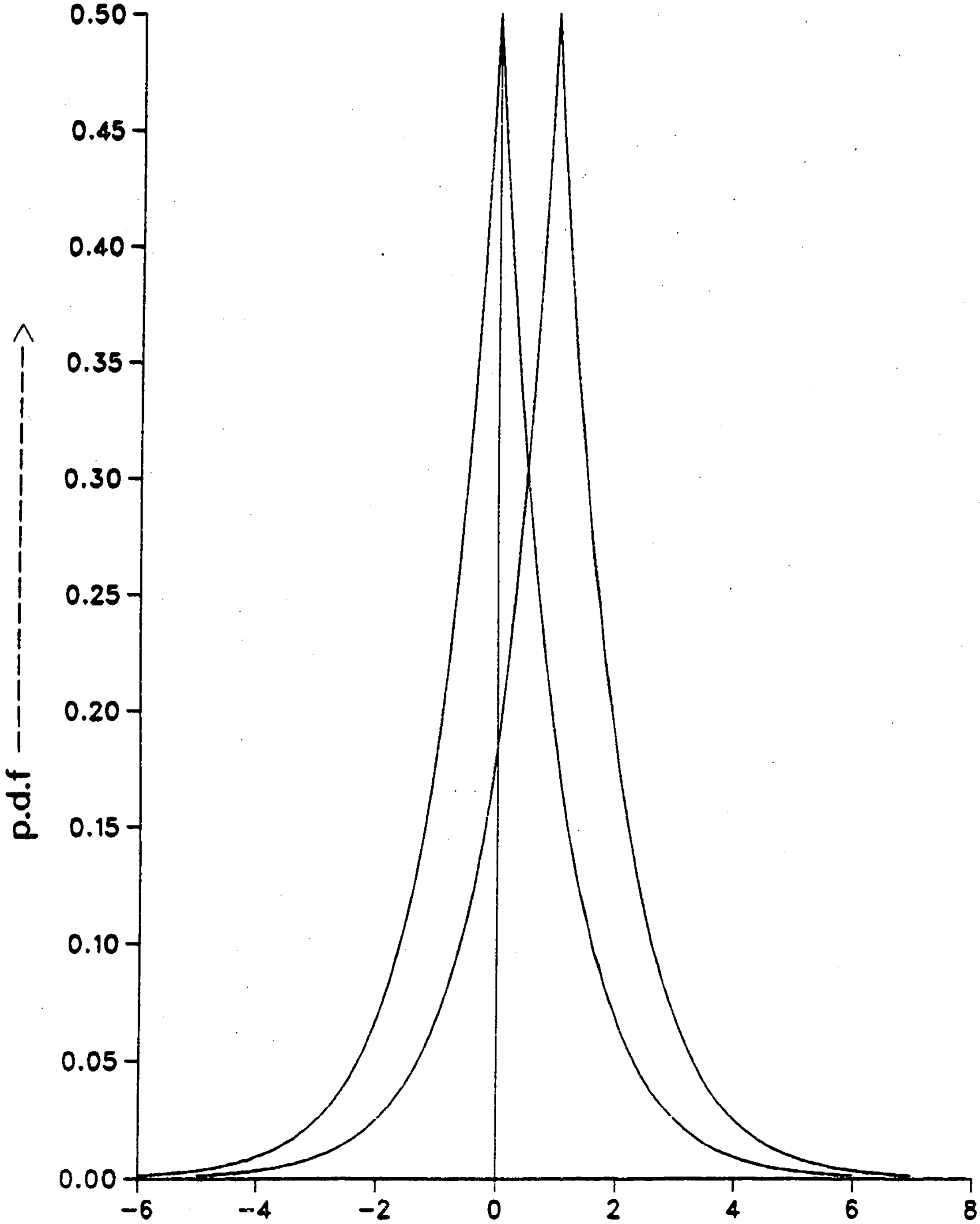


FIG.10 : LA(0,1) AND LA(2,1)

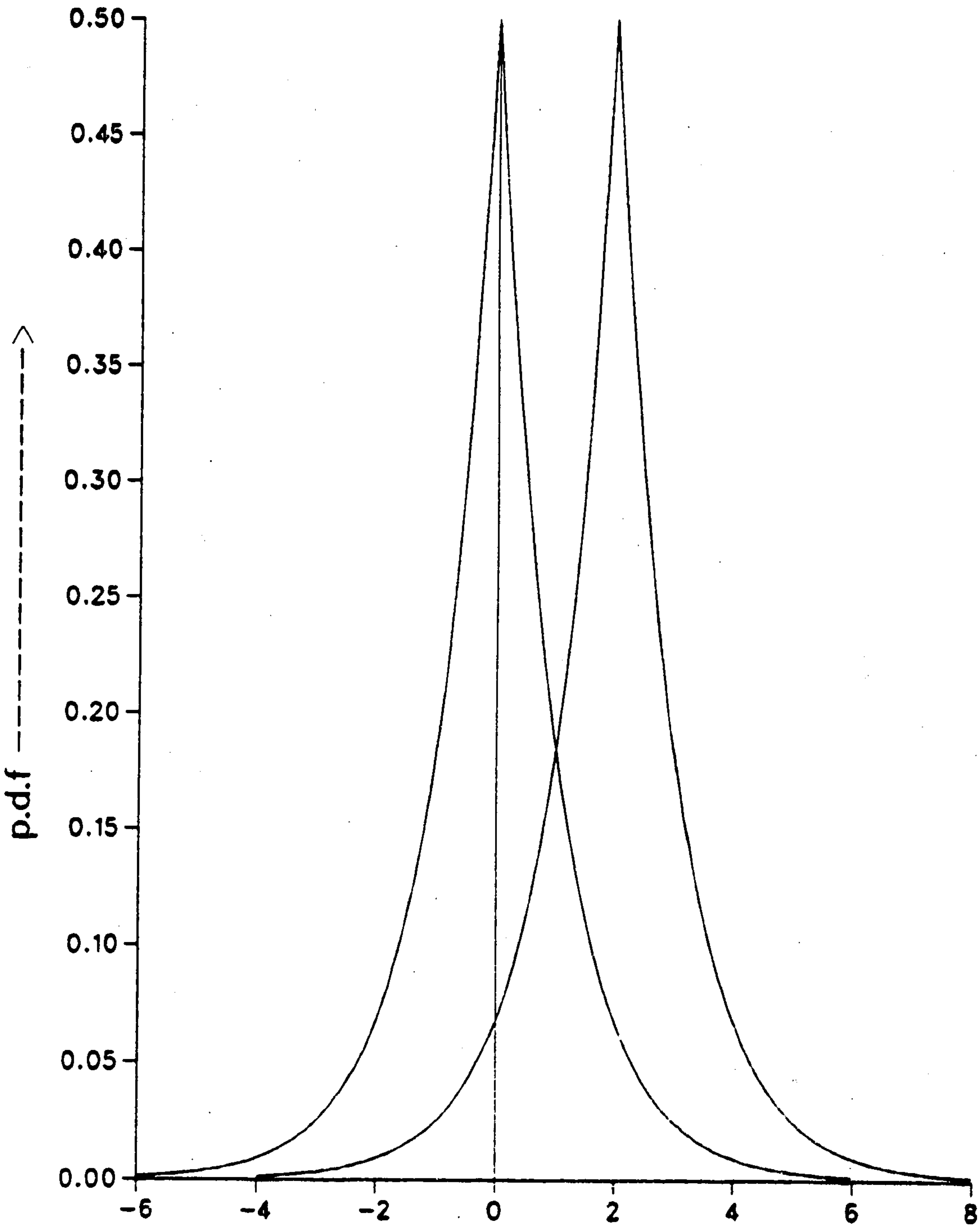


FIG.11 : LA(0,1) AND LA(3,1)

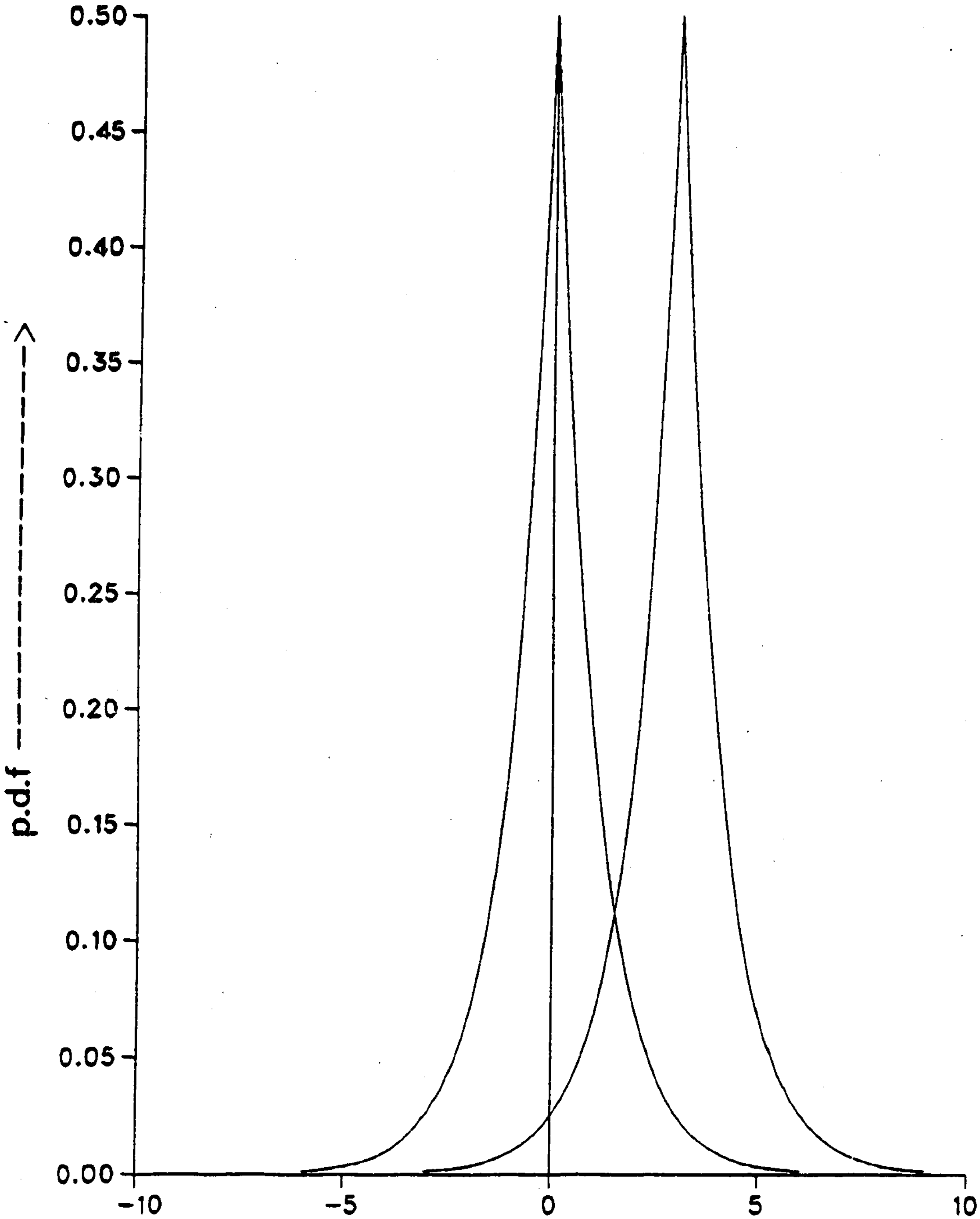


FIG.12 : LG(0,1) AND LG(2,1)

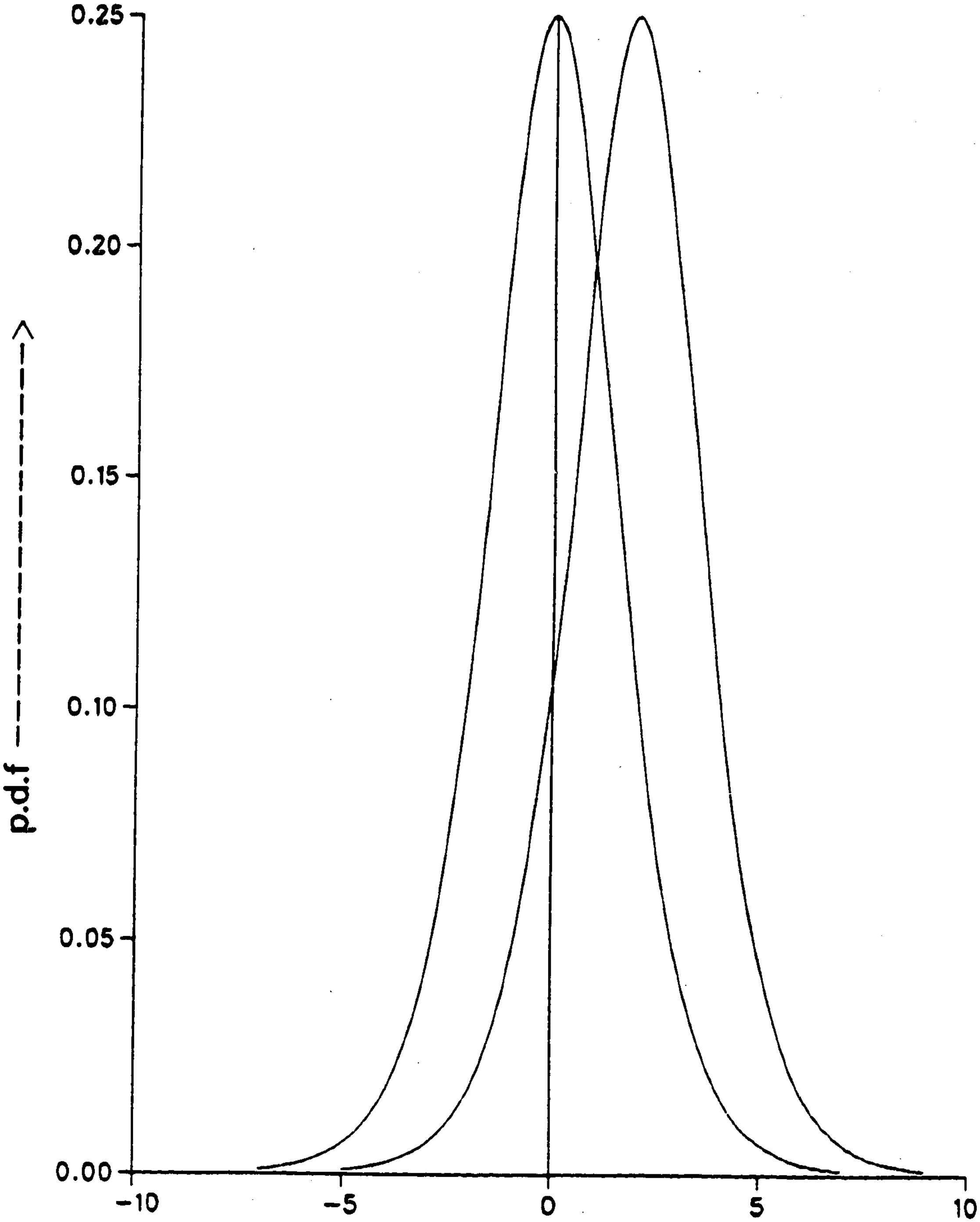


FIG.13 : EX(0,1) AND GA(2,1)

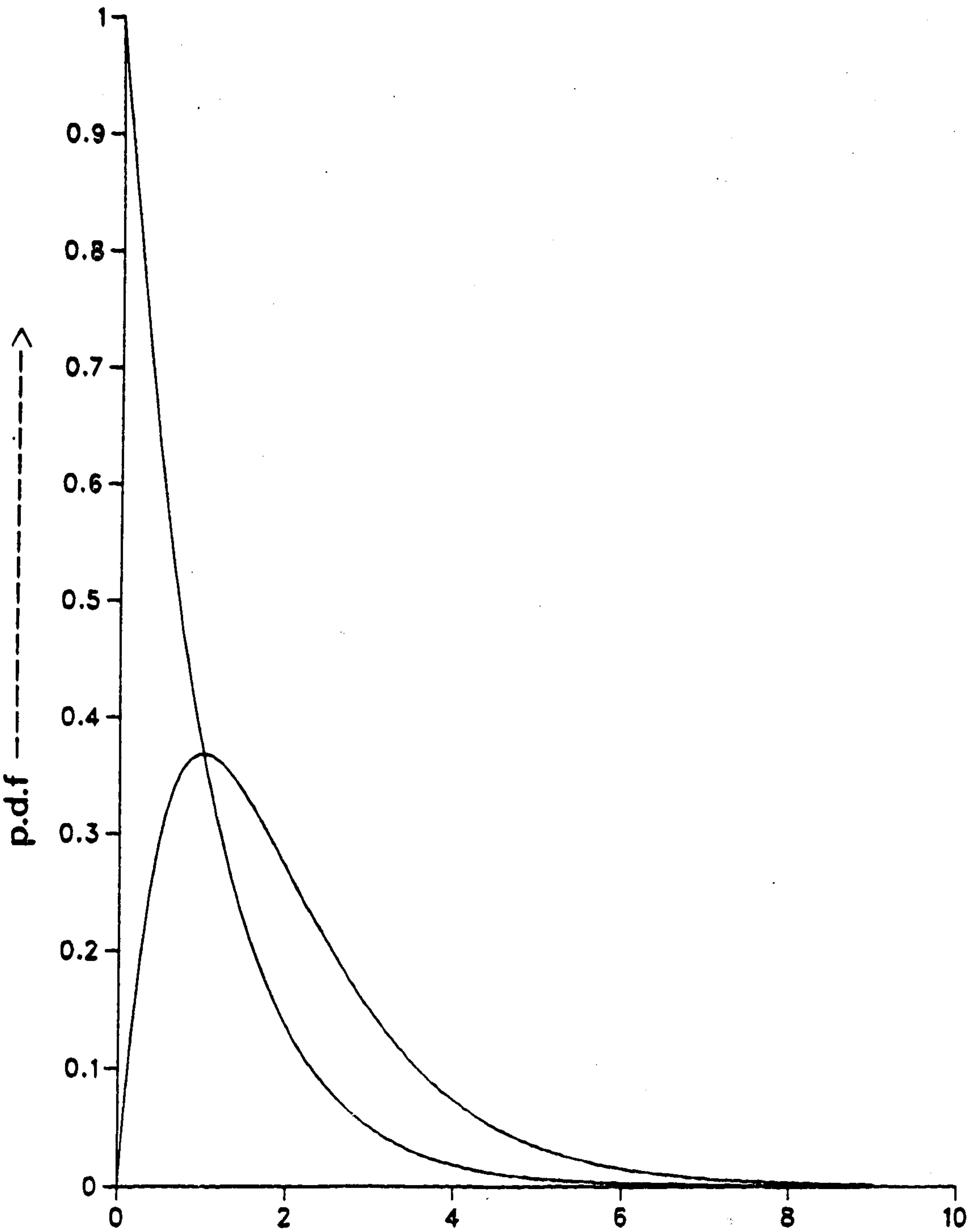


FIG.14 : EX(0,1) AND WE(2,1)

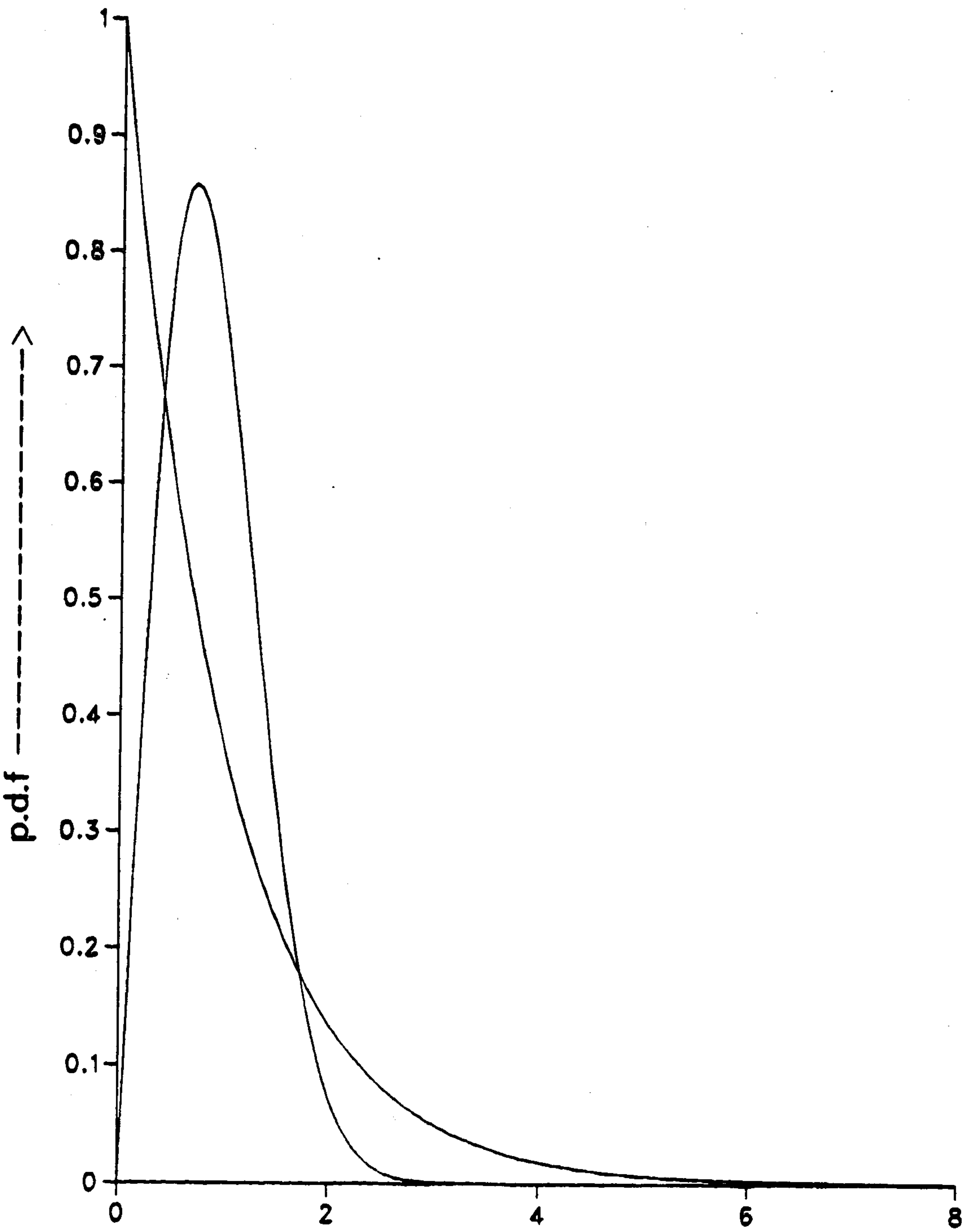
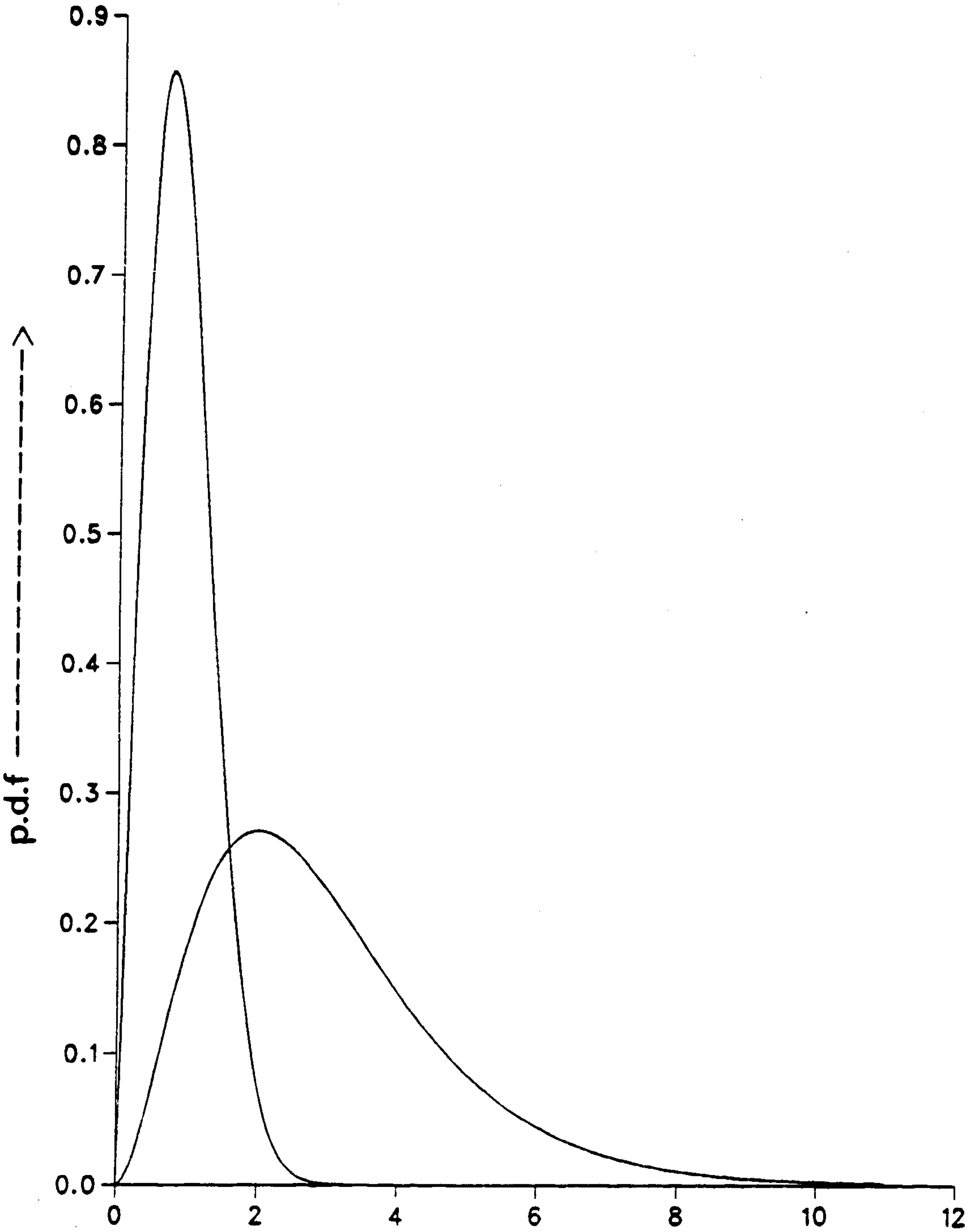


FIG.15 : WE(2,1) AND GA(3,1)



BIBLIOGRAPHY

ANDERSON, T.W. (1966). Some nonparametric multivariate procedures based on Statistical equivalent blocks. *Proc. 1st Internat. Symp. Anal.* Ed. P.R. Krishnaiah. Academic Press, New York. 5-127.

BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. and STONE, C.J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, Calif.

BROFFITT, J.D., RANDLES, R.H. and HOGG, R.V. (1976). Distribution-free partial discriminant analysis. *Jour. Amer. Statist. Assoc.* **73** 934-939.

CACOULLOS, T., ED. (1973). *Discriminant Analysis and Applications.* Academic Press, New York.

CHOI, S.C. (1972). Classification of mulitiply observed data. *Biometrische Zeitschrift,* **14** 8-11.

COVER, T.M. (1968). Estimation by the Nearest Neighbor rule. *IEEE Trans. Inform. Theory,* **IT-14** 50-55.

COVER, T.M. (1969). Learning in pattern recognition. *Methodologies of Pattern Recognition.* Ed. S. Watanabe. Academic Press, New York. 111-132.

COVER, T.M. and HART, P.E. (1967). Nearest Neighbor pattern classification. *IEEE Trans. Inform. Theory,* **IT-13** 21-26.

DAS GUPTA, S. (1962). On the optimum properties of some classification rules. *Inst. Statist. Mimeo,* No. **333** Univ. of N.C. Chapel Hill.

DAS GUPTA, S. (1964). Nonparametric classification rules. *Sankhyá* A. **26** 25-30.

DAS GUPTA, S. (1973). Theories and methods in classification: a review. *Discriminant Analysis and Applications.* Ed. T. Cacoullos. Academic Press, New York. 77-137.

DAS GUPTA, S. and LIN, H.E. (1980). Nearest Neighbor rules for Statistical classification based on ranks. *Sankhyá* A. **42** 219-230.

DEVROYE, L. (1981b). On the asymptotic probabilty of error in nonparametric discrimination. *Ann. Statist.* **9** 1320-1327.

FIX, E. and HODGES, J.L. (1951). Nonparametric discrimination: Consistency properties. *U.S. Air Force School of Aviation Medicine,* Report No. **4** Randolph Field, Texas.

FIX, E. and HODGES, J.L. (1952). Discriminatory analysis: nonparametric discrimination: Small Sample performance. *U.S. Air Force School of Aviation Medicine,* Report No. **11** Randolph Field, Texas.

FRITZ, J. (1975). Distribution-free exponential error bound for Nearest Neighbor pattern classification. *IEEE Trans. Inform. Theory,* **IT-21** 552-557.

HOEFFDING, W. (1963). Probability inequalities for the sums of bounded random variables. *Jour. Amer. Statist. Assoc.* **58** 13-30.

JOHNS, M.V. (1961). An empirical Bayes approach to nonparametric two-way classification. *Studies in Item Analysis and Prediction.* Ed. H. Solomon. Stanford University Press, Stanford, California.

LACHENBRUCH, P.A. (1975). *Discriminant Analysis.* Hafner Press, New York.

MATUSITA, K. (1956). Decision rule, based on the distance, for the classification problem. *Ann. Inst. Statist. Math.* **8** 67-77.

PETERSON, D.W. (1970). Some convergence properties of a nearest neighbor decision rule. *IEEE Trans. Inform. Theory,* **IT-16** 26-31.

QUESENBERRY, C.P. and GESSAMAN, M.P. (1968). Nonparametric discrimination using tolerance regions. *Ann. Math. Statist.* **39** 664-673.

RANDLES, R.H., BROFFITT, J.D. and RAMBERG, J.S. (1978). Discrimant analysis based on ranks. *Jour. Amer. Statist. Assoc.* **73** 379-384.

RYAN, B., JOINER, B.L., and RYAN, T.A., JR. (1985). *Minitab Handbook.* Duxbury Press, Boston.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons, New York.

VAN RYZIN, J. (1966). Bayes risk consistency of classification procedures using density estimation. *Sankhyá* A. **26** 25-30.

WAGNER, T.J. (1971). Convergence of the nearest neighbor rule. *IEEE Trans. Inform. Theory,* **IT-17** 566-571.

WELCH, B.L. (1939). Note on discriminant functions, *Biometrika,* **31** 218-220.

XIRU, C. (1985). Exponential posterior error bound for the k-NN discrimination rule. *Scientia Sinica.* A **XXVII** 673-682.

APPENDIX

PROGRAM A

This program has two parts, program A1 and program A2. For the given training samples X's and Y's , the program A1 (written in Minitab (see Ryan et. al.(1966)) generates 200 Z's (100 from each population), the observations to be classified. In order to compute, average PMC of RNN rule (Das Gupta & Lin's procedure) for each Z, it combines X's, Y's and Z's and arranges them in increasing order; then prints them under Order.1 (if Z is an X) and Order.2 (if Z is an Y) in codes1, 2 and 3, where 1, 2 and 3 respectively denote the X, Y and Z observations. At the same time,to compute average PMC of NN rule (Cover & Hart's procedure) for each Z, it computes the distances from all X's and Y's and then arrange the distances in increasing order and prints them under Cover.1 (if Z an X) and Cover.2 (if Z is an Y) in codes1 and 2,where 1 and 2 denote the distances from X's and Y's respectively.
Program A2 (written in Fortran) reads the output from program A1 and makes decision in each case according to the rules and writes them. Finally, it calculates the average PMC for both RNN and NN rules simultaneously.

PROGRAM A1

```
Name C1='SmplX' C2='SmplY' C3='Smpl1Z' C4='Smpl2Z'
Name C11='Order.1' C13='Order.2' C23='Cover.1' C25='Cover.2'
Let K4=100
Read 'DATA.XY.F' C1 C2
STORE 'RAVI'
NOECHO
Random 1 observation put into C3;
NORMAL mean=0 , s.d.=1.
Random 1 observation put into C4;
NORMAL mean=2 , s.d.=1.
STACK C1 C2 C3 into C5;
Subscripts C6.
STACK C1 C2 C3 into C7;
Subscripts C8.
SORT C5 carry C6 put into C10 C11
SORT C7 carry C8 put into C12 C13
LET C14=ABSOLUTE(C1-C3)
LET C15=ABSOLUTE(C2-C3)
LET C18=ABSOLUTE(C1-C4)
LET C19=ABSOLUTE(C2-C4)
STACK C14 C15 into C16;
Subscripts C17.
STACK C18 C19 into C20;
Subscripts C21.
SORT C16 carry C17 put into C22 C23
SORT C20 carry C21 put into C24 C25
Delete 3:20 of C23
Delete 3:20 of C25
PRINT C11
PRINT C13
PRINT C23
PRINT C25
END
EXECUTE 'RAVI' K4
Stop
```

PROGRAM A2

```
*******************************************************
*
*
*    PROGRAM TPMC
*
* THIS PROGRAM CALCULATES AVERAGE PMC OF
* RNN RULE & NN RULE ( K=1 ) SIMULTANOUSLY
*
*******************************************************


*******************************************************
*********
*********        MAIN  PROGRAM
*********
*******************************************************


      CHARACTER*80 A,B*405
       OPEN(UNIT=20,FILE='SUBHAS',STATUS='OLD')
       OPEN(UNIT=30,FILE='SCAN',STATUS='OLD')
      CLOSE(UNIT=30,STATUS='DELETE')
       OPEN(UNIT=30,FILE='SCAN',STATUS='NEW')
      IORDER=0
      JORDER=0
      CRT=0
      CWR=0
      WRONG=0.
      RIGHT=0.

**********************************************************
*
* CHANGE THE VALUE OF "M" FOR TESTING DIFFERENT
* NUMBER  OF 3S
                    M=1
**********************************************************
      PRINT*,' '
      PRINT*,' '
      PRINT*,' '
      PRINT*,'*********************************************'
      PRINT*,'*                                           *'
      PRINT*,'*                                           *'
      PRINT*,'*  WELCOME TO "DEB.SCAN", SUBHAS!!!         *'
      PRINT*,'*                                           *'
      PRINT*,'*                                           *'
      PRINT*,'*********  PROGRAM RUN STARTS **********'
      PRINT*,' '
100   READ(20,'(A80)',END=500)A
      KC1=INDEX(A,'Cover.1        ')
      KC2=INDEX(A,'Cover.2        ')
      K=INDEX(A,'Order.1    ')
      KK=INDEX(A,'Order.2    ')
```

```
        IF(K.NE.0.OR.KK.NE.0)THEN
              ITYPE=1
              IORDER=1
        ELSE IF(KC1.NE.0.OR.KC2.NE.0)THEN
              ITYPE=0
              JORDER=1
        END IF

        IF(ITYPE.EQ.1.OR.ITYPE.EQ.0)THEN
  WRITE(30,*)'1          *****************************'
  WRITE(30,*)'           *                           *'
  WRITE(30,*)'           * PROGRAM TPMC OUTPUT *'
  WRITE(30,*)'           * ----------          *'
  WRITE(30,*)'           * CODED BY: S. BAGUI        *'
  WRITE(30,*)'           * DEPT. OF STATISTICS       *'
  WRITE(30,*)'           * UNIV. OF ALBERTA          *'
  WRITE(30,*)'           * DATE   : 89/04/30   *'
  WRITE(30,*)'           * CODE   : FORTRAN    *'
  WRITE(30,*)'           *****************************'
  WRITE(30,*)' '
  WRITE(30,*)' '
  WRITE(30,*)' *************************************'
  WRITE(30,*)' **** PROGRAM OUTPUT STARTS ****'
  WRITE(30,*)' *************************************'
  WRITE(30,*)' '
  WRITE(30,*)A

  IF(ITYPE.EQ.1)THEN
  PRINT*,'-> NOW PROCESSING (DASGUPTA) SAMPLE NUMBER:',
 * IORDER
  WRITE(30,*)' (DASGUPTA) SAMPLE NUMBER: ',IORDER
  END IF

  IF(ITYPE.EQ.0)THEN
  WRITE(30,*)' (COVER) SAMPLE NUMBER: ',JORDER
  PRINT*,'-> NOW PROCESSING (COVER) SAMPLE NUMBER:',
 * JORDER
  END IF

  WRITE(30,*)' '
   READ(A(8:8),'(I1)')NORDER
  WRITE(30,*)' '
        GO TO 110

        END IF
  GO TO 100


110   DO 200 I=1,10000
   CALL READER(A,NORDER,NORD,B,ISTART,RIGHT,
  *WRONG,IORDER,M,CRT,CWR,JORDER,ITYPE,ITP)
   IF(ITYPE.EQ.1)CALL SEE(B,ISTART,NORDER,RIGHT,
  * WRONG, M)

        ITYPE=ITP

   IF(ITYPE.EQ.1)IORDER = IORDER + 1
   IF(ITYPE.EQ.0)JORDER=JORDER+1
```

```
      WRITE(30,*)' '
      WRITE(30,*)'****************************************'
      WRITE(30,*)' '
      WRITE(30,*)A

      IF(ITYPE.EQ.1)THEN
      PRINT*,'-> NOW PROCESSING (DASGUPTA) SAMPLE NUMBER:',
     * IORDER
      WRITE(30,*)' (DASGUPTA) SAMPLE NUMBER: ',IORDER
      END IF

      IF(ITYPE.EQ.0)THEN
      WRITE(30,*)' (COVER) SAMPLE NUMBER: ',JORDER
      PRINT*,'-> NOW PROCESSING (COVER) SAMPLE NUMBER :',
     * JORDER
      END IF

      WRITE(30,*)' '
      NORDER = NORD
 200  CONTINUE
 500  STOP
      END


*********************************************************
********
********    READER
********
*********************************************************


      SUBROUTINE READER(A,NORDER,NORD,B,ISTART,RIGHT,
     *WRONG,IORDER,M,CRT,CWR,JORDER,ITYPE,ITP)
      INTEGER NUM(50)
      CHARACTER*(*)A,B
      ISTART=0

 100  READ(20,'(A80)',END=555)A
      K=INDEX(A,'Order.1   ')
      KK=INDEX(A,'Order.2   ')
      KC1=INDEX(A,'Cover.1        ')
      KC2=INDEX(A,'Cover.2        ')
      J=INDEX(A,'*** Minitab')

      IF(J.NE.0.)THEN
 555  IF(ITYPE.EQ.1)CALL SEE(B,ISTART,NORDER,RIGHT,
     * WRONG, M)
      WRITE(30,*)' '
      WRITE(30,*)'****************************************'
      WRITE(30,*)' '
      WRITE(30,*)'                        SUMMARY: '
      WRITE(30,*)' '
      WRITE(30,*)'  NOTE: '
      WRITE(30,*)' #1. THE POPULATION DATA WAS READ FROM ',
     *'FILE: SUBHAS'
      WRITE(30,*)'#2. THE NUMBER OF 3s IN EACH POPULATION ',
     *'IS: ',M
```

```
      WRITE(30,44)
 44   FORMAT('                #3. IF THE NUMBER OF 3s ARE CHANGED, ',
     *'ALTER THE VALUE OF '/
     *'        "M" IN LINE  32 OF FILE "DEB.SCAN", AND THEN ',
     *'RECOMPILE '/
     *'          DEB.SCAN, BEFORE RUNNING BATCHFILE.')
      WRITE(30,*)' '
      WRITE(30,*)'                           ------------'
      WRITE(30,*)' '
      WRITE(30,*)' '
      WRITE(30,*)'           DASGUPTA SAMPLE ANALYSIS'
      WRITE(30,*)'           ------------------'
      WRITE(30,*)' '
      WRITE(30,*)' TOTAL NUMBER OF  SAMPLES ARE : ',IORDER
      WRITE(30,*)'TOTAL NUMBER OF CORRECT CLASSIFICATIONS ARE',
     *' : ',RIGHT
      WRITE(30,*)'TOTAL NUMBER OF WRONG CLASSIFICATIONS ARE  ',
     *' : ',WRONG
      WRITE(30,*)' '
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'MIS-CLASSIFICATION IS    : ',WRONG/IORDER
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'CORRECT CLASSIFICATION IS : ',RIGHT/IORDER
      WRITE (30,*)' '
      WRITE(30,*)' '
      WRITE(30,*)' '
      WRITE(30,*)'           COVER SAMPLE ANALYSIS'
      WRITE(30,*)'           ---------------'
      WRITE(30,*)' '
      WRITE(30,*)' TOTAL NUMBER OF  SAMPLES ARE : ',JORDER
      WRITE(30,*)'TOTAL NUMBER OF CORRECT CLASSIFICATIONS ARE',
     *' : ',CRT
      WRITE(30,*)'TOTAL NUMBER OF WRONG CLASSIFICATIONS ARE  ',
     *' : ',CWR
      WRITE(30,*)' '
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'MIS-CLASSIFICATION IS    : ',CWR/JORDER
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'CORRECT CLASSIFICATION IS : ',CRT/JORDER
      WRITE (30,*)' '
      WRITE(30,*)'**********************************************'
      WRITE(30,*)'****  N(0,1)  N(2,1)  N=50  R(1) XY.F ******'
      WRITE(30,*)'**********************************************'
      PRINT*,' '
      PRINT*,'************* PROGRAM RUN ENDS **************'
      PRINT*,'*                                            *'
      PRINT*,'*        RESULTS ARE IN FILE "SCAN"          *'
      PRINT*,'*                                            *'
      PRINT*,'**********************************************'
500   STOP
      END IF


         IF(K.NE.0.OR.KK.NE.0.OR.KC1.NE.0.OR.KC2.NE.0)THEN
      IF(K.NE.0.OR.KK.NE.0)ITP=1
      IF(KC1.NE.0.OR.KC2.NE.0)ITP=0
      READ(A(8:8),'(I1)')NORD
         GO TO 103
```

```
      END IF

      IF(ITYPE.EQ.0)THEN
           DO 112 I=1,80
             IF(A(I:I).NE.' ')THEN
               READ(A(I:I),'(I1)')NVAL
             GO TO 113
            END IF
 112       CONTINUE
 113      IF(NVAL.EQ.NORDER)THEN
      WRITE(30,*)' '
      WRITE(30,*)'FIRST ELEMENT OF THE SAMPLE IS : ',NVAL
      WRITE(30,*)' '
      WRITE(30,*)'  ***  CORRECT CLASSIFICATION!!!! '
      WRITE(30,*)'   '
           CRT=CRT+1
          ELSE
      WRITE(30,*)' '
      WRITE(30,*)' FIRST ELEMENT OF THE SAMPLE IS : ',NVAL
      WRITE(30,*)' '
      WRITE(30,*)'  ***  MISCLASSIFICATION!!!! '
      WRITE(30,*)'   '
           CWR=CWR+1
          END IF
     ITYPE=2
     END IF


     IF(ITYPE.EQ.1)THEN
      INUM = 0
     DO 101 I=1,80
           IF(A(I:I).NE.' ')THEN
             INUM=INUM+1
              READ(A(I:I),'(I1)')NUM(INUM)
            ISTART=ISTART+1
              WRITE(B(ISTART:ISTART),'(I1)')NUM(INUM)
          END IF
 101   CONTINUE
     END IF

     GO TO 100
 103   RETURN
     END

*********************************************************
********
********   SEE
********
*********************************************************


     SUBROUTINE SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
     CHARACTER*(*)B,STRING*5,STORE*5
    N1=0
    N2=0
    STRING='33333'
    STORE='00000'
```

```
      DO 80 I=6-M,5
      K5=0
      K1=0

60    K5=INDEX(B(K1+1:ISTART),STRING(I:))
      K1=K1+K5
            IF(K5.EQ.0)THEN
              GO TO 80
            ELSE
              IF(K1.EQ.1)THEN
                   READ(B(K1+6-I:K1+6-1),'(I1)') IR
                   IF(IR.EQ.1)N1=N1+1*(6-I)
                   IF(IR.EQ.2)N2=N2+1*(6-I)
              ELSE IF(K1.EQ.ISTART-5+I)THEN
                   READ(B(K1-1:K1-1),'(I1)') IL
                   IF(IL.EQ.1)N1=N1+1*(6-I)
                   IF(IL.EQ.2)N2=N2+1*(6-I)
              ELSE
                   READ(B(K1-1:K1-1),'(I1)')IL
                   IF(IL.EQ.1)N1=N1+1*(6-I)
                   IF(IL.EQ.2)N2=N2+1*(6-I)
                   READ(B(K1+6-I:K1+6-I),'(I1)')IR
                   IF(IR.EQ.1)N1=N1+1*(6-I)
                   IF(IR.EQ.2)N2=N2+1*(6-I)
              END IF
                   B(K1:K1+5-I)=STORE(I:)

                   GO TO 60
            ENDIF

80    CONTINUE

      WRITE(30,*)'  # RNN FROM POPULATION.1 IS : ',N1
      WRITE(30,*)'  # RNN FROM POPULATION.2 IS : ',N2

      WR=0.
      RI=0.


      IF(NORDER.EQ.1)THEN
            IF(N1.LT.N2)WR=1.
            IF(N1.EQ.N2)WR=0.5
            IF(N1.EQ.N2)RI=0.5
            IF(N1.GT.N2)RI=1.
         ELSE
            IF(N2.LT.N1)WR=1.
            IF(N2.EQ.N1)WR=0.5
            IF(N2.EQ.N1)RI=0.5
            IF(N2.GT.N1)RI=1.
         END IF

      WRITE(30,*)' '
      IF(N1.GT.N2)WRITE(30,*)'---> # RNN OF POPULATION.1 ',
     *' >  # RNN OF POPULATION.2'
      IF(N2.GT.N1)WRITE(30,*)'---> # RNN OF POPULATION.2 ',
     *' >  # RNN OF POPULATION.1'
      IF(N1.EQ.N2)WRITE(30,*)'---> # RNN OF POPULATION.1 ',
     *' =  # RNN OF POPULATION.2'
```

```
 WRITE(30,*)' '
  IF(WR.EQ.1.)WRITE(30,*)'** MIS-CLASSIFICATION!!!!!!'
  IF(WR.EQ.0.5)WRITE(30,*)'** ERROR FUNCTION IS HALF!'
  IF(RI.EQ.1.)WRITE(30,*)'** CORRECT CLASSIFICATION!!'
  WRONG=WRONG+WR
  RIGHT=RIGHT+RI
 RETURN
END
```

PROGRAM B

This program calculates the average PMC for the first stage RNN rule when up to five multiple observations (K) are classified. It consists of two parts , program B1 and program B2. Similar to the program A , program B2 reads the output from program B1 and makes a decision at each case and finally gives the average PMC of the MRNN rule. In order to compute the average PMC for MRNN rule for different values of K, one should repeat the program B changing the value of K.

PROGRAM B1

```
Name C1='SmplX' C2='SmplY' C3='smpl1Z' C4='Smpl2Z'
Name C11='Order.1' C13='Order.2'
LET K4=100
READ 'DATA.XY.F' C1 C2
STORE 'MALA'
NOECHO
Random 5 observations put into C3;
NORMAL mean=0, s.d.=1.
Random 5 observations put into C4;
NORMAL mean=2, s.d.=1.
STACK C1 C2 C3 into C5;
Subscripts C6.
STACK C1 C2 C4 into C7;
Subscripts C8.
SORT C5 carry C6 put into C10 C11
SORT C7 carry C8 put into C12 C13
PRINT C11
PRINT C13
END
EXECUTE 'MALA' K4
STOP
```

PROGRAM B2

```
*********************************************************
*
*
*   PROGRAM TPMC
*
*   THIS PROGRAM CALCULATES AVERAGE PMC OF
*   FIRST-STAGE ( K=1,2,3,4,5 ) RNN RULE
*
*********************************************************


*********************************************************
*********
*********        MAIN  PROGRAM
*********
*********************************************************


     CHARACTER*80 A,B*405
      OPEN(UNIT=20,FILE='SUBHAS',STATUS='OLD')
      OPEN(UNIT=30,FILE='SCAN',STATUS='OLD')
     CLOSE(UNIT=30,STATUS='DELETE')
      OPEN(UNIT=30,FILE='SCAN',STATUS='NEW')
     IORDER=1
     WRONG=0.
     RIGHT=0.

*********************************************************
*
* CHANGE THE VALUE OF "M" FOR TESTING DIFFERENT
*  NUMBER OF 3S
                    M=5
*********************************************************


100  READ(20,'(A80)',END=500)A
     K=INDEX(A,'Order.1   ')
     KK=INDEX(A,'Order.2   ')
          IF(K.NE.0.OR.KK.NE.0)THEN
     READ(A,'(I1)')NORDER
     WRITE(30,*)'          ******************************'
     WRITE(30,*)'          *                            *'
     WRITE(30,*)'          *  PROGRAM TPMC OUTPUT  *'
     WRITE(30,*)'          *  ------------       *'
     WRITE(30,*)'          *  CODED BY : S. BAGUI     *'
     WRITE(30,*)'          *  DEPT OF STATISTICS      *'
     WRITE(30,*)'          *  UNIV. OF ALBERTA             *'
     WRITE(30,*)'          *  DATE       : 89/04/15       *'
     WRITE(30,*)'          *  CODE       : FORTRAN        *'
     WRITE(30,*)'          ******************************'
     WRITE(30,*)' '
     WRITE(30,*)' '
     WRITE(30,*)' ***************************************'
```

```
WRITE(30,*)' **** PROGRAM OUTPUT STARTS ****'
WRITE(30,*)' *****************************************'
WRITE(30,*)' '
WRITE(30,*)A
WRITE(30,*)' SAMPLE NUMBER: ',IORDER
WRITE(30,*)' '
 READ(A(8:8),'(I1)')NORDER
WRITE(30,*)' '
      GO TO 110
      END IF
GO TO 100


110  DO 200 I=1,10000
   CALL READER(A,NORDER,NORD,B,ISTART,RIGHT,
  * WRONG, IORDER, M)
   CALL SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
   IORDER = IORDER + 1
   WRITE(30,*)' '
   WRITE(30,*)'*************************************'
   WRITE(30,*)' '
   WRITE(30,*)A
   WRITE(30,*)' SAMPLE NUMBER: ',IORDER
   WRITE(30,*)' '
   NORDER = NORD
200  CONTINUE
500  STOP
   END


*********************************************************
********
********     READER
********
*********************************************************


   SUBROUTINE READER(A,NORDER,NORD,B,ISTART,RIGHT,
  * WRONG, IORDER, M)
   INTEGER NUM(50)
   CHARACTER*(*)A,B
   ISTART=0

100  READ(20,'(A80)',END=500)A
   K=INDEX(A,'Order.1   ')
   KK=INDEX(A,'Order.2   ')
   J=INDEX(A,'*** Minitab')

   IF(J.NE.0.)THEN
   CALL SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
   WRITE(30,*)' '
   WRITE(30,*)'*****************************************'
   WRITE(30,*)' '
   WRITE(30,*)'                        SUMMARY:  '
   WRITE(30,*)' '
   WRITE(30,*)'  NOTE:  '
   WRITE(30,*)' #1. THE POPULATION DATA WAS READ FROM',
  *'FILE: SUBHAS'
```

```
      WRITE(30,*)' #2. THE NUMBER OF 3s IN EACH POPULATION',
     *'IS: ',M
      WRITE(30,44)
 44   FORMAT('              #3. IF THE NUMBER OF 3s ARE CHANGED,',
     *'ALTER THE VALUE OF '/
     *'         "M" IN LINE  32 OF FILE "DEB.SCAN", AND THEN',
     *'RECOMPILE '/
     *'              DEB.SCAN, BEFORE RUNNING BATCHFILE.')
      WRITE(30,*)' '
      WRITE(30,*)'                        ---------- '
      WRITE(30,*)' '
      WRITE(30,*)' TOTAL NUMBER OF  SAMPLES ARE : ',IORDER
      WRITE(30,*)'TOTAL NUMBER OF CORRECT CLASSIFICATIONS ARE',
     *' : ',RIGHT
      WRITE(30,*)'TOTAL NUMBER OF WRONG CLASSIFICATIONS ARE',
     *' : ',WRONG
      WRITE(30,*)' '
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'MIS-CLASSIFICATION IS    : ',WRONG/IORDER
      WRITE(30,*)'   TOTAL PROBABILITY OF ',
     *'CORRECT CLASSIFICATION IS : ',RIGHT/IORDER
      WRITE (30,*)' '
      WRITE(30,*)'*******************************************'
      WRITE(30,*)'***** N(0,1)  N(2,1) XY.F  R(5) *********'
      WRITE(30,*)'*******************************************'
500   STOP
      END IF


          IF(K.NE.0.OR.KK.NE.0)THEN
      READ(A,'(I1)')NORD
      READ(A(8:8),'(I1)')NORD
          GO TO 103
      END IF


      INUM = 0
      DO 101 I=1,80
          IF(A(I:I).NE.' ')THEN
            INUM=INUM+1
             READ(A(I:I),'(I1)')NUM(INUM)
           ISTART=ISTART+1
             WRITE(B(ISTART:ISTART),'(I1)')NUM(INUM)
          END IF
101   CONTINUE
      GO TO 100
103   RETURN
      END

*****************************************************
********
********   SEE
********
*****************************************************


      SUBROUTINE SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
      CHARACTER*(*)B,STRING*5,STORE*5
```

```
      N1=0
      N2=0
      STRING='33333'
      STORE='00000'

      DO 80 I=6-M,5
      K5=0
      K1=0

60    K5=INDEX(B(K1+1:ISTART),STRING(I:))
      K1=K1+K5
          IF(K5.EQ.0)THEN
           GO TO 80
          ELSE
            IF(K1.EQ.1)THEN
                READ(B(K1+6-I:K1+6-1),'(I1)') IR
                IF(IR.EQ.1)N1=N1+1*(6-I)
                IF(IR.EQ.2)N2=N2+1*(6-I)
            ELSE IF(K1.EQ.ISTART-5+I)THEN
                READ(B(K1-1:K1-1),'(I1)') IL
                IF(IL.EQ.1)N1=N1+1*(6-I)
                IF(IL.EQ.2)N2=N2+1*(6-I)
           ELSE
                READ(B(K1-1:K1-1),'(I1)')IL
                IF(IL.EQ.1)N1=N1+1*(6-I)
                IF(IL.EQ.2)N2=N2+1*(6-I)
                READ(B(K1+6-I:K1+6-I),'(I1)')IR
                IF(IR.EQ.1)N1=N1+1*(6-I)
                IF(IR.EQ.2)N2=N2+1*(6-I)
           END IF
                B(K1:K1+5-I)=STORE(I:)

                GO TO 60
          ENDIF

80    CONTINUE

     WRITE(30,*)' # RNN FROM POPULATION.1 IS : ',N1
     WRITE(30,*)' # RNN FROM POPULATION.2 IS : ',N2

     WR=0.
     RI=0.


     IF(NORDER.EQ.1)THEN
          IF(N1.LT.N2)WR=1.
          IF(N1.EQ.N2)WR=0.5
          IF(N1.EQ.N2)RI=0.5
          IF(N1.GT.N2)RI=1.
        ELSE
          IF(N2.LT.N1)WR=1.
          IF(N2.EQ.N1)WR=0.5
          IF(N2.EQ.N1)RI=0.5
          IF(N2.GT.N1)RI=1.
        END IF

    WRITE(30,*)' '
    IF(N1.GT.N2)WRITE(30,*)'---> # RNN OF POPULATION.1',
```

```
*'>  # RNN OF POPULATION.2'
 IF(N2.GT.N1)WRITE(30,*)'---> # RNN OF POPULATION.2',
*' >  # RNN OF POPULATION.1'
 IF(N1.EQ.N2)WRITE(30,*)'---> # RNN OF POPULATION.1',
*' =  # RNN OF POPULATION.2'
 WRITE(30,*)' '
  IF(WR.EQ.1.)WRITE(30,*)'** MIS-CLASSIFICATION!!'
  IF(WR.EQ.0.5)WRITE(30,*)'** ERROR FUNCTION IS HALF!'
  IF(RI.EQ.1.)WRITE(30,*)'** CORRECT CLASSIFICATION!!'
 WRONG=WRONG+WR
 RIGHT=RIGHT+RI
RETURN
END
```

PROGRAM C

It also has two parts as program C1 and program C2. The second part program C2 reads the output from program C1 and calculates average PMC for first stage RNN rule and Sub-Sample procedure simultanously. We also added two more Fortran programs , program P1 and Program P2 to get permutations Of N elements taken 2 and 3 at a time respectively.

PROGRAM P1

```
      PROGRAM  PERM2
       DIMENTION  Z(20)
        OPEN( UNIT=15, FILE='INPUT', STATUS='OLD')
       OPEN( UNIT=16, FILE='OUT2',  STATUS='OLD')
       CLOSE( UNIT=16,  STATUS='DELETE')
        OPEN( UNIT=16, FILE='OUT2', STATUS='NEW')
      DO  10  I=1,10
10       READ ( 15, * )  Z(I)
      DO  20  I=1,9
      DO  20  J=(I+1), 10
       WRITE  (16,13)  Z(I), Z(J)
20     WRITE (16,13)  Z(J), Z(I)
13     FORMAT(1X, 2(F10.6,2X))
      STOP
      END
```

PROGRAM P2

```
      PROGRAM PERM3
      DIMENSION Z(20), P(3)
      OPEN( UNIT=25, FILE='INPUT', STATUS='OLD')
      OPEN( UNIT=26, FILE='OUT3', STATUS='OLD')
      CLOSE( UNIT=26, STATUS='DELETE')
      OPEN( UNIT=26, FILE='OUT3', STATUS='NEW')
      DO 10 I=1,10
10    READ (25,*) Z(I)
      DO 20 I=1, 8
      DO 20 J=(I+1), 9
      DO 20 K=(J+1), 10
      WRITE (26,30) Z(I), Z(J), Z(K)
      P(1)=Z(I)
      P(2)=Z(J)
      P(3)=Z(K)
      M=1
      DO 17 L=1,3
      IF ( M.LE.5 ) THEN
         CALL INT12(P,3)
         WRITE (26,30) ( P(I1), I1=1,3)
         M=M+1
         IF ( M.LE.5 ) THEN
           CALL INT23(P,3)
           WRITE (26,30) (P(I1), I1=1,3)
           M=M+1
         END IF
      END IF
17    CONTINUE
20    CONTINUE
30    FORMAT (1X, 3(F10.6,2X))
      STOP
      END

      SUBROUTINE INT12(R,3)
      DIMENTION R(3), T(1)
      T(1)=R(1)
      R(1)=R(2)
      R(2)=T(1)
      END

      SUBROUTINE INT23(R,3)
      DIMENTION R(3), T(1)
      T(1)=R(2)
      R(2)=R(3)
      R(3)=T(1)
      END
```

PROGRAM C1

```
Read 'OUTPX3' C1 C2 C3
Read 'OUTPY3' C4 C5 C6
Read 'INPUTX' C13
Read 'INPUTY' C14
LET K4=100
NAME C34='Cover.1', C36='Cover.2', C22='Order.1', C24='Order.2'
STORE 'MRN.3'
NOECHO
Random 1 C7 C8 C9;
NORMAL mean=0, s.d.=1.
Random 1 C10 C11 C12;
NORMAL mean=2, s.d.=1.
STACK C7 C8 C9 into C15
STACK C10 C11 C12 into C16
STACK C13 C14 C15 into C17;
Subscripts C18.
STACK C13 C14 C16 into C19;
Subscripts C20.
SORT C17 carry C18 put into C21 C22
SORT C19 carry C20 put into C23 C24
LET C25=SQRT((C1-C7)**2 + (C2-C8)**2 + (C3-C9)**2)
LET C26=SQRT((C4-C7)**2 + (C5-C8)**2 + (C6-C9)**2)
LET C27=SQRT((C1-C10)**2 + (C2-C11)**2 + (C3-C12)**2)
LET C28=SQRT((C4-C10)**2 + (C5-C11)**2 + (C6-C12)**2)
STACK C25 C26 put into C29;
Subscripts C30.
STACK C27 C28 put into C31;
Subscripts C32.
SORT C29 carry C30 put into C33 C34
SORT C31 carry C32 put into C35 C36
DELETE 4:1440 C34
DELETE 4:1440 C36
PRINT C22
PRINT C24
PRINT C34
PRINT C36
END
EXECUTE 'MRN.3' K4
STOP
```

PROGRAM C2

```
*****************************************************
*
*
*    PROGRAM TPMC
*
* THIS PROGRAM CALCULATES AVERAGE PMC OF
* 1ST STAGE RNN RULE & SUB-SAMPLE PROCEDURE
* ( K=1,2,3 ) SIMULTANOUSLY
*****************************************************


*****************************************************
*********
*********        MAIN  PROGRAM
*********
*****************************************************


      CHARACTER*80 A,B*405
       OPEN(UNIT=20,FILE='SUBHAS',STATUS='OLD')
       OPEN(UNIT=30,FILE='SCAN',STATUS='OLD')
      CLOSE(UNIT=30,STATUS='DELETE')
       OPEN(UNIT=30,FILE='SCAN',STATUS='NEW')
      IORDER=1
      JORDER=0
      CRT=0
      CWR=0
      WRONG=0.
      RIGHT=0.

*******************************************************
*
* CHANGE THE VALUE OF "M" FOR TESTING DIFFERENT
* NUMBER OF 3S
                    M=1
*******************************************************
     PRINT*,' '
     PRINT*,' '
     PRINT*,' '
     PRINT*,'******************************************'
     PRINT*,'*                                        *'
     PRINT*,'*                                        *'
     PRINT*,'* WELCOME TO "DEB.SCAN", SUBHAS!!!  *'
     PRINT*,'*                                        *'
     PRINT*,'*                                        *'
     PRINT*,'********* PROGRAM RUN STARTS ********'
     PRINT*,' '
100  READ(20,'(A80)',END=500)A
     KC1=INDEX(A,'Cover.1        ')
     KC2=INDEX(A,'Cover.2        ')
     K=INDEX(A,'Order.1   ')
     KK=INDEX(A,'Order.2   ')
```

```
        IF(K.NE.0.OR.KK.NE.0)THEN
              ITYPE=1
              IORDER=1
        ELSE IF(KC1.NE.0.OR.KC2.NE.0)THEN
              ITYPE=0
              JORDER=1
        END IF

        IF(ITYPE.EQ.1.OR.ITYPE.EQ.0)THEN
   WRITE(30,*)'1          ********************************'
   WRITE(30,*)'           *                              *'
   WRITE(30,*)'           *  PROGRAM TPMC OUTPUT         *'
   WRITE(30,*)'           *  ---------------             *'
   WRITE(30,*)'           * CODED BY: S. BAGUI           *'
   WRITE(30,*)'           * DEPT. OF STATISTICS          *'
   WRITE(30,*)'           * UNIV. OF ALBERTA             *'
   WRITE(30,*)'           * DATE   : 89/05/15            *'
   WRITE(30,*)'           * CODE   : FORTRAN             *'
   WRITE(30,*)'           ********************************'
   WRITE(30,*)' '
   WRITE(30,*)' '
   WRITE(30,*)' **************************************'
   WRITE(30,*)'  *** PROGRAM OUTPUT STARTS ***'
   WRITE(30,*)' **************************************'
   WRITE(30,*)' '
   WRITE(30,*)A

   IF(ITYPE.EQ.1)THEN
   PRINT*,'-> NOW PROCESSING (DASGUPTA) SAMPLE NUMBER:',
  * IORDER
   WRITE(30,*)' (DASGUPTA) SAMPLE NUMBER: ',IORDER
  END IF

   IF(ITYPE.EQ.0)THEN
   WRITE(30,*)' (COVER) SAMPLE NUMBER: ',JORDER
   PRINT*,'-> NOW PROCESSING (COVER) SAMPLE NUMBER:',
  * JORDER
  END IF

  WRITE(30,*)' '
   READ(A(8:8),'(I1)')NORDER
  WRITE(30,*)' '
        GO TO 110

        END IF
  GO TO 100


110 DO 200 I=1,10000
   CALL READER(A,NORDER,NORD,B,ISTART,RIGHT,
  *WRONG,IORDER,M,CRT,CWR,JORDER,ITYPE,ITP)
   IF(ITYPE.EQ.1)CALL SEE(B,ISTART,NORDER,RIGHT,
  * WRONG, M)

        ITYPE=ITP

   IF(ITYPE.EQ.1)IORDER = IORDER + 1
   IF(ITYPE.EQ.0)JORDER=JORDER+1
```

```
      WRITE(30,*)' '
      WRITE(30,*)'****************************************'
      WRITE(30,*)' '
      WRITE(30,*)A

      IF(ITYPE.EQ.1)THEN
      PRINT*,'-> NOW PROCESSING (DASGUPTA) SAMPLE NUMBER:',
     * IORDER
      WRITE(30,*)' (DASGUPTA) SAMPLE NUMBER: ',IORDER
      END IF

      IF(ITYPE.EQ.0)THEN
      WRITE(30,*)' (COVER) SAMPLE NUMBER: ',JORDER
      PRINT*,'-> NOW PROCESSING (COVER) SAMPLE NUMBER :',
     * JORDER
      END IF

      WRITE(30,*)' '
      NORDER = NORD
 200  CONTINUE
 500  STOP
      END


***********************************************************
********
********    READER
********
***********************************************************


      SUBROUTINE READER(A,NORDER,NORD,B,ISTART,RIGHT,
     *WRONG,IORDER,M,CRT,CWR,JORDER,ITYPE,ITP)
      INTEGER NUM(50)
      CHARACTER*(*)A,B
      ISTART=0

 100  READ(20,'(A80)',END=555)A
      K=INDEX(A,'Order.1   ')
      KK=INDEX(A,'Order.2   ')
      KC1=INDEX(A,'Cover.1        ')
      KC2=INDEX(A,'Cover.2        ')
      J=INDEX(A,'*** Minitab')

      IF(J.NE.0.)THEN
 555  IF(ITYPE.EQ.1)CALL SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
      WRITE(30,*)' '
      WRITE(30,*)'*******************************************'
      WRITE(30,*)' '
      WRITE(30,*)'                       SUMMARY: '
      WRITE(30,*)' '
      WRITE(30,*)'  NOTE: '
      WRITE(30,*)' #1. THE POPULATION DATA WAS READ FROM ',
     *'FILE: SUBHAS'
      WRITE(30,*)' #2. THE NUMBER OF 3s IN EACH POPULATION ',
     *'IS: ',M
      WRITE(30,44)
```

```
44  FORMAT('            #3. IF THE NUMBER OF 3s ARE CHANGED, ',
   *'ALTER THE VALUE OF '/
   *'         "M" IN LINE  32 OF FILE "DEB.SCAN", AND THEN ',
   *'RECOMPILE '/
   *'                DEB.SCAN, BEFORE RUNNING BATCHFILE.')
    WRITE(30,*)' '
    WRITE(30,*)'                  ------------- '
    WRITE(30,*)' '
    WRITE(30,*)' '
    WRITE(30,*)'   FIRST- STAGE RNN --  SAMPLE ANALYSIS'
    WRITE(30,*)'   --------------------- '
    WRITE(30,*)' '
    WRITE(30,*)'TOTAL NUMBER OF  SAMPLES ARE : ',IORDER
    WRITE(30,*)'TOTAL NUMBER OF CORRECT CLASSIFICATIONS  ARE',
   *' : ',RIGHT
    WRITE(30,*)'TOTAL NUMBER OF WRONG CLASSIFICATIONS ARE  ',
   *' : ',WRONG
    WRITE(30,*)' '
    WRITE(30,*)'  TOTAL PROBABILITY OF ',
   *'MIS-CLASSIFICATION IS    : ',WRONG/IORDER
    WRITE(30,*)'   TOTAL PROBABILITY OF ',
   *'CORRECT CLASSIFICATION IS : ',RIGHT/IORDER
    WRITE (30,*)' '
    WRITE(30,*)' '
    WRITE(30,*)' '
    WRITE(30,*)'   SUB - SAMPLE ANALYSIS'
    WRITE(30,*)'   ------------------ '
    WRITE(30,*)' '
    WRITE(30,*)'TOTAL NUMBER OF  SAMPLES ARE : ',JORDER
    WRITE(30,*)'TOTAL NUMBER OF CORRECT CLASSIFICATIONS  ARE',
   *' : ',CRT
    WRITE(30,*)'TOTAL NUMBER OF WRONG CLASSIFICATIONS ARE  ',
   *' : ',CWR
    WRITE(30,*)' '
    WRITE(30,*)'   TOTAL PROBABILITY OF ',
   *'MIS-CLASSIFICATION IS    : ',CWR/JORDER
    WRITE(30,*)'   TOTAL PROBABILITY OF ',
   *'CORRECT CLASSIFICATION IS : ',CRT/JORDER
    WRITE (30,*)' '
    WRITE(30,*)'**********************************************'
    WRITE(30,*)'** MRN.SUB  N(0,1) N(2,1) N=10  R(1) XY.F **'
    WRITE(30,*)'**********************************************'
    PRINT*,' '
    PRINT*,'*************** PROGRAM RUN ENDS ***************'
    PRINT*,'*                                              *'
    PRINT*,'*        RESULTS ARE IN FILE "SCAN"            *'
    PRINT*,'*                                              *'
    PRINT*,'************************************************'
500 STOP
    END IF


        IF(K.NE.0.OR.KK.NE.0.OR.KC1.NE.0.OR.KC2.NE.0)THEN
    IF(K.NE.0.OR.KK.NE.0)ITP=1
    IF(KC1.NE.0.OR.KC2.NE.0)ITP=0
     READ(A(8:8),'(I1)')NORD
        GO TO 103
    END IF
```

```
      IF(ITYPE.EQ.0)THEN
            DO 112 I=1,80
              IF(A(I:I).NE.' ')THEN
                READ(A(I:I),'(I1)')NVAL
              GO TO 113
             END IF
 112        CONTINUE
 113      IF(NVAL.EQ.NORDER)THEN
      WRITE(30,*)' '
      WRITE(30,*)' FIRST ELEMENT OF THE SAMPLE IS : ',NVAL
      WRITE(30,*)' '
      WRITE(30,*)' ***  CORRECT CLASSIFICATION!!!! '
      WRITE(30,*)'  '
           CRT=CRT+1
          ELSE
      WRITE(30,*)' '
      WRITE(30,*)' FIRST ELEMENT OF THE SAMPLE IS : ',NVAL
      WRITE(30,*)' '
      WRITE(30,*)'  ***  MISCLASSIFICATION!!!! '
      WRITE(30,*)'  '
           CWR=CWR+1
         END IF
    ITYPE=2
    END IF


    IF(ITYPE.EQ.1)THEN
    INUM = 0
    DO 101 I=1,80
          IF(A(I:I).NE.' ')THEN
            INUM=INUM+1
             READ(A(I:I),'(I1)')NUM(INUM)
           ISTART=ISTART+1
             WRITE(B(ISTART:ISTART),'(I1)')NUM(INUM)
          END IF
101   CONTINUE
   END IF

   GO TO 100
103   RETURN
   END

*******************************************************
********
********   SEE
********
*******************************************************


    SUBROUTINE SEE(B,ISTART,NORDER,RIGHT,WRONG,M)
    CHARACTER*(*)B,STRING*5,STORE*5
    N1=0
    N2=0
    STRING='33333'
    STORE='00000'

    DO 80 I=6-M,5
```

```
      K5=0
      K1=0

 60   K5=INDEX(B(K1+1:ISTART),STRING(I:))
      K1=K1+K5
           IF(K5.EQ.0)THEN
            GO TO 80
           ELSE
             IF(K1.EQ.1)THEN
                   READ(B(K1+6-I:K1+6-1),'(I1)') IR
                   IF(IR.EQ.1)N1=N1+1*(6-I)
                   IF(IR.EQ.2)N2=N2+1*(6-I)
             ELSE IF(K1.EQ.ISTART-5+I)THEN
                   READ(B(K1-1:K1-1),'(I1)') IL
                   IF(IL.EQ.1)N1=N1+1*(6-I)
                   IF(IL.EQ.2)N2=N2+1*(6-I)
            ELSE
                   READ(B(K1-1:K1-1),'(I1)')IL
                   IF(IL.EQ.1)N1=N1+1*(6-I)
                   IF(IL.EQ.2)N2=N2+1*(6-I)
                   READ(B(K1+6-I:K1+6-I),'(I1)')IR
                   IF(IR.EQ.1)N1=N1+1*(6-I)
                   IF(IR.EQ.2)N2=N2+1*(6-I)
            END IF
                   B(K1:K1+5-I)=STORE(I:)

                  GO TO 60
          ENDIF

80   CONTINUE

     WRITE(30,*)'  # RNN FROM POPULATION.1 IS : ',N1
     WRITE(30,*)'  # RNN FROM POPULATION.2 IS : ',N2

     WR=0.
     RI=0.


     IF(NORDER.EQ.1)THEN
             IF(N1.LT.N2)WR=1.
             IF(N1.EQ.N2)WR=0.5
             IF(N1.EQ.N2)RI=0.5
             IF(N1.GT.N2)RI=1.
          ELSE
             IF(N2.LT.N1)WR=1.
             IF(N2.EQ.N1)WR=0.5
             IF(N2.EQ.N1)RI=0.5
             IF(N2.GT.N1)RI=1.
          END IF

    WRITE(30,*)' '
    IF(N1.GT.N2)WRITE(30,*)'----> # RNN OF POPULATION.1 ',
   *' >  # RNN OF POPULATION.2'
    IF(N2.GT.N1)WRITE(30,*)'----> # RNN OF POPULATION.2 ',
   *'  >  # RNN OF POPULATION.1'
    IF(N1.EQ.N2)WRITE(30,*)'----> # RNN OF POPULATION.1 ',
   *'  =  # RNN OF POPULATION.2'
    WRITE(30,*)' '
```

```
 IF(WR.EQ.1.)WRITE(30,*)'** MIS-CLASSIFICATION!!!!!'
 IF(WR.EQ.0.5)WRITE(30,*)'** ERROR FUNCTION IS HALF!'
 IF(RI.EQ.1.)WRITE(30,*)'** CORRECT CLASSIFICATION!!!'
 WRONG=WRONG+WR
 RIGHT=RIGHT+RI
RETURN
END
```