INVESTIGATIONS OF NONPARAMETRIC DISCRIMINATION THEORY AND PROCEDURES APPLICABLE TO PROBLEMS IN ELECTRONIC COMPONENT SCREENING

Research Sponsored by the
National Aeronautics and Space Administration
Research Grant NsG-562

Final Report

ON DISCRIMINATION PROBLEMS

by

C. P. Quesenberry and Margaret Gessaman

Department of Mathematics
Montana State University

Bozeman, Montana

August, 1966

TABLE OF CONTENTS

ON DISCRIMINATION PROBLEMS

CHAPTER I

THE DISCRIMINATION PROBLEM

1.0 Introduction

This paper reports the results of a research effort intended to obtain discrimination procedures which would be useful for the screening of electronic components and which would be valid for more general models than the usual multivariate normal discriminant function techniques frequently used. The basic model used here allows more decisions than the usual "forced classification" model. In that one of the natural ways to judge nonparametric procedures is by comparison with "best" parametric techniques, and since the parametric techniques for the model used here have not been investigated, to the knowledge of this writer, it was felt necessary to present some parametric results, also.

The remainder of this chapter describes the discrimination problem and the model to be used in this paper. Chapter 2 presents a (theoretical) solution for the case of two known distributions along with some particular examples. Chapter 3 introduces definitions of consistency for sequences based on samples.

Chapter 4 proposes a rather general approach for constructing non-parametric discrimination procedures. This approach makes use of work by many writers on nonparametric tolerance regions. Particular procedures are considered which are consistent with best known distribution procedures and some procedures with (mainly) intuitive appeal are suggested.

1.1 The General Problem

The discrimination problem may be stated in its general form as follows. A unit is obtained in some fashion and it is assumed that it arose from some one of k populations. A random variable Z is associated with the unit and if the unit is from population $j(j = 1,\ldots,k)$ the random variable Z has probability distribution P_j on a Euclidean measure space X(G). The problem is to use the value z observed for the random variable Z on the unit obtained to try to decide which population gave rise to the unit in hand.

The usual model for this problem, which may be called the forced classification model, allows k possible decisions. That is, the decision space allowed, Δ_o say, has k elements which are defined as:

$$\delta_j: \text{ decide Z has distribution } P_j;\ j = 1,2,\ldots k. \tag{1.1.1}$$

A discrimination procedure for this model is a decision function d which maps points of the sample space X to points of the decision space Δ_o. The discrimination procedure can be specified as a covering partition set $\{S_1,\ldots,S_k\}$ of subsets of X and the decision function d is given by

$$d(z) = \delta_j \text{ if } z \in S_j,\ j = 1,\ldots,k. \tag{1.1.2}$$

An error will be committed when Z has the distribution P_j and $d(z) = \delta_i$ with $j \neq i$. For a particular value of j there are (k - 1) possible decisions which constitute errors and each of these may, and sometimes should, be considered separately. The probabilities of these errors are of primary

concern to the statistician and he is concerned with methods of controlling them. For a given value of j the probabilities of the various errors may be added to obtain the probability of the compound event (error) that P_j is not correctly identified. Let q_j denote the probability that Z is misclassified when its distribution is actually P_j, i.e.

$$q_j = 1 - \Pr\left\{d(z) \neq \delta_j: \ P_j \text{ is distribution of } Z\right\} \qquad (1.1.3)$$

For the forced classification model with decision space Δ_o whose elements are $\delta_j (j = 1,\ldots,k)$ given by (1.1.1), it is not possible to require that each q_j be no greater than a value α_j, for $\alpha_j (j = 1,\ldots,k)$ a specified constant in the open interval (0,1). In the next section a model will be proposed which allows the possibility of controlling the probabilities of these errors.

The discrimination problem may itself be broken down into subproblems by considering the information that is available about the distributions $P_1,\ldots,P_k$. The information available itself comes in two forms: (1) assumptions about the forms of the distributions, (2) samples from the distributions. Three particular subproblems have been suggested by Fix and Hodges (1951):

(i) The distributions $P_1,\ldots,P_k$ are all completely known.

(ii) The distributions are known except for values of parameters, and samples are available from all distributions.

(iii) The distributions are only known to possess densities and possibly satisfy some types of continuity assumptions. Samples are available from all distributions.

These subproblems clearly do not exhaust all possibilities but do represent important cases. For example, a case that could arise would be for the assumptions of (i) to apply to part of the distributions while those of (iii) apply to the remaining ones. Attention in this paper will be directed primarily toward subproblems (i) and (iii), however the developments for these cases will frequently provide guidance for the treatment of other subproblems.

1.2 A Discrimination Model

In this section a discrimination model is described which is a generalization of the forced classification model mentioned above. The distributional assumptions of the problem remain unchanged, i.e. that a random variable Z has probability distribution P and that P identifies with P_j for some $j = 1,2,\ldots,k$. On the basis of an observation z, a decision is to be made regarding the value of j in $\{1,2,\ldots,k\}$ for which $P = P_j$. The forced classification model assumes the decision space Δ_o with elements δ_j as given by (1.1.1). We shall in this paper assume a larger decision space Δ which has elements

$$\left.\begin{array}{ll} \delta_{j_1\cdots j_s}: & \text{decide that } P \in \{P_{i_1},\ldots,P_{i_s}\} \text{ for } s \in \{1,\ldots,k-1\} \\ \delta_o: & \text{reserve judgment} \end{array}\right\} \quad (1.2.1)$$

where $(i_1,\ldots,i_s)$ is a subset of s distinct elements of $\{1,\ldots,k\}$. By "reserve judgment" it is meant that no decision whatever is to be made concerning the distribution P of Z. Observe that $\Delta_o \subset \Delta$.

The points added to Δ_o to obtain Δ represent partial decisions which are often realistic in practice. For example, if k = 3 and two of the distributions were quite similar but the other was wildly different it would be possible to conclude that Z was distributed according to one of the first two and eliminate the last from consideration. The inclusion of a reserve judgement decision will allow one to provide a positive control of the probabilities q_j of (1.1.3). (The definition is still valid since $\Delta_o \subset \Delta$.)

Let d denote a decision function which maps the sample space X to Δ. An error will be made when $P = P_j$ and $d(z) = \delta_{i_1 \ldots j_s}$ with $j \notin \{i_1,\ldots,i_s\}$. For $Q_j = \{x: d(x) = \delta_{i_1 \ldots i_s} \text{ with } j \notin \{i_1,\ldots,i_s\}\}$ the probability $q_j = \Pr\{Z \in Q_j : P = P_j\}$.

A discrimination procedure may be specified as a decision function d from X to Δ, or since Δ has $(2^k - 1)$ elements as a partition $\{S_{i_1 \ldots i_s}\}$ of X.

Df. 1.2.1 A discrimination procedure d is said to be of size - $(\alpha_1,\ldots,\alpha_k)$ if

$$q_j \leq \alpha_j$$

for α_j a constant in (0,1) for all j = 1,...,k.

Df. 1.2.2 A discrimination procedure d is said to be of exact size - $(\alpha_1,\ldots,d_k)$ if

$$q_j = \alpha_j \text{ for all } j = 1,\ldots,k.$$

Among the class of decision functions which map X onto Δ and therefore give decision procedures we shall in the succeeding developments of this paper restrict attention to these functions which give size - $(\alpha_1,\ldots,\alpha_k)$ procedures, or, in some cases, exact size - $(\alpha_1,\ldots,\alpha_k)$ procedures. There are usually many procedures which will meet a size requirement, however, and to obtain a unique procedure we shall have to impose further restrictions. When subproblems (i), (ii), and (iii) are considered in subsequent sections of this paper the major problem is to formulate such criteria for particular problems and to apply the criteria to obtain a procedure.

We next set out a construction procedure in rather general terms which gives a size - $(\alpha_1,\ldots\alpha_k)$ procedure.

1.3 A Construction of a Procedure

It was mentioned in the last section that a discrimination procedure is essentially a partition of the sample space X into $(2^k - 1)$ subsets $S_{i_1\ldots i_s}$ corresponding to the decisions of (1.2.1).

If A_j for $j = 1,\ldots,k$, is some measurable subset of X, then, of course, A_j and its complement $\bar{A}_j$ $(= X - A_j)$ constitute a partition of X into two subsets. The following theorem is essentially the same as one given by Lehmann (1957).

Theorem 1.3.1 The relationship

$$S^*_{i_1\ldots i_s} = \bar{A}_{i_1}\ldots\bar{A}_{i_s}A_{i_{s+1}}\ldots A_{i_k} \quad \text{for } s = 0,\ldots,k \tag{1.3.1}$$

defines a 1 - 1 correspondence between the disjoint covering classes of sets $\left\{ S^*_{i_1 \ldots i_s} \right\}$ and classes of subsets $A_1, \ldots, A_k$ of X.

Pf. (See Lehmann, 1957, p. 5)

We now define a class of sets $\left\{ S_{i_1 \ldots i_s} \right\}$:

$$\left.\begin{aligned} S_{i_1 \ldots i_s} &= S^*_{i_1 \ldots i_s} \text{ if } s = 1, \ldots, k-1 \\ S_o &= (A_1 \ldots A_k) \cup (\bar{A}_1 \ldots \bar{A}_k) \end{aligned}\right\} \tag{1.3.2}$$

Then the class of S-sets is obviously uniquely determined by the sets $\left\{ A_1, \ldots, A_k \right\}$, however more than one class of sets $\left\{ A_1, \ldots, A_k \right\}$ may give the same class of S-sets.

The class of S-sets constitutes a covering partition of X and we can use it for a discrimination procedure. Define the procedure:

$$d(z) = \begin{cases} \delta_{i_1 \ldots i_s} & \text{if } z \in S_{i_1 \ldots i_s} \\ \delta_o & \text{if } z \in S_o \end{cases} \tag{1.3.3}$$

Theorem 1.3.2 If $P_j(Z_j \in A_j) \leq \alpha_j$ for all $j = 1, \ldots, k$, then the discrimination procedure given by (1.3.3) is size - $(\alpha_1, \ldots, \alpha_k)$.

Pf. Suppose $P = P_1$.

Then an error will be made when z falls in the set

$$B_1 = A_1 \cap \left(\bigcup_{s=1}^{k-1} \left(\bigcup \bar{A}_{i_1} \ldots \bar{A}_{i_s} A_{i_{s+1}} \ldots A_{i_{k-1}} \right) \right), \tag{1.3.4}$$

where the second union is over all combinations $(i_1,\ldots,i_s)$ of size s that can be taken from the integers $\{2,\ldots,k\}$, and $\{i_{s+1},\ldots,i_{k-1}\}$ is the remaining set.

Then $$B_1 \subset A_1,$$

and $$P_1(B_1) \leq P_1(A_1) \leq \alpha_1.$$

For any $j \in \{2,\ldots,k\}$, by symmetry

$$P(B_j) \leq \alpha_j$$

i.e. the procedure is size - $(\alpha_1,\ldots,\alpha_k)$. Observe that the set B_1 of (1.3.4) can be written

$$B_1 = A_1 \cap (X - A_2 \ldots A_k)$$

$$= A_1 \cap (\bigcup_2^k \bar{A}_j) \quad .$$

When the sets A_j $(j = 1,\ldots,k)$ can be obtained such that $P_j\{A_j\}$ is bounded by α_j, this theorem shows that the procedure defined by (1.3.2) and (1.3.1) from the A sets is a size - $(\alpha_1,\ldots,\alpha_k)$ procedure. The A sets can of course be chosen in many different ways and would determine many procedures of the specified size. From the class of all size - $(\alpha_1,\ldots,\alpha_k)$ procedures we wish to choose one that will in some sense be optimal, or have, at least, some desirable characteristics. In the next chapter we shall study the problem in considerable detail for subproblem (i) when there are two categories (k = 2). This will provide a standard for comparison for procedures for the other subproblems when k = 2.

CHAPTER II

KNOWN CONTINUOUS DISTRIBUTIONS

2.1 Introduction

In this chapter we consider subproblem (i), i.e. the case for all of the distributions completely known. We further assume that the distributions are all absolutely continuous with respect to Lebesque measure which implies existence of densities f_j for the distributions $P_j (j = 1,\ldots,k)$. A rather complete theoretical solution is given for the case $k = 2$.

2.2 Two Distributions

With two distributions ($k = 2$) the problem simplifies as follows. The decision space Δ has three elements:

d_1: classify Z as a P_1 random variable,

d_2: classify Z as a P_2 random variable,

d_o: reserve judgment.

A discrimination procedure is a function d which maps the sample space X to Δ and since Δ contains three elements it may also be specified by sets S_1, S_2, and S_o where

$$S_j = \left\{ z: d(z) = \delta_j \right\}, \quad j = 1,2,0. \tag{2.2.1}$$

We will sometimes denote the procedure by $d(S_1,S_2)$, since S_1 and S_2 determine a procedure and conversely.

Let π_1 and π_2 denote the a priori probabilities that Z has the

distributions P_1 and P_2, respectively. Assume that π_1 and π_2 are both nonzero, for otherwise there is no discrimination problem. There are then only two ways for errors to arise and these are for z to fall in S_1 when $P = P_2$, and for z to fall in S_2 when $P = P_1$. The probabilities of these errors are $P_2(S_1)$ and $P_1(S_2)$, respectively.

We have defined size for discrimination procedures, Df. 1.2.1, and for this special case the definition says that for given $\alpha_j \in (0,1)$, $j = 1,2$, a procedure is of size - (α_1,α_2) if $P_2(S_1) \leq \alpha_2$ and $P_1(S_2) \leq \alpha_1$, and of exact size - (α_1,α_2) if both equalities hold. Now, the overall probability of an error of the first type mentioned above is $\pi_2 P_2(S_1)$ and of the other is $\pi_1 P_1(S_2)$. Then the overall probability of any error is $\pi_1 P_1(S_2) + \pi_2 P_2(S_2)$. Also, the overall probability of a reserve judgment conclusion is $\pi_1 P_1(S_o) + \pi_2 P_2(S_o)$. Then since $\{S_1,S_2,S_o\}$ is a disjoint covering of the sample space X, we have it that $P_j(S_1) + P_j(S_2) + P_j(S_o) = 1$ for $j = 1,2$. We define the power of a size - (α_1,α_2) discrimination procedure as the overall probability that a reserve judgment conclusion is not reached.

Df. 2.2.1 The power of a discrimination procedure $d(S_1,S_2)$ is the overall probability that a reserve judgment conclusion is not reached, i.e.

$$Q(S_o,\pi_1) = \pi_1(P_1(S_1) + P_1(S_2)) + \pi_2(P_2(S_1) + P_2(S_2)). \qquad (2.2.2)$$

Immediately

$$Q(S_o,\pi_1) = 1 - \pi_1 P_1(S_o) - \pi_2 P_2(S_o).$$

Df. 2.2.2 A size - (α_1,α_2) discrimination procedure $d(S_1,S_2)$ is said to be more powerful than a size - (α_1,α_2) discrimination procedure $d^*(S_1^*,S_2^*)$ if

$$Q(S_o,\pi_1) > Q(S_o^*, \pi_1).$$

Df. 2.2.3 A size - (α_1,α_2) discrimination procedure $d(S_1,S_2)$ is said to be a most powerful size - (α_1,α_2) discrimination procedure if there exists no other size - (α_1,α_2) procedure that is more powerful than $d(S_1,S_2)$.

From (1.3.1) and (1.3.2) the sets S_1,S_2 and S_o are related to partitions $\{A_1,\bar{A}_1\}$ and $\{A_2,\bar{A}_2\}$ by

$$\left.\begin{aligned} S_1 &= \bar{A}_1 A_2 \\ S_2 &= A_1 \bar{A}_2 \\ S_o &= A_1 A_2 \cup \bar{A}_1 \bar{A}_2 \end{aligned}\right\} \qquad (2.2.3)$$

For A_1 and A_2 subsets of X with $P_1(A_1) \leq \alpha_1$ and $P_2(A_2) \leq \alpha_2$, the procedure $d(S_1,S_2)$ is a size - (α_1,α_2) procedure from Theorem (1.3.2).

We wish to choose A_1 and A_2, if possible, in such a way as to make $d(S_1,S_2)$ a most powerful procedure. The next theorem provides the solution to this problem. A lemma whose assertions are used in the proof of the theorem is presented first. Proofs of its parts follow directly from the definitions of size and power and (2.2.3).

Lemma 2.2.1 Let $d(S_1,S_2)$ be a size - (α_1,α_2) discrimination procedure for $\alpha_j \in (0,1)$; $j = 1,2$.

(a) $d(S_1,S_2)$ has power one for discrimination if and only if $P_1(S_o) = P_2(S_o) = 0$.

(b) A sufficient condition that $d(S_1,S_2)$ be a most powerful size - (α_1,α_2) discrimination procedure is that for $d^*(S_1^*, S_2^*)$ any size - (α_1,α_2) discrimination procedure, $P_1(S_o) \leq P_1(S_o^*)$ and $P_2(S_o) \leq P_2(S_o^*)$.

(c) If there exists a size - (α_1,α_2) discrimination procedure $d^*(S_1^*, S_2^*)$ such that $P_1(S_o) \geq P_1(S_o^*)$ and $P_2(S_o) \geq P_2(S_o^*)$ where at least one inequality is strict, then $d(S_1,S_2)$ is not a size -(α_1,α_2) most powerful discrimination procedure.

We now give the main theorem of this chapter.

Theorem 2.2.1 Let P_1 and P_2 be distinct probability distributions defined on k-dimensional Euclidean space, R^k, with probability density functions f_1 and f_2, respectively, with respect to Lebesque measure λ on R^k. Also, assume the random variable $W = f_1(X)/f_2(X)$ is a continuous random variable defined a.e.λ.

(i) For $\alpha_j \in (0,1)$, $j = 1,2$, there exist positive constants C_1 and C_2 and sets A_1 and A_2 which define a discrimination procedure as in (2.2.3) and such that

$$P_1(A_1) = \alpha_1, \; P_2(A_2) = \alpha_2 \qquad (2.2.4)$$

where

$$\left.\begin{aligned} A_1 &= \left\{ x : f_1(x) < C_1 f_2(x) \right\} \\ A_2 &= \left\{ x : f_1(x) > C_2 f_2(x) \right\} \end{aligned}\right\} \qquad (2.2.5)$$

(ii) Let $d(S_1,S_2)$ be a discrimination procedure constructed as in (2.2.5) from sets A_1 and A_2 satisfying (2.2.4) and (2.2.5).

(a) If $C_1 \leq C_2$, then $d(S_1,S_2)$ is a most powerful size - (α_1,α_2) discrimination procedure. It is exact size - (α_1,α_2).

(b) If $C_1 > C_2$, then $d(S_1,S_2)$ is not a most powerful size - (α_1,α_2) discrimination procedure.

(iii) Let C_1 and C_2 be the constants in (i).

(a) Let $C_1 \leq C_2$ and suppose $d^*(S_1^*,S_2^*)$ is a most powerful size - (α_1,α_2) discrimination procedure. Then $d^*(S_1^*,S_2^*)$ is exact size - (α_1,α_2) and $\lambda(S_1^* \cup S_1) = \lambda(S_1^* S_1)$ and $\lambda(S_2^* \cup S_2) = \lambda(S_2^* S_2)$.

(b) If $C_1 > C_2$, then there exists size - (α_1,α_2) discrimination procedures with power one. Such a procedure can be constructed as in (2.2.3) with sets A_1 and A_2 as in (2.2.5) with the C's there replaced by a common value C^* in the interval $\left[C_2,C_1\right]$.

Proof:

(i) This is shown in the proof of the Neyman-Pearson Fundamental Lemma given by Lehmann (1959).

(ii) (a) For $C_1 \leq C_2$

1) A_1A_2 is empty and $S_o = \bar{A}_1\bar{A}_2$,

2) $S_2 = A_1\bar{A}_2 = A_1$ and $P_1(S_2) = P_1(A_1) = \alpha_1$,

3) $S_1 = \bar{A}_1A_2 = A_2$ and $P_2(S_1) = P_2(A_2) = \alpha_2$.

Let $d^*(S_1^*,S_2^*)$ be any size - (α_1,α_2) discrimination procedure. In order to show that $d(S_1,S_2)$ is a most powerful exact size - (α_1,α_2) procedure, by Lemma (2.2.1) (b) it suffices to show that

$$P_1(S_o) \leq P_1(S_o^*) \text{ and } P_2(S_o) \leq P_2(S_o^*). \tag{2.2.6}$$

Now $$P_1(S_1) + P_1(S_2) + P_1(S_o) = P_1(S_1^*) + P_1(S_2^*) + P_1(S_o^*) = 1,$$

and $$P_1(S_2) = \alpha_1 \geq P_1(S_2^*).$$

So $$P_1(S_1) + P_1(S_o) \leq P_1(S_1^*) + P_1(S_o^*). \tag{2.2.7}$$

We now show that

$$P_1(S_1^*) \leq P_1(S_1), \tag{2.2.8}$$

and the first inequality of (2.2.6) will follow. To show (2.2.7) let $\Psi(x)$ be the characteristic function of the set $S_1 = A_2$ and let $\phi(x)$ be the characteristic function of the set S_1^*. Put

$$S^+ = \{x: \Psi(x) - \phi(x) > 0\}.$$

$$S^- = \{x: \Psi(x) - \phi(x) < 0\}.$$

If $x \in S^+$, then $\psi(x) = 1$ and $f_1(x) - C_2 f_2(x) > 0$. If $x \in S^-$, then $\psi(x) = 0$ and $f_1(x) - C_2 f_2(x) < 0$. In each case $(\psi(x) - \phi(x))(f_1(x) - C_2 f_2(x)) \geq 0$.

Consider the integral

$$\int_X (\psi - \phi)(f_1 - C_2 f_2) d\lambda = \int_{S^+ \cup S^-} (\psi - \phi)(f_1 - C_2 f_2) d\lambda \geq 0. \quad (2.2.9)$$

But

$$\int_X (\psi - \phi)(f_1 - C_2 f_2) d\lambda = P_1(S_1) - P_1(S_1^*) - C_2(P_2(S_1) - P_2(S_1^*)),$$

and

$$P_1(S_1 - P_1(S_1^*) \geq C_2(P_2(S_1) - P_2(S_1^*)) \geq 0,$$

since

$$P_2(S_1) = \alpha_2 \text{ and } P_2(S_1^*) \leq \alpha_2.$$

Therefore (2.2.7) holds and the first inequality of (2.2.6) is established. The second inequality of (2.2.6) follows by symmetry.

(b) For $C_1 > C_2$, this will follow from (iii)(b).

(iii) (a) Let $C_1 \leq C_2$ and suppose $d^*(S_1^*, S_2^*)$ is a most powerful size - (α_1, α_2) procedure. Again denote $d(S_1, S_2)$ the procedure determined by (2.2.4), (2.2.5) and (2.2.3). Statements 1), 2), and 3) of (ii)(a) still hold. Let S^+ and S^- be defined as in (ii)(a) and put

$$S' = \{x : f_1(c) - C_2 f_2(x) = 0\}$$

$$S = (S^+ \cup S^-) \cap S'$$

Then

$$(\psi(x) - \phi(x))(f_1(x) - C_2 f_2(x)) > 0 \text{ for } x \in S,$$

and $$\int_X (\psi - \phi)(f_1 - C_2 f_2)\, d\lambda = \int_S (\psi - \phi)(f_1 - C_2 f_2)\, d\lambda$$

$$= P_1(S_1) - P_1(S_1^*) - C_2(P_2(S_1) - P_2(S_1^*))$$

$$> 0,$$

unless $\lambda(S) = 0$. If $\lambda(S) > 0$, then $P_1(S_1) > P_1(S_1^*)$ and from (2.2.7) $P_1(S_o) < P_1(S_o^*)$. From the proof of (ii)(a) it is known that $P_2(S_o) \leq P_2(S_o^*)$. By Lemma 2.2.1(c) $d^*(S_1^*, S_2^*)$ is not most powerful, a contradiction. Thus $\lambda(S) = \lambda(S^+ \cup S^-) = 0$.

But $$\lambda(S_1 \cup S_1^*) = \lambda(S_1 S_1^* \cup S_1 \bar{S}_1^* \cup \bar{S}_1 S_1^*)$$

$$= \lambda(S_1 S_1^*) + \lambda(S_1 \bar{S}_1^* \cup \bar{S}_1 S_1^*)$$

$$= \lambda(S_1 S_1^*) + \lambda(S^+ \cup S^-)$$

$$= \lambda(S_1 S_1^*).$$

This completes (iii)(a).

It may also be observed that

$$\alpha_1 = P_1(A_1) = P_2(S_2) = P_1(S_2^*) \leq \alpha_1,$$

$$\alpha_2 = P_2(A_2) = P_2(S_1) = P_2(S_1^*) \leq \alpha_2,$$

i.e. that $d^*(S_1^*, S_2^*)$ is exact size - (α_1, α_2).

(iii)(b) For $C_1 > C_2$, let α^* be a point in the closed interval $\left[C_2, C_1\right]$, and

$$A_1^* = \left\{ x : f_1(x) < C^* f_2(x) \right\},$$
$$A_2^* = \left\{ x : f_1(x) > C^* f_2(x) \right\}.$$

Then (2.2.3), gives

$$S_1^* = \bar{A}_1^* A_2^* = A_2^*,$$
$$S_2^* = A_1^* \bar{A}_2^* = A_1^*.$$

But $C^* \leq C_1$, and $A_1^* \subset A_1$. Also, $C^* \geq C_2$, and $A_2^* \subset A_2$. Therefore the procedure $d^*(S_1^*, S_2^*)$ is size - (α_1, α_2). Also, $\lambda(S_o^*) = 0$ and the power of $d^*(S_1^*, S_2^*)$ is therefore one.

This theorem gives a rather complete theoretical solution to the discrimination problem as formulated in this paper for the two category continuous distributions case. Applications of the theorem to particular cases (distributions) may present practical problems. We consider applying the theorem to a few particular cases. The examples chosen are felt to be of some importance in their own right, and in any case they should help to clarify the ideas involved in the model and the theorem. Some of these procedures will later be used for standards for comparison for nonparametric procedures.

2.3 Examples

Example 2.3.1 Families with Monotone Likelihood Ratios

For θ a real-valued parameter, let $\{P_\theta : \theta \in \Omega\}$ be a family of absolutely continuous probability distributions with strictly monotone (increasing) likelihood ratio in a real-valued continuous statistic $T(x)$, i.e. for $\theta_1 < \theta_2$, $f_{\theta_2}(x)/f_{\theta_1}(x)$ is a strictly monotone (increasing) function of the random variable $T(x)$. Put $f_{\theta_j} = f_j$, $j = 1,2$. Assume that the ratio $f_1(x)/f_2(x)$ satisfies the hypotheses of Theorem 2.2.1, and let C_1 and C_2 be the constants defined by (2.2.4) and (2.2.5) for given $\alpha_j \in (0,1)$, $j = 1,2$. Since $f_1(x)/f_2(x)$ is strictly increasing in $T(x)$, the sets A_1 and A_2 can be obtained directly from $T(x)$. Then $f_1(x)/f_2(x) < C_1$ if and only if $T(x) < C_1^*$, for C_1^* a unique real number. We have it that $A_1 = \{x : T(x) < C_1^*\}$ and similarly there is a C_2^* such that $A_2 = \{x : T(x) < C_2^*\}$. If P_j^T denotes the distribution induced on the space of T from P_j, $j = 1,2$, then

$$P_1^T\{t : t < C_1^*\} = \alpha_1 \text{ and } P_2^T\{t : t > C_2^*\} = \alpha_2.$$

Then $$S_1 = \{t : t > C_2^*\} \text{ and } S_2 = \{t : t < C_1^*\}.$$

Then C_1^* is the α_1th percentile of the P_1^T distribution and C_2^* is the $(1 - \alpha_2)$nd percentile of the P_2^T distribution. Denote $C_1^* = t^1_{\alpha_1}$ and $C_2^* = t^2_{1-\alpha_2}$. If $C_1 > C_2$, or equivalently, $t^1_{\alpha_1} > t^2_{1-\alpha_2}$, there are

size - (α_1,α_2) discrimination procedures with power 1. Such a procedure can be constructed as follows. Let $t^* \in \left[t^2_{1-\alpha_2}, t^1_{\alpha_1} \right]$ and put $S_1^* = \left\{ t:t > t^* \right\}$, $S_2^* = \left\{ t:t < t^* \right\}$. Then $S_o = \left\{ t:t = t^* \right\}$ and $P_j^T(S_o) = 0$, $j = 1,2$. Therefore, the power of this procedure is one. Also,

$$P_1^T(S_2^*) \leq P_1^T \left\{ t:t < t^1_{\alpha_1} \right\} = \alpha_1 \text{ and } P_2^T(S_1^*) \leq P_2^T \left\{ t:t > t^2_{1-\alpha_2} \right\} = \alpha_2 \text{ and}$$

the procedure is size - (α_1,α_2). The choice of the value t^* in the interval $\left[t^2_{1-\alpha_2}, t^1_{\alpha_1} \right]$ is an interesting problem in its own right. We will consider one choice in Appendix I. The next example considers a special case of the model in this example.

Example 2.3.2 Particular Univariate Normals

(i) Let P_1 and P_2 be univariate normal distributions with common variance 1, and means 2 and 0, respectively. The density ratio is

$$f_1(x)/f_2(x) = \exp \left\{ -\frac{1}{2} \left((x - 2)^2 - x^2 \right) \right\} = \exp (2x - 2),$$

which is strictly increasing in $T(x) = x$. We can then define A_1 and A_2 in terms of x. Put $A_1 = \left\{ x:x < C_1 \right\}$ and $A_2 = \left\{ x:x > C_2 \right\}$. If $\alpha_1 = \alpha_2 = 0.1$, then $C_1 = 0.718$ and $C_2 = 1.282$. Here $C_1 < C_2$ and the procedure is given by

$$S_1 = \left\{ x:x > 1.282 \right\}$$

$$S_2 = \left\{ x:x < 0.718 \right\}$$

$$S_o = \left\{ x:0.718 \leq x \leq 1.282 \right\}.$$

In other words,

(1) If $z > 1.282$, classify Z as a P_1 random variable,

(2) If $z < 0.718$, classify Z as a P_2 random variable,

(3) If $z \in [0.718, 1.282]$, reserve judgment, i.e. do not classify Z.

The actual power for discrimination can be evaluated in this case and is

$$Q(S_o,\pi_1) = (0.85)\pi_1 + (0,86)\pi_2 = 0.86$$

for any $\pi_1 \in (0,1)$. This is, of course, the maximum possible power for a size - (.1,.1) procedure.

(ii) The model is the same as in (i) except that we let the means of the probability distributions P_1 and P_2 be 3 and 0, respectively. Again, taking $(\alpha_1,\alpha_2) = (.1,.1)$, we obtain the procedure $d(S_1,S_2)$ with

$$S_1 = \{x:x > 1.718\},$$

$$S_2 = \{x:x < 1.282\},$$

$$S_o = \{x:1.282 \leq x \leq 1.718\}.$$

Here $C_1 = 1.718$ and $C_2 = 1.282$, and $C_1 > C_2$. This is the case of (iii) (b) of Theorem (2.2.1). The actual power for discrimination of $d(S_1,S_2)$ is $1 - (.057)\pi_1 + (.057)\pi_2 = 1 - (.057) = .943$. We can define another procedure

$$S_1^* = \{x:x > 1.5\},$$

$$S_2^* = \{x:x < 1.5\},$$

$$S_o^* = \{x:x = 1.5\}.$$

This is a size - (.1,.1)(exact size - (.067,.067)) procedure with discrimination power 1. It is the most powerful exact size - (.067,.067) procedure since it satisfies (ii)(a) of Theorem (2.2.1) for $(\alpha_1,\alpha_2) = (.067,.067)$.

Example 2.3.3 Univariate Normals, General Case

Let $\{P_\theta:\theta \in \Omega\}$ be the family of univariate normal distributions where θ is the vector (μ,σ^2). Then P_j is a $N(\mu_j, \sigma_j^2)$ distribution. The density ratio has the form

$$f_1(x)/f_2(x) = k \exp\{ax^2 + bx + c\}.$$

This ratio is strictly monotone in the exponent $ax^2 + bx + c$, whose graph is a parabola. The ratio is then strictly monotone for x in the interval from $-\infty$ to the point corresponding to the vertex of the parabola and strictly monotone in the opposite sense from this point to $+\infty$. The vertex is at the point $x_o = (\mu_2\sigma_1^2 - \mu_1\sigma_2^2)/(\sigma_1^2 - \sigma_2^2)$. If $a = [(\sigma_1^2 - \sigma_2^2)/2\sigma_1^2\sigma_2^2] < 0$, the ratio increases strictly from $-\infty$ to x_o and decreases strictly from x_o to $+\infty$. Then

$$\bar{A}_1 = \{x:x_o - b_1 \leq x \leq x_o + b_1\}$$

where b_1 is such that $P_1(\bar{A}_1) = 1 - \alpha_1$. Also,

$$A_2 = \{x:x_o - b_2 < x < x_o + b_2\}$$

and b_2 is such that $P_2(A_2) = \alpha_2$. The A sets are determined analogously

if $a > 0$. If $a = 0$, the problem reduces to the preceding example, i.e. $\sigma_1^2 = \sigma_2^2$.

We insert particular numbers to illustrate the computation. Suppose $(\mu_1, \sigma_1^2) = (0,1)$ and $(\mu_2, \sigma_2^2) = (2,25)$. Take $\alpha_1 = \alpha_2 = 0.1$. The density ratio is

$$f_1(x)/f_2(x) = 5 \exp\left\{ - \frac{2}{25} (6x^2 + x - 1 \right\}.$$

This ratio increases from $-\infty$ to $- 0.083$ and decreases strictly from $- 0.083$ to $+ \infty$. The sets A_1 and A_2 are

$$A_1 = \{x:x < - 1.733\} \cup \{x:x > 1.567\},$$

$$A_2 = \{x:- 0.768 < x < 0.602\}.$$

The discrimination procedure is then given by the sets

$$S_1 = \{x:- 0.768 < s < 0.602\}$$

$$S_2 = \{x:x < - 1.733\} \cup \{x:x > 1.567\}$$

$$S_o = \{x:- 1.733 \leq x \leq - 0.768\} \cup \{x:0.602 \leq x \leq 1.567\}.$$

In this case $C_1 < C_2$ and this procedure is the most powerful size - (.1,.1) procedure. It is exact size - (.1,.1).

Example 2.3.4 Multivariate Normals

Let P_1 be a MVN $(\underline{\mu}_1, \Sigma)$ distribution and P_1 be a MVN $(\underline{\mu}_2, \Sigma)$ distribution.

Then the likelihood ratio is monotone in the linear function $T(\underline{x}) = \underline{x}^1 \sum^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)^1 \sum^{-1} (\underline{\mu} - \underline{\mu}_2)$. This problem can then be treated by the approach of Example 2.3.1. The induced distributions of T, P_1^T and P_2^T are univariate normals with common variance. These are the distributions of the linear discriminant analysis for forced classification.

CHAPTER III

DISTRIBUTIONS NOT COMPLETELY KNOWN

3.1 Introduction

If all of the distributions $P_1,\ldots,P_k$ associated with the k populations are not completely known, then a sample $(X_{j1},\ldots,X_{jn_j})$ must be available for every unknown distribution P_j if any discrimination procedure is to be found, in most cases. We assume that none of the distributions are entirely known and that a sample is available from each. Then we seek a procedure based on the statistic $(z; x_{11},\ldots,x_{1n_1};\ldots;x_{k1},\ldots,x_{kn_k})$. Such a procedure then will be a function of the sample sizes, or may be viewed for varying sample sizes as a sequence of discrimination procedures, say $\left\{d_{n_1\ldots n_k}\right\}$. As a means for comparing sequences of procedures we introduce some definitions.

3.2 Consistency of Sequences of Procedures

Let $\left\{d_n\right\}$ and $\left\{d_m^*\right\}$ denote two sequences of decision procedures.

Df. 3.2.1 The sequence of decision procedures $\left\{d_m^*\right\}$ is said to be consistent with the sequence of decision procedures $\left\{d_n\right\}$ if for any $j = 1,\ldots,k$

$$\Pr\left\{d_m^* = d_n:\ P = P_j\right\} \xrightarrow{P} 1, \text{ as } m,n \longrightarrow \infty.$$

We shall usually be interested in comparing a sequence $\left\{d_n\right\}$ with a particular procedure, d_o say, and we will need to show that

$$\Pr\left\{d_n = d_o:\ P_j\right\} \xrightarrow{P} 1 \text{ for all } j = 1,\ldots,k, \text{ as } n \longrightarrow \infty.$$

In searching for a best procedure based on samples from the distributions, the natural limit to consider is often a procedure obtained from known distributions. This is so from the following considerations. Under the assumptions of subproblem (i), the observation z is sufficient for the discrimination problem. But with the distributions incompletely known, the statistic $(z; x_{11},\ldots,x_{1n_1};\ldots;x_{k1},\ldots,x_{kn_k})$ is a sufficient statistic for the discrimination problem. But any procedure that is a function of $(z; x_{11},\ldots,x_{1n_1};\ldots;x_{k1},\ldots,x_{kn_k})$ can be duplicated by a procedure based on z alone by incorporating randomization, if necessary. Therefore if a procedure is best according to some criteria among all procedures based on z, it will also be as good as any procedure based on $(z; x_{11},\ldots,x_{1n_1};\ldots;x_{k1},\ldots,x_{kn_k})$, (cf. Fix and Hodges, 1951). In the subsequent parts of this paper we shall in some of the examples seek procedures valid under the conditions of subproblem (iii), and which will be consistent with the procedures given in the preceding chapter.

CHAPTER IV

A NONPARAMETRIC MODEL

4.1 The Problem and Model

In this chapter we shall consider the discrimination problem under subproblem (iii), i.e. under the assumption that the members of the family of distributions $\{P_\theta : \theta \in \Omega\}$ are all absolutely continuous with respect to Lebesque measure.

The procedures considered here differ from those previously considered in one fundamental property. Those considered previously were deterministic in that the sets $\{S_{i_1 \ldots i_s}\}$ were fixed partitions of the sample space. This is not the case for the procedures of this chapter for the partitions will themselves be random, i.e. depend upon samples, and the probabilities of the various decisions will themselves be random variables. We choose the partitions in such a way that these probabilities have known distributions.

The model assumed is as follows. Let $\{P_\theta : \theta \in \Omega\}$ devote the class of probability measures defined on a Euclidean measure space X(G) for which each P_θ is absolutely continuous with respect to Lebesque measure. Let $\{\theta_1, \ldots, \theta_k\}$ denote a set of k distinct elements of Ω, and suppose that a sample $(X_{j1}, \ldots, x_{jn_j})$ is available from each P_{θ_j}, $j = 1, \ldots, k$. Let Z be a random variable with distribution P_θ and assume that $\theta = \theta_j$ for some $j = 1, \ldots, k$. Then on the basis of an observation z on Z it is required to make a decision as to which values in the set $\{\theta_1, \ldots, \theta_k\}$ that θ might take.

We again use the decision space Δ of section (1.2) which allows partial and reserve judgment decisions.

4.2 Definition of a Procedure

Let α_j $(0,1)$ for each $j = 1,2,\ldots k$ be a specific constant and put $a_j = \left[\alpha_j(n_j + 1)\right]$, i.e. the greatest integer in $\alpha_j(n_j + 1)$. Using the theory of coverages (cf. references in Bibliography), a nonparametric tolerance region containing a_j sample blocks in constructed on the sample space X using the sample $(x_{j1},\ldots,x_{jn_j})$ for each $j = 1,\ldots,k$. These subsets, $A_1,\ldots,A_k$, are then used to define a discrimination procedure $\left\{S_{i_1\ldots i_s}\right\}$ by the relations (1.3.1) and (1.3.2).

The reader should bear in mind that the sets A_j, $j \in \left\{1,\ldots,k\right\}$, are nonparametric tolerance regions and as such their probabilities are random variables. The set A_j contains a_j sample blocks and $P_{\theta_j}(A_j)$ has a beta-distribution with parameters $(a_j,\ n_j - a_j + 1)$ (cf. Wilks, 1963). This distribution has mean

$$\frac{a_j}{n_j + 1} = \frac{\left[\alpha_j(n_j + 1)\right]}{n_j + 1} = \alpha_j - o(1/n_j)$$

with $o(1/n_j) \geq 0$,

and variance

$$\frac{a_j(n_j - a_j + 1)}{(n_j + 1)^2(n_j+2)} = \frac{\left[\alpha_j(n_j + 1)\right]\ (n_j + 1 - \left[\alpha_j(n_j + 1)\right])}{(n_j + 1)^2\ (n_j + 2)}.$$

If $\alpha_j(n_j + 1)$ is an integer, then the mean is α_j and the variance is $\alpha_j(1 - \alpha_j)/(n_j + 2)$. In any case, an application of the Tchebycheff inequality yields

$$P_{\theta_j}(A_j) \xrightarrow{P} \alpha_j \text{ as } n_j \longrightarrow \infty \text{ for all } j = 1,\ldots,k. \qquad (4.2.1)$$

As in section 1.3, an error is made when z falls in the set $A_j \cap \left(\bigcup_{\substack{i=1 \\ i\neq j}}^{k} \bar{A}_i\right)$. The probability of this event is then bounded by the random variable $P_{\theta_j}(A_j)$ with the above beta distribution when P_{θ_j} is the distribution of Z. If it turns out that $A_j \subset \bigcup_{\substack{i=1 \\ i\neq j}}^{k} \bar{A}_i$ with probability 1, then the probability of this error is the random variable $P_{\theta_j}(A_j)$.

The procedure set out in this section provides a method whereby a procedure can be obtained that allows some control, in the above sense, of the probabilities of the errors. This property derives immediately from having chosen the regions A_j to be tolerance regions with a_j blocks as obtained from the jth sample. These regions can still be chosen in many different ways and it would be desirable to choose them in such a manner as

to obtain "good" procedures, in some sense. It does not appear possible to find a rule for choosing regions that will give good procedures for any set of distributions in $\{P_\theta : \theta \in \Omega\}$ of section (4.1). In practice it often occurs that the statistician may suspect that the distributions are of some particular subclass of $\{P_\theta : \theta \in \Omega\}$, and this will provide guidance in choosing the regions to be used. Properties of the subclass may be used to obtain regions with good properties if the distributions are from the subclass, and even if they are not the procedure would have the size properties discussed above.

4.3 The Control of Size

The procedure set out in the last section has the property that when Z has distribution P_1 the probability that a mistake will be made is bounded by the random variable $P_1(A_1)$ with a beta-distribution with parameters $(a_1, n_1 - a_1 + 1)$. Similar statements hold for the other errors. This beta-distribution is completely determined by α_1 and n_1 and can be used to study the probability of errors for various α_1's and n_1's. It is natural to consider the percentiles of this distribution.

Let $\bar{q}(p; a_1, n_1)$ be such that

$$p = \left\{ \int_0^{\bar{q}} x^{a_1 - 1} (1 - x)^{n_1 - a_1} dx \right\} \Big/ B(a_1, n_1 - a_1 + 1)$$

for $0 < p < 1$ and $B(x_1,x_2)$ the complete beta-function. Then the probability of an error is less than $\bar{q}$ with probability p. For specific α_1 and n_1 we can evaluate these probabilities from Pearson's (1934) tables. Murphy (1948) has given graphs which may be convenient here. From these graphs, for example, with $n_1 = 100$, $\alpha_1 = .1$, the probability is approximately .9 that the probability of error is less than .14.

These relations may be useful in planning sample sizes, or for given sample sizes to study the effect of requiring different procedure sizes.

4.4 Two Consistent Procedures

In this section we set out procedures which are consistent with the procedures of Examples 2.3.1 and 2.3.3, i.e. for the most powerful procedure for the family of densities with monotone likelihood ratio and the general univariate normal family.

Example 4.4.1 The Monotone Likelihood Ratio Family

We consider constructing a nonparametric procedure to be consistent with the known distribution procedure for the distributions of Example 2.3.1. This is the class of distributions with real parameter θ possessing the property that for $\theta_1 < \theta_2$ the density ratio $f_{\theta_2}(x)/f_{\theta_1}(x)$ is strictly increasing in a real-valued statistic $T(x)$.

For this case the sets A_1 and A_2 are defined as follows. Let $t_{ij} = T(x_{ij})$, for x_{ij} the jth observation from the ith sample, $i = 1,2$; $j = 1,\ldots,n_i$. Let $t_{i(1)},\ldots,t_{i(n_i)}$ devote the n_i ordered values of t_{ij} for a fixed i. Then, if $t_{1(a_1)} \leq t_{2(n_2 - a_2 + 1)}$, we put

$$A_1 = \left\{x:T(x) < t_{1(a_1)}\right\},$$

and

$$A_2 = \left\{x:T(x) > t_{2(n_2 - a_2 + 1)}\right\}.$$

If $t_{1(n_1 - a_1 + 1)} > t_{2(a_2)}$, put

$$A_1 = \left\{x:T(x) < (t_{1(a_1)} + t_{(n_2 - a_2 + 1)})/2\right\},$$

$$A_2 = \left\{x:T(x) > (t_{1(a_1)} + t_{2(n_2 - a_2 + 1)})/2\right\}.$$

Then the procedure $d(S_1,S_2)$ obtained by (2.2.3) from these A-sets has the following properties.

(1) If $C_1 \leq C_2$, the procedure is consistent with the exact size - (α_1,α_2) most powerful procedure of Example 2.3.1.

(2) If $C_1 > C_2$, the procedure is consistent with the particular procedure of Example 2.3.1 for this case when t^* there is taken to be $(t^1_{\alpha_1} + t^2_{1-\alpha_2})/2$.

Proof of (1):

By a well-known result

$$t_{1(a_1)} \xrightarrow{P} t^1_{\alpha_1} \text{ and } t_{2(n_2 - a_2 + 1)} \xrightarrow{P} t^2_{1 - \alpha_2},$$

(Cf. Wilks, 1952, p. 272).

Now the procedure $d(S_1,S_2)$ of this example and that of example 2.3.1, $d^*(S^*_1,S^*_2)$, differ only when $t_1(a_1) \neq t^1_{\alpha_1}$ or $t_{2(n_2 - a_2 + 1)} \neq t^2_{1 - \alpha_2}$.

Then $\Pr\{d(S_1,S_2) = d^*(S_1^*,S_2^*)\} \xrightarrow{P} 1$ as $n_1 \longrightarrow \infty$ and $n_2 \longrightarrow \infty$.

Proof of (2):

The result quoted in (1) establishes this, immediately.

Example 4.4.2 General Univariate Normal Distributions

The distributions P_1 and P_2 are taken to be the same as in Example 2.3.3, i.e. P_j is a $N(\mu_j, \sigma_j^2)$ distribution for $j = 1,2$. It is required to construct tolerance regions B_j from the samples $(x_{j1},\ldots,x_{jn_j})$, $j = 1,2$, which will give a procedure consistent with the procedure $d(S_1,S_2)$ of Example 2.3.3. That procedure depended upon the quantities a, x_o, α_1, α_2.

The sample means $\bar{X}_j$ and variances s_j^2 are consistent estimators of the means μ_j and variances σ_j^2, $j = 1,2$. From these consistent estimators $\hat{a}$ and $\hat{x}_o$ can be constructed for a and x_o, if $\sigma_1^2 \neq \sigma_2^2$, i.e. there is an $\hat{a}$ and an $\hat{x}_o$ such that

$$\hat{a} \xrightarrow{P} a, \tag{4.4.1}$$

and

$$\hat{x}_o \xrightarrow{P} x_o. \tag{4.4.2}$$

Suppose $\hat{a} > 0$. Then take B_1 to be the interval $(x_{1(r_1)}, x_{1(r_2)})$, i.e. the open interval determined by the r_1st and r_2nd order statistics subject to the conditions:

$$\left.\begin{array}{ll} \text{(a)} & r_2 - r_1 = a_1, \\ \text{(b)} & \text{the quantity } \left|(x_{1(r_2)} - \hat{x}_o) - (\hat{x}_o - x_{1(r_1)})\right| \end{array}\right\} \qquad (4.4.3)$$

is a minimum.

In words, take B_1 to be the union of the a_1 consecutive sample blocks for which $\left|x_{1(r_1)} + x_{1(r_2)} - 2\hat{x}_o\right|$ is a minimum. If $x_{1(1)} < \hat{x}_o < x_{1(n_1)}$, this will require that $(x_{1(r_1)}, x_{1(r_2)})$ be the interval containing $\hat{x}_o$ for which the difference of the distances of the end points from $\hat{x}_o$ is a minimum.

We construct $\bar{B}_2$ in similar fashion from the second sample, i.e. $\bar{B}_2$ is an interval $(x_{2(r_3)}, x_{2(r_4)})$ with r_3 and r_4 chosen such that:

$$\left.\begin{array}{ll} \text{(a)} & r_4 - r_3 = n_2 - a_2 + 1, \\ \text{(b)} & \text{the quantity } \left|r_{2(r_3} + r_{2(r_4)} - 2\hat{x}_o\right| \text{is a minimum.} \end{array}\right\} \qquad (4.4.3)'$$

We then define a decision procedure $d'(S_1', S_2'; n_1, n_2)$ as follows:

(i) If $x_{1(r_1)} \geq x_{2(r_3)}$ and $x_{1(r_2)} \leq x_{2(r_4)}$,

$$S_1' = \bar{B}_1 B_2 = \left\{x : x \leq x_{2(r_3)}\right\} \cup \left\{x : x \geq x_{2(r_4)}\right\},$$

$$S_2' = B_1 \bar{B}_2 = \left\{x : x_{1(r_1)} < x < x_{1(r_2)}\right\}.$$

(ii) If $x_{1(r_1)} < x_{2(r_3)}$ and $x_{1(r_2)} \leq x_{2(r_4)}$,

$$S_1' = \left\{x : x \leq (x_{1(r_1)} + x_{2(r_3)})/2\right\} \cup \left\{x : x \geq x_{2(r_4)}\right\},$$

$$S_2' = \left\{x : ((x_{1(r_1)} + x_{2(r_3)})/2) \leq x \leq x_{1(r_2)}\right\}.$$

(iii) If $x_{1(r_1)} \geq x_{2(r_3)}$ and $x_{1(r_2)} > x_{2(r_4)}$,

$$S_1' = \left\{x : x \leq x_{2(r_3)}\right\} \cup \left\{x : x \geq (x_{2(r_4)} + x_{1(r_2)})/2\right\},$$

$$S_2' = \left\{x : x_{1(r_1)} < x < (x_{2(r_4)} + x_{1(r_2)})/2\right\}.$$

(iv) If $x_{1(r_1)} < x_{2(r_3)}$ and $x_{1(r_2)} > x_{2(r_4)}$,

$$S_2' = \left\{x : ((x_{1(r_1)} + x_{2(r_3)})/2) \leq (x_{1(r_2)} + x_{2(r_4)})/2\right\},$$

$$S_1' = \bar{S}_2' = X - S_2'.$$

Proposition 4.4.1

(a) For $C_1 \leq C_2$, the procedure $d'(S_1', S_2'; n_1, n_2)$ of this example will be consistent with the procedure $d(S_1, S_2)$ of Example 2.3.3 as $n_1 \longrightarrow \infty$ and $n_2 \longrightarrow \infty$.

(b) For $C_1 > C_2$, this procedure will be consistent with a size - (α_1, α_2) procedure with power one as $n_1 \longrightarrow \infty$ and $n_2 \longrightarrow \infty$.

(Here C_1 and C_2 are those of Theorem 2.2.1.)

Proof: It will suffice to show that the end points of the intervals B_1 and $\bar{B}_2$ converge in probability to the end points of A_1 and $\bar{A}_2$ of the procedure $d(S_1,S_2)$. We show this for B_1 and A_1, and $\bar{B}_2$ and $\bar{A}_2$ can be done in the same fashion.

Now, A_1 is the interval (x_1,x_2) where x_1 and x_2 are determined by the properties:

$$\left.\begin{array}{ll} \text{(a)} & F_1(x_2) - F_1(x_1) = \alpha_1 \\ \text{(b)} & x_1 + x_2 = 2x_o. \end{array}\right\} \qquad (4.4.4)$$

The set B_1 is the interval $(x_{1(r_1)}, x_{1(r_2)})$ where $x_{1(r_1)}$ and $x_{1(r_2)}$ are order statistics from the sample $(x_{11},\ldots,x_{1n_1})$ selected to satisfy (a) and (b) of (4.4.3).

Since B_1 is a tolerance region containing $a_1 = \left[n_1\alpha_1\right]$ blocks, $P_1(B_1)$ is a beta random variable with parameters $(a_1, n_1 - a_1 + 1)$ and

$$P_1(B_1) = F_1(x_{1(r_2)}) - F_1(x_{1(r_1)}) \xrightarrow{P} \alpha_1 \text{ as } n_1 \longrightarrow \infty. \qquad (4.4.5)$$

Put $$e_j = F_1(x_j),\ j = 1,2.$$

and $$u_j = \left[n_1 e_j\right], \text{ i.e. the greatest integer in } n_1 e_j.$$

Now $$\alpha_1 = e_2 - e_1,$$

$$\longrightarrow \qquad n_1\alpha_1 = n_1 e_2 - n_1 e_1.$$

So $$u_2 = a_1 + u_1 + w, \; w = 0,1.$$

Also, $$x_{1(u_j)} \xrightarrow{p} x_j, \; j = 1,2. \tag{4.4.6}$$

Combining (4.4.2) and (4.4.6)

$$x_{1(u_2)} - \hat{x} \xrightarrow{p} x_2 - x_o$$

$$\hat{x} - x_{1(u_1)} \xrightarrow{p} x_o - x_1$$

$$\longrightarrow \qquad x_1(u_2) + x_{1(u_1)} - 2\hat{x} \xrightarrow{p} 0$$

by (4.4.4)(b).

Now, from the manner in which $x_{1(r_1)}$ and $x_{1(r_2)}$ are chosen, (4.4.3)(b), we know that

$$\left| x_{1(r_1)} + x_{1(r_2)} - 2\hat{x} \right| \leq \left| x_{1(u_2)} + x_{1(u_1)} - 2\hat{x} \right|$$

Therefore,

$$x_{1(r_1)} + x_{1(r_2)} - 2\hat{x} \xrightarrow{p} 0$$

$$\longrightarrow \qquad x_{1(r_1)} + x_{1(r_2)} \xrightarrow{p} 2x_o \tag{4.4.7}$$

The relations (4.4.5), (4.4.7) and (4.4.4) and the functional properties of F_1 (a normal distribution function) are sufficient to show that

$$x_{1(r_j)} \longrightarrow x_j, \; j = 1,2, \text{ as } n_1 \longrightarrow \infty,$$

as was to be shown.

4.5 Selection of Tolerance Regions in the General Case

In practice it may be required to construct a discrimination procedure in situations where information concerning the distributions is not sufficient to suggest a likely parametric family on which to calibrate the procedure, as was done in the examples of the last section. Also, in most cases we will not know optimal procedures, and even if we did it is likely that any consistent nonparametric procedures would be way complicated partitions of the sample space which would be difficult to use in practice.

In selecting the tolerance regions A_j, $j = 1,\ldots,k$, which determine the procedure by (1.3.1) and (1.3.2), almost any information about the distributions can be utilized. It appears reasonable to select A_j in such a manner that the density of the distribution P_j is expected to be relatively small on A_j. This will not lead to procedures with optimal properties even for large samples but should in many cases give reasonably good procedures.

For example, suppose that the distributions are bivariate and all are thought to be unimodal. Then a reasonable choice for each A_j would be to take it to be the complement of a tolerance region $\bar{A}_j$ which is chosen as a bounded convex region containing $(n_j - a_j + 1)$ blocks. This can be accomplished in many ways and one which is very easy to apply and appears to give good results is to use the region whose boundary is made up of eitht (or less) straight line segments suggested by Tukey (1947, p. 532).

For higher dimensional distributions, similar regions bounded by hyperplanes can be used.

Fraser (1953, p. 45) suggests an approach to forming a tolerance region for a bimodal distribution. Writers on tolerance regions have given results useful in a variety of situations.

We give an (artificial) numerical example to illustrate how a procedure can be constructed. The data was generated by drawing samples of size $n_1 = n_2 = 81$ from bivariate normal distributions P_1 and P_2 with mean vectors

$$(\mu_{11}, \mu_{12}) = (0,0), \ (\mu_{21}, \mu_{22}) = (3,0),$$

and dispersion matrices

$$\sum_1 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \quad , \quad \sum_2 = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix} \quad .$$

Eight-sided regions of the type mentioned above were formed for $\alpha_1 = \alpha_2 = 0.1$ $(a_1 = a_2 = 8)$. The regions are shown in Figure 1. From the samples drawn 19 observations from P_1 and 14 from P_2 are in the region S_o.

From Murphy's chart, the probability is approximately .90 that the conditional probability of either error is less than 0.14.

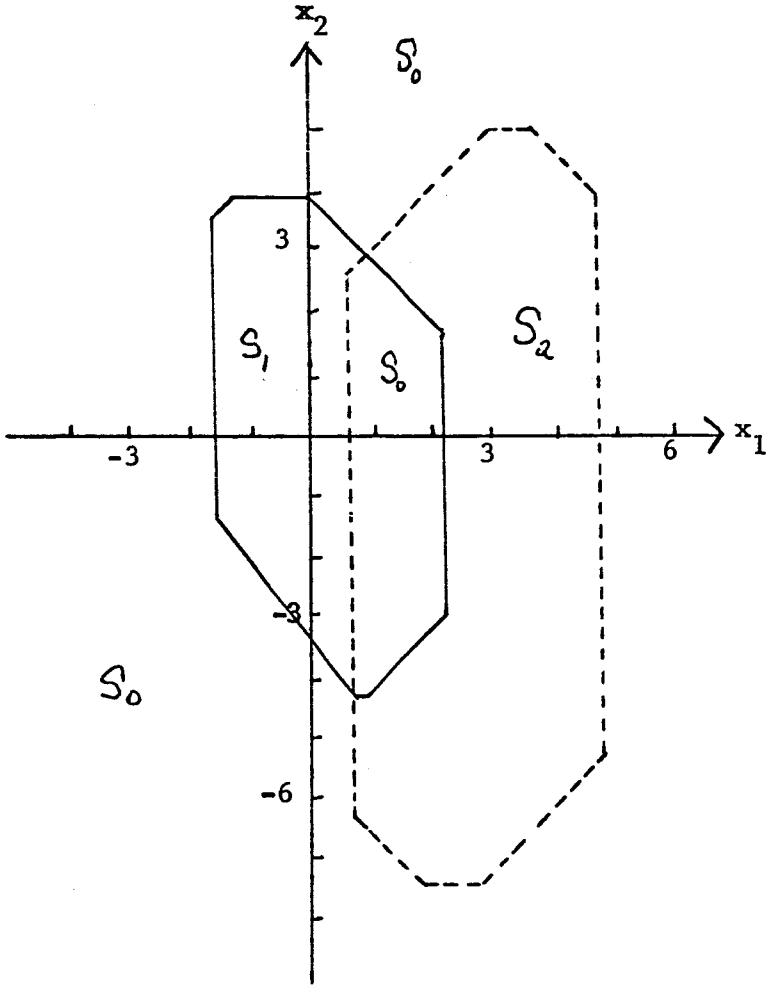

Figure 1: Tolerance Regions for Numerical Example.

Appendix 1

In (iii)(b) of Theorem 2.2.1 it is shown that for the case when $C_1 > C_2$ that any value C^* in the interval (C_2,C_1) can be used for both C_1 and C_2 in (2.2.5) and a discrimination procedure of size - (α_1,α_2) and power 1 is obtained. We consider here the problem of choosing a particular value C^* in (C_2,C_1). This is the happy situation where we can obtain a procedure with some exact size - (α_1^*,α_2^*) with either $\alpha_1^* < \alpha_1$ or $\alpha_2^* < \alpha_2$, or both. Now, in choosing the values α_1 and α_2 to begin with the statistician would have considered not only their absolute values but also their relative values. It seems reasonable to choose C^*, if possible, so that a procedure of exact size - (α_1^*,α_2^*) is obtained with $\alpha_1^*/\alpha_2^* = \alpha_1/\alpha_2$.

Theorem A.1 In Theorem 2.2.1, when $C_2 < C_1$ there is a unique C^* in the interval (C_2,C_1) such that if C^* is used for the C's in (2.2.5) and a procedure is formed by (2.2.3) it will be an exact size - (α_1^*,α_2^*) procedure with power 1 and $\alpha_1^*/\alpha_2^* = \alpha_1/\alpha_2$.

Proof: Put $G_1(C) = \Pr\left\{ x:f_1(x)/f_2(x) < C:P_1 \right\}$,

and $G_2(C) = \Pr\left\{x:f_1(x)/f_2(x) > C: P_2 \right\}$,

and $H(C) = G_1(C)/G_2(C)$.

Then $H(C)$ is a continuous strictly increasing function of C. But

$$H(C_2) < \frac{\alpha_1}{\alpha_2} \text{ and } H(C_1) > \frac{\alpha_1}{\alpha_2},$$

and therefore there is a unique solution C^*, $C_2 < C^* < C_1$, of the equation

$$H(C) = \alpha_1/\alpha_2.$$

BIBLIOGRAPHY

Fix, E. and Hodges, J. L., Jr. (1951), Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties, USAF School of Aviation Medicine, Project No. 21-49-004, Report No. 4.

Fraser, D. A. S. (1951), Sequentially determined statistically equivalent blocks, Ann. Math. Stat., Vol. 22, pp. 372-381.

Fraser, D. A. S. (1953), Nonparametric tolerance regions, Ann. Math. Stat., Vol. 24, pp. 44-55.

Fraser, D. A. S. and Guttman, I. (1956), Tolerance regions, Ann. Math. Stat., Vol. 27, pp. 162-179.

Kemperman, J. H. B. (1956), Generalized tolerance limits, Ann. Math. Stat., Vol. 27, pp. 180-186.

Lehmann, E. L. (1957), A theory of some multiple decision problems, I, Vol. 28, pp. 1-25.

Lehmann, E. L. (1959), Testing Statistical Hypotheses, Wiley, New York.

Murphy, R. B. (1948), Nonparametric tolerance limits, Ann. Math. Stat., Vol. 17, pp. 377-408.

Pearson, K., editor (1934), Tables of the Incomplete Beta Function, Cambridge University Press.

Tukey, J. W. (1947), Nonparametric estimation, II. Statistically equivalent blocks and tolerance regions--the continuous case, Ann. Math. Stat., Vol. 18, pp. 529-539.

Wald, A. (1943), An extension of Wilks' method for setting tolerance limits, Ann. Math. Stat., Vol. 14, pp. 45-55.

Wilks, S. S. (1942), Statistical prediction with special reference to the problem of tolerance limits, Ann. Math. Stat., Vol. 13, pp. 400-409.

Wilks, S. S. (1962), Mathematical Statistics, Wiley, New York.