The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 664-673

NONPARAMETRIC DISCRIMINATION USING TOLERANCE REGIONS[1]

By C. P. Quesenberry and M. P. Gessaman

North Carolina State University and Montana State University

0. Introduction and summary. A method is given which can be used to construct procedures for discriminating among distributions on a Euclidean space with continuous distribution functions. The decision space used includes "partial" decisions and the probabilities of errors are random variables with beta distributions. Emphasis is upon control of the distribution of the conditional overall probabilities of errors. These procedures can be used in a wide class of discrimination problems, such as, for example, discriminating among multivariate normal distributions with unknown, unequal dispersion matrices.

A number of other writers have suggested nonparametric discrimination procedures. The first work in this area, to the knowledge of these writers, was by Fix and Hodges [2]. Since the work by those authors has remained unpublished, a brief statement of its approach and results is given in Section 4 for comparison. Procedures have been suggested also by Stoller [13], Anderson [1] and Kendall [8].

1. The general model. An observation z is obtained on a random variable Z with probability distribution P on a Euclidean measure space $X(\mathfrak{A})$. It is assumed that P identifies with one of k distributions $P_1, \cdots, P_k$ which are distinct members of a class $\mathcal{P}$ of distributions defined on $X(\mathfrak{A})$. The class $\mathcal{P}$ will here be taken to be a subclass of the class of distributions with continuous distribution functions. The distribution function for P_j will be denoted by F_j.

The problem is to use an observation z and any information available on the distributions $P_1, \cdots, P_k$ to make a decision as to which of these distributions may have given rise to the observation. Information about the distributions may be in the form of assumptions about their properties (definition of $\mathcal{P}$), and in samples from the individual distributions.

The class of decisions which will be used here is defined as follows:

Let Δ denote the class of decisions with elements given by

$$(1.1)\quad \begin{aligned} &\delta_{i_1 \cdots i_s}: \text{ means decide that } P \,\varepsilon\, \{P_{i_1}, \cdots, P_{i_s}\} \text{ for } s = 1, \cdots, k-1 \\ &\delta_0: \text{ means reserve judgment} \end{aligned}$$

where $(i_1, \cdots, i_s)$ is a subset of the set $\{1, \cdots, k\}$.

By *reserve judgment* it is meant that no decision whatever is to be made concerning the distribution P of Z on $X(\mathfrak{A})$.

Let d denote a function which maps the sample space X to Δ. Such a function will here be called a *discrimination function* or *procedure*. It will be observed that

Received 30 March 1967; revised 5 September 1967.

[1] Work supported by NASA Research Grant NsG-562.

much of the literature on discrimination considers a subset of Δ containing only $\delta_1, \cdots, \delta_k$. A discrimination procedure d maps X to the finite set Δ with $2^k - 1$ elements. It is essentially a partition of the sample space into subsets given by

$$(1.2) \qquad S_{i_1,\cdots,i_s} = \{x : d(x) = \delta_{i_1\cdots i_s}\} \quad \text{for } s = 1, \cdots, k-1,$$
$$S_0 = \{x : d(x) = \delta_0\}.$$

If a discrimination procedure d is used, an error will be made when $P = P_j$ and $d(z) = \delta_{i_1\cdots i_s}$ with $j \notin \{i_1, \cdots, i_s\}$. Put

$$Q_j = \bigcup_{s=1}^{k-1} (\cup S_{i_1\cdots i_s})$$

where the first union is over all subsets of size s from the set $\{1, \cdots, j-1, j+1, \cdots, k\}$. Then

$$(1.3) \qquad q_j = P_j(Q_j), \qquad j = 1, \cdots, k,$$

is the probability of an error when Z has distribution P_j, i.e. $P = P_j$.

Definition 1.1. Let α_j be a fixed number in the open interval $(0, 1)$ for each $j = 1, \cdots, k$. A discrimination procedure is said to be of *size*-$(\alpha_1, \cdots, \alpha_k)$ if

$$(1.4) \qquad q_j \leqq \alpha_j \quad \text{for every} \quad j = 1, \cdots, k.$$

A procedure is of *exact size*-$(\alpha_1, \cdots, \alpha_k)$ if equality holds for every $j = 1, \cdots, k$.

2. A nonparametric procedure. In this section $\mathcal{P}$ will be the entire class of distributions which have continuous distribution functions on X. It is assumed that there is a sample $(x_{j1}, \cdots, x_{jn_j})$ available from each distribution P_j, $j = 1, \cdots, k$. Let $(\alpha_1, \cdots, \alpha_k)$ be constants in the open interval $(0, 1)$, and let a_j denote the largest integer in the quantity $\alpha_j(n_j + 1)$, i.e., $a_j = [\alpha_j(n_j + 1)]$. Using the theory of coverages (cf. [4], [5], [6], [7], [10], [12], [14], [16]) and the jth sample, construct a nonparametric tolerance region A_j containing a_j blocks on X for the distribution P_j. Each set A_j formed from the jth sample and its complementary set $\bar{A}_j = X - A_j$ constitutes a two-set partition of the sample space X. A product partition is formed from these partitions as follows.

Definition 2.1. The sets $S_{i_1\cdots i_s}$ and S_0 are defined by

$$(2.1) \qquad S_{i_1\cdots i_s} = \bar{A}_{i_1} \cdots \bar{A}_{i_s} A_{i_{s+1}} \cdots A_{i_k} \text{ for } s = 1, \cdots, k-1,$$
$$S_0 = (A_1 \cdots A_k) \cup (\bar{A}_1 \cdots \bar{A}_k),$$

where $\{i_1, \cdots, i_s\}$ is any subset of s elements of $\{1, \cdots, k\}$, except $\varnothing$ (null set) or the whole set.

Definition 2.2. The discrimination procedure d^* is defined by

$$(2.2) \qquad d^*(z) = \delta_{i_1\cdots i_s} \quad \text{if} \quad z \in S_{i_1\cdots i_s},$$
$$= \delta_0 \quad \text{if} \quad z \in S_0,$$

where the S-sets are given by (2.1).

An error will be made whenever $P = P_j$, i.e. P_j is the correct distribution, but z falls in the set

$$B_j = A_j(\bigcup_{s=1}^{k-1} (\cup \bar{A}_{i_1} \cdots \bar{A}_{i_s} \bar{A}_{i_{s+1}} \cdots A_{i_{k-1}})), \tag{2.3}$$

where the first union is over all combinations $(i_1, \cdots, i_s)$ of size s that can be taken from the set $\{1, \cdots, j-1, j+1, \cdots, k\}$, and $\{i_{s+1}, \cdots, i_{k-1}\}$ is in each case the remainder set. Observe that B_j can be written

$$B_j = A_j(\bigcup_{i=1, i\neq j}^{k} \bar{A}_i) = A_j(X - A_1 \cdots A_{j-1}A_{j+1} \cdots A_k).$$

From either expression, $B_j \subset A_j$, and

$$P_j(B_j) \leqq P_j(A_j), \quad \text{for every} \quad j = 1, \cdots, k. \tag{2.4}$$

The probability of the set A_j as measured by P_j, i.e. $P_j(A_j)$, is a beta random variable with parameters $(a_j, n_j - a_j + 1)$. Its distribution has mean

$$a_j/(n_j + 1) = \alpha_j + O(1/n_j), \qquad O(1/n_j) \geqq 0, \tag{2.5}$$

and variance

$$a_j(n_j - a_j + 1)/(n_j + 1)^2(n_j + 2). \tag{2.6}$$

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

$$P_j(A_j) \rightarrow_P \alpha_j \quad \text{as} \quad n_j \rightarrow \infty \quad \text{for all} \quad j = 1, \cdots, k. \tag{2.7}$$

The procedure d^* has the property that when Z has distribution P_j the probability that a mistake will be made is bounded by the random variable $P_j(A_j)$ with a beta-distribution with parameters $(a_j, n_j - a_j + 1)$. Similar statements hold for the other errors. This control is a consequence of having taken the number of blocks to go into A_j to be a_j. The regions A_j can be constructed in many ways. This flexibility in the choice of the regions A_j can be used to obtain procedures with other desired properties.

From (2.4), for large sample sizes the procedure is approximately size-$(\alpha_1, \cdots, \alpha_k)$. For any size samples with a specific α_j and n_j, Pearson's tables [11] can be used to give a probability statement that the probability $P_j(A_j)$ is less than a particular value. Murphy [10] has given graphs which are convenient here. From these graphs, for example, with $n_j = 100$, $\alpha_j = .1$, the probability is approximately .9 that the probability of error is less than .14.

3. Selection of tolerance regions. When considering procedures based on samples from distributions, a natural limit is provided by procedures based on completely known distributions. The next definition is a slight generalization of one given in [2].

DEFINITION 3.1. Two sequences of discrimination procedures $\{d_n'\}$ and $\{d_m''\}$ are said to be *consistent* as $n, m \rightarrow \infty$ if

$$P_j\{d_n' = d_m''\} \rightarrow_P 1 \quad \text{for} \quad j = 1, \cdots, k.$$

Interest here will be in comparing a sequence $d_{n_1,\cdots,n_k}$ with a particular procedure d_0, and it will be required to show that as $n_1, \cdots, n_k \to \infty$,

(3.1) $$P_j\{d_{n_1, \cdots, n_k} = d_0\} \to_P 1 \quad \text{for} \quad j = 1, \cdots, k.$$

The next theorem provides a discrimination procedure based on known distributions for $k = 2$ against which nonparametric procedures of Section 2 can be compared. It is a direct extension of a result of Welch [15]. Parts (i) and (ii) follow from the Neyman-Pearson lemma as given in [9] and (iii) is obvious from geometry.

THEOREM 3.1. *Let P_1 and P_2 be distinct absolutely continuous probability measures defined on a Euclidean measure space with* pdf's *f_1 and f_2 with respect to Lebesgue measure λ. Also assume $W = f_1(X)/f_2(X)$ is an absolutely continuous* rv *defined* a.e. λ.

(i) *For $0 < \alpha_j < 1, j = 1, 2$; there exist essentially unique constants c_1 and c_2 such that the sets*

$$A_1 = \{x : f_1(x) \leqq c_1 f_2(x)\}, \qquad A_2 = \{x : f_1(x) > c_2 f_2(x)\},$$

have probabilities

$$P_1(A_1) = \alpha_1, \quad \textit{and} \quad P_2(A_2) = \alpha_2 .$$

Put $H_1 = \bar{A}_1 A_2$, $H_2 = A_1 \bar{A}_2$, $H_0 = A_1 A_2 \cup \bar{A}_1 \bar{A}_2$. Let

$$d(H_1, H_2) = \delta_i \qquad \textit{if } z \in H_i, \qquad\qquad i = 0, 1, 2.$$

(ii) *If $c_1 \leqq c_2$, then $d(H_1, H_2)$ is the unique* (a.e.) *size-(α_1, α_2) procedure with the property that the probability of reserving judgment is a minimum for both P_1 and P_2,* i.e. *$P_1(H_0)$ and $P_2(H_0)$ are simultaneously minimized. The procedure $d(H_1, H_2)$ is exact size-(α_1, α_2).*

(iii) *If $c_2 < c_1$, let c^* be any number in the interval $[c_2, c_1]$. Then, if both c_1 and c_2 are replaced by c^* in the definitions of A_1 and A_2 in* (i), *the resulting procedure $d(H_1, H_2)$ will be size-(α_1, α_2), but not exact size-(α_1, α_2), and $P_1(H_0) = P_2(H_0) = 0$.*

In the next two examples nonparametric procedures d^* of Section 2 will be constructed which are consistent with the procedures given by Theorem 3.1.

EXAMPLE 3.1. *Families with monotone density ratio.* For θ a real-valued parameter, let $\mathcal{P} = \{P_\theta : \theta \varepsilon R_1\}$ be a family of absolutely continuous probability distributions with strictly monotone likelihood ratio in a real-valued statistic $T(x)$, i.e. if $\theta_2 < \theta_1$, then $f_1(x)/f_2(x)$ is strictly increasing in $T(x)$. The procedure can be constructed on the space of T (the real line) rather than the original space. Let P_j^T denote the distribution induced on the real line by T from P_{θ_j} and let c_1 be the lower α_1 percentage point of P_1^T and c_2 be the upper α_2 percentage point of P_2^T. Then a procedure of Theorem 3.1 is given by:

(i) If $c_1 \leqq c_2$,

$$\begin{aligned} d &= \delta_1 && \text{if} \quad T(z) > c_2, \\ &= \delta_2 && \text{if} \quad T(z) \leqq c_1, \\ &= \delta_0 && \text{if} \quad c_1 < T(z) \leqq c_2. \end{aligned}$$

(ii) *If* $c_1 > c_2$, put $c = (c_1 + c_2)/2$, and

$$d = \delta_1 \quad \text{if} \quad t > c,$$
$$= \delta_2 \quad \text{if} \quad t \leqq c.$$

We now set out a choice of tolerance regions which makes the procedure d^* of Section 2 consistent with the above parametric procedure. Let $t_{ij} = T(x_{ij})$ for x_{ij} the jth observation from population i (= 1, 2). Let $t_{i(1)}, \cdots, t_{i(n_i)}$ denote the n_i ordered values of the t_{ij} for each i. Let $A_1 = (-\infty, t_{1(a_1)})$ and $A_2 = (t_{2(n_2-a_2+1)}, +\infty)$. The procedure is then:

(i) If $t_{1(a_1)} \leqq t_{2(n_2-a_2+1)}$,

$$d^* = d_1 \quad \text{if} \quad T(z) > t_{2(n_2-a_2+1)},$$
$$= d_2 \quad \text{if} \quad T(z) \leqq t_{1(a_1)},$$
$$= d_0 \quad \text{if} \quad t_{1(a_1)} < T(z) \leqq t_{2(n_2-a_2+1)}.$$

(ii) If $t_{1(a_1)} > t_{2(n_2-a_2+1)}$, put $t^* = (t_{1(a_1)} + t_{2(n-a_2+1)})/2$ and

$$d^* = d_1 \quad \text{if} \quad T(z) \geqq t^*,$$
$$= d_2 \quad \text{if} \quad T(z) < t^*.$$

The consistency of the procedure d^* with d follows from the fact that $t_{1(a_1)} \rightarrow_P c_1$ and $t_{2(n-a_2+1)} \rightarrow_P c_2$.

EXAMPLE 3.2. *Univariate normal distributions.* Let P_1 and P_2 be univariate normal distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 with $\sigma_1^2 \neq \sigma_2^2$. If $a = \sigma_1^2 - \sigma_2^2 > 0$, the density ratio decreases from $-\infty$ to $x_0 = (\mu_2\sigma_1^2 - \mu_1\sigma_2^2)/(\sigma_1^2 - \sigma_2^2)$, and increases from x_0 to $+\infty$. The sets A_1 and $\bar{A}_2$ of Theorem 3.1 are

$$A_1 = \{x : x_0 - b_1 < x < x_0 + b_1\}, \tag{3.2}$$
$$\bar{A}_2 = \{x : x_0 - b_2 < x < x_0 + b_2\},$$

with b_j such that $P_j(A_j) = \alpha_j, j = 1, 2$. If $a < 0$ the sets A_1 and A_2 are chosen in similar fashion. Again, let d denote the procedure using A_1 and A_2 as in Theorem 3.1.

We now set out a choice of tolerance regions which will make the nonparametric procedure d^* of Section 2 consistent with the above procedure d based on densities.

If $\sigma_1^2 \neq \sigma_2^2$, the sample means $\bar{X}_j$ and sample variances s_j^2 can be used to construct consistent estimators $\hat{a}$ and $\hat{x}_0$ for a and x_0, i.e. there are estimators $\hat{a}$ and $\hat{x}_0$ of a and x_0 such that

$$\hat{a} \rightarrow_P a, \tag{3.3}$$

and

$$\hat{x}_0 \rightarrow_P x_0, \tag{3.4}$$

as $n_1 \to \infty$ and $n_2 \to \infty$. When $\hat{a} > 0$, 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:

(3.5) (a) $r_2 - r_1 = a_1$,

(b) the quantity $|x_{1(r_1)} + x_{1(r_2)} - 2\hat{x}_0|$ is a minimum.

In words, B_1 is the union of those a_1 consecutive blocks for which $|x_{1(r_1)} + x_{1(r_2)} - 2\hat{x}_0|$ is a minimum. If $x_{1(1)} < \hat{x}_0 < x_{1(n_1)}$, then $(x_{1(r_1)}, x_{1(r_2)})$ is the interval containing $\hat{x}_0$ for which the difference of the distances of the end points from $\hat{x}_0$ is a minimum.

The complement of B_2, i.e. $\bar{B}_2$, is constructed in a similar fashion. Let $\bar{B}_2$ be the open interval $(x_{2(r_3)}, x_{2(r_4)})$ with r_3 and r_4 determined by:

(3.6) (a) $r_4 - r_3 = n_2 - a_2 + 1$,

(b) $|x_{2(r_3)} + x_{2(r_4)} - 2\hat{x}_0|$ is a minimum.

When $\hat{a} < 0$, the sets B_1 and B_2 are defined similarly, but with $\bar{B}_1$ and B_2 the intervals centered on $\hat{x}_0$.

With $k = 2$ here, the procedure d^* is determined completely by the sets S_1 and S_2 of (2.1), since $S_0 = X - S_1 - S_2$. We give the procedure d^* by specifying these sets for all cases.

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

$$S_1 = \bar{A}_1 A_2 = \{x : x \leqq x_{2(r_3)}\} \cup \{x : x \geqq x_{2(r_4)}\},$$

$$S_2 = A_1 \bar{A}_2 = \{x : x_{1(r_1)} < x < x_{1(r_2)}\}.$$

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

$$S_1 = \{x : x \leqq (x_{1(r_1)} + x_{2(r_3)})/2\} \cup \{x : x \geqq x_{2(r_4)}\},$$

$$S_2 = \{x : ((x_{1(r_1)} + x_{2(r_3)})/2) \leqq x \leqq x_{1(r_2)}\}.$$

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

$$S_1 = \{x : x \leqq x_{2(r_3)}\} \cup \{x \geqq (x_{2(r_4)} + x_{1(r_2)})/2\},$$

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

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

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

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

If $\hat{a} < 0$, the procedure is constructed in similar fashion.

THEOREM 3.2. *The procedure d^* above is consistent with the procedure d of this example, as $n_1 \to \infty$ and $n_2 \to \infty$.*

PROOF. It is sufficient to show that the end points of B_1 and $\bar{B}_2$ converge in probability to the end-points of A_1 and $\bar{A}_2$ given in (3.2). We show this for B_1 and $\bar{B}_2$ is done similarly.

If $a > 0$, A_1 is the interval (x_1, x_2) where x_1 and x_2 are determined by the properties

$$(3.7)\quad \text{(a)} \qquad F_1(x_2) - F_1(x_1) = \alpha_1,$$

$$\text{(b)} \qquad x_1 + x_2 = 2x_0.$$

If $\hat{a} > 0$, 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 $(x_{11}, \cdots, x_{1n_1})$ selected to satisfy (a) and (b) of (3.5).

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

$$(3.8)\qquad P_1(B_1) = F_1(x_{1(r_2)}) - F_1(x_{1(r_1)}) \rightarrow_P \alpha_1 \quad \text{as} \quad n_1 \rightarrow \infty.$$

Put $e_j = F_1(x_j), j = 1, 2$, and $u_j = [n_1e_j]$, i.e. the greatest integer in n_1e_j. Now $\alpha_1 = e_2 - e_1$, and $n_1\alpha_1 = n_1e_2 - n_1e_1$. So $u_2 = a_1 + u_1 + w$, $w = 0, 1$. Also,

$$(3.9)\qquad x_{1(u_j)} \rightarrow_P x_j, \quad j = 1, 2 \text{ as } n_j \rightarrow \infty.$$

Combining (3.4) and (3.9)

$$x_{1(u_j)} - \hat{x}_0 \rightarrow_P x_j - x_0, \quad j = 1, 2 \Rightarrow x_{1(u_1)} + x_{1(u_2)} - 2\hat{x}_0 \rightarrow_P x_1 + x_2 - 2x_0 = 0,$$

by (3.7) (b), as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$.

Now, from the manner in which $x_{1(r_1)}$ and $x_{1(r_2)}$ are chosen (3.5)(b), we know that

$$|x_{1(r_1)} + x_{1(r_2)} - 2\hat{x}_0| \leqq |x_{1(u_2)} + x_{1(u_1)} - 2\hat{x}_0|.$$

Therefore, $x_{1(r_1)} + x_{1(r_2)} - 2\hat{x}_0 \rightarrow_P 0$ and

$$(3.10)\qquad x_{1(r_1)} + x_{1(r_2)} \rightarrow_P 2x_0.$$

The relations (3.3), (3.8), (3.10) and (3.7) and the functional properties of F_1 (a normal distribution function) are sufficient to show that

$$x_{1(r_j)} \rightarrow_P x_j, \quad j = 1, 2 \qquad \text{as} \quad n_1, n_2 \rightarrow \infty,$$

as was to be shown. This completes Example 3.2.

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

In selecting the tolerance regions A_j, $j = 1, \cdots, k$, which determine the procedure d^*, 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 easy to apply and appears to give good results is to use the region whose boundary is made up of eight (or less) straight line segments suggested by Tukey [12]. For higher dimensional distributions, similar regions bounded by hyperplanes can be used.

Fraser [5] 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 the construction of a procedure. 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

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}.$$

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 Murphy's [10] chart,

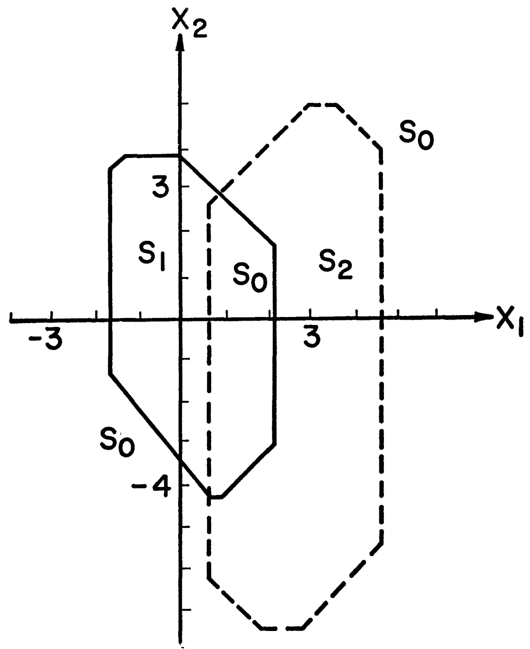

FIG. 1

the probability is approximately .90 that the conditional probability of either error is less than 0.14. If an observed z falls in S_1 it is assigned to P_1, if it falls in S_2 it is assigned to P_2 and if it falls in S_0 it is not assigned to either P_1 or P_2.

4. Discussion. The procedures described in this paper are the first completely distribution free discrimination procedures known to these writers. These procedures provide a control of the probabilities of errors for all distributions with continuous distribution functions and for all sample sizes. They can be chosen so as to have consistency properties for some families of distributions. Some words of warning are in order. The size control is possible because of the use of the decision space Δ which contains "partial" decisions, including a reserve judgment category. Whether this space is reasonable or not in a particular problem must receive careful consideration. If the distributions are "close" together or if the tolerance regions are unhappily chosen the probability of reserving judgment can be large. In this situation if the procedure was used to screen a sequence of observations $z_1, \cdots, z_m$, then the expected number which would not be classified can be large. None of these procedures are consistent for all possible distributions.

Fix and Hodges [2] suggested that an observation z be classified by considering the $h(n_1, \cdots, n_k)$ observations that are "closest" to it and assigning it to the distribution which contributes the largest number of these h observations. This procedure is shown to be consistent for the class of distributions which have continuous densities at the point z, if h is appropriately chosen. The same authors studied small sample properties for some special cases in a subsequent paper [3].

In comparison with the procedures of this paper the Fix-Hodges procedures will have the advantage of consistency against larger classes of distributions. We believe the procedures proposed here are somewhat easier to use in practice, particularly if they are to be used to screen a sequence $z_1, z_2, \cdots, z_m$ of observations. Once the calibration samples are available, the classification regions are determined and we simply observe which region each z-observation falls into. For one- or two-dimensional distributions this can be done by a simple graphing method that could be carried out by unskilled personnel.

REFERENCES

[1] ANDERSON, T. W. (1966). Some nonparametric multivariate procedures based on statistically equivalent blocks, *Proceedings of an International Symposium on Multivariate Analysis*. Academic Press, New York.

[2] 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.

[3] FIX, E. and HODGES, J. L., JR. (1952). Discriminatory Analysis, Nonparametric Discrimination: Small Sample Performance. USAF School of Aviation Medicine, Project No. 21-49-004, Report No. 11.

[4] FRASER, D. A. S. (1951). Sequentially determined statistically equivalent blocks. *Ann. Math. Statist.* **22** 372–381.

[5] FRASER, D. A. S. (1953). Nonparametric tolerance regions. *Ann. Math. Statist.* **24** 44–55.

[6] FRASER, D. A. S. and GUTTMAN, I. (1956). Tolerance regions. *Ann. Math. Statist.* **27** 162–179.

[7] KEMPERMAN, J. H. B. (1956). Generalized tolerance limits. *Ann. Math. Statist.* **27** 180–186.

[8] KENDALL, M. G. (1966). Discrimination and classification. *Proceedings of an International Symposium on Multivariate Analysis*. Academic Press, New York.

[9] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[10] MURPHY, R. B. (1948). Nonparametric tolerance limits. *Ann. Math. Statist.* **17** 377–408.

[11] PEARSON, K., editor (1934). *Tables of the Incomplete Beta Function*. Cambridge Univ. Press.

[12] TUKEY, J. W. (1947). Nonparametric estimation, II. Statistically equivalent blocks and tolerance regions—the continuous case. *Ann. Math. Statist.* **18** 529–539.

[13] STOLLER, D. S. (1954). Univariate two-population distribution-free discrimination, *J. Amer. Statist. Assoc.* **49** 770–777.

[14] WALD, A. (1943). An extension of Wilks' method for setting tolerance limits. *Ann. Math. Statist.* **14** 45–55.

[15] WELCH, B. L. (1939). Note on discriminant functions. *Biometrika* **31** 218–220.

[16] WILKS, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. *Ann. Math. Statist.* **13** 400–409.

[17] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.