A GENERAL THEORY OF DISCRIMINATION WHEN THE INFORMATION ABOUT ALTERNATIVE POPULATION DISTRIBUTIONS IS BASED ON SAMPLES

By C. Radhakrishna Rao

Indian Statistical Institute

1. Introduction. The problem of discrimination, that is of assigning an observed individual to its proper group, admits a simple solution when the distributions of measurements in the alternative populations are completely specified. Research in this direction originated with the use of the linear discriminant function introduced in 1936 by Fisher [3]. In 1939 Welch [24] showed that a general discriminant function in the case of two alternatives is the likelihood ratio of the two hypotheses, and is deducible either from Bayes' theorem with given a priori probabilities or by the use of a lemma by Neyman and Pearson [11] when the errors for the two hypotheses are minimised in any given ratio.

A general theory of decision functions when the alternatives are finite or infinite was developed by Wald [19] in 1939 and further generalized by him in 1949 [23]. In 1945 von Mises [9] obtained, in the case of a finite number of alternatives, the solution to the problem of minimising the maximum error, which is the general theme of Wald's work. Explicit solutions of Bayes' form, with given a priori probabilities or ratio of errors for the alternative groups, and the construction and use of a doubtful region were discussed by the author [13] in 1948. Related problems and the extension to problems of selection have been treated in a subsequent series of papers [15], [16].

In all these cases the alternative population distributions are assumed to be completely specified. The decision rule consists in setting up a correspondence between values observed in a sample and the alternative population distributions. In practice it is rarely possible to specify completely the distributions, but they may be estimable on the basis of independent samples from each of the alternative distributions.

Let $S_1, \cdots, S_k$ be independent samples from k alternative populations which may be partially specified, as when the functional forms of the probability densities are given but with unspecified parameters, or completely unspecified. After a sample S is drawn from a population known a priori to be one of the above set of k populations, the problem is to infer from which population the sample S has been drawn. The decision rule should be in the form of associating S with one of the samples $S_1, \cdots, S_k$, and declaring that S has come from the same population as the sample with which it is associated.

The usual practice is to estimate the alternative distributions on the basis of the sample information, and to use them in the solution which is strictly applicable when the alternatives are completely specified. This is probably the right

Received 8/6/53, revised 1/22/54.

approach when estimation is based on large samples. Fix and Hodges [4] have shown that this procedure is consistent under certain conditions, that is, with probability tending to unity it gives the same results as when the alternatives are known, provided the sample sizes are indefinitely increased. This procedure can be shown to be asymtotically the best in the sense of Wald [20].

No systematic attempt seems to have been made to offer solutions for finite samples. Wald [22] proposed to solve this problem in the case of two alternatives by obtaining the distribution of the estimated likelihood ratio or the linear discriminant function of Fisher. Even if the distribution problem is satisfactorily solved, it cannot be applied in practice since it involves unknown parameters.

In this paper some general methods have been developed with the help of which the discrimination problem can be solved, utilizing only the sample information. This theory is immediately applicable when the alternative distributions have given functional forms but with unspecified parameters. The nonparametric cases can be treated in a similar manner, but no attempt has been made in this paper to offer explicit solutions.

2. Statement of the problem. Let $p_1(x \mid \theta_1), \cdots, p_k(x \mid \theta_k)$ be k probability densities with known functional forms but unknown parameters. In the representation of the function $p(x \mid \theta)$, x stands for all the measurements and θ for all the unknown parameters. We have, in general, to deal with p-variate populations so that x stands for a vector of p stochastic variables. Samples of sizes $n_1, \cdots, n_k$ are available from these k populations. The observations from the ith population, for $i = 1, \cdots, k$, are denoted by

$$S_i: \quad x_j^i = (x_{1j}^i, \cdots, x_{pj}^i), \qquad j = 1, \cdots, n_i. \tag{2.1}$$

An individual known a priori to belong to one of the k groups has the measurements

$$S: \quad x = (x_1, \cdots, x_p). \tag{2.2}$$

The problem is to assign this individual to its proper group on the basis of the information supplied only by the observations (2.1) and (2.2), without making any assumption about the unknown parameters. The problem is similar when, instead of p measurements on a single individual, the sample S in (2.2) consists of p measurements on each of n individuals drawn from that population. The problem is to decide on the population from which S has arisen, using the information supplied by S and $S_1, \cdots, S_k$ of (2.1).

3. Some observations on the solution when the parameters are known. If the a priori probabilities of the observed individual belonging to the k groups are $\pi_1, \cdots, \pi_k$, then the Bayes' solution which minimises the errors of wrong classification is to assign individuals with measurements x to the ith group if $\pi_i p_i(x \mid \theta_i)$ has the highest value in the set

$$\pi_1 p_1(x \mid \theta_1), \cdots, \pi_k p_k(x \mid \theta_k). \tag{3.1}$$

The solution which assigns the individual to the ith group if $a_i p_i(x \mid \theta_i)$ has the highest value in the set

$$a_1 p_1(x \mid \theta_1), \cdots, a_k p_k(x \mid \theta_k) \tag{3.2}$$

has the property of minimising the frequencies of wrong classification for the various groups in a ratio determined by the procedure (3.2). This ratio is a function of $a_1, \cdots, a_k$; if possible, the constants may be chosen for any specified ratio [13].

When $\pi_1, \cdots, \pi_k$ are unknown or when the consideration of a priori probabilities is irrelevant, we have to depend on solution (3.2). One method is to choose the constants such that the errors are in an equal ratio, using the criterion of minimax [9], [23]. Another method is to choose $a_i = 1$ for all i, using the principle of maximum likelihood. The latter method gives an unbiased division of the space, that is, the probability with respect to the density p_j of all observation points assigned to the ith population is the highest for $j = i$. All Bayes' solutions do not have this property except in the case of two alternative populations. Also, it is not evident whether the minimax solution is always unbiased in the above sense. Some criterion has to be developed for the choice of a rule from the subclass of Bayes' solutions which are unbiased.

4. Large sample theory. The observations (2.1) and (2.2) considered in Section 2 can be represented by a point in a space of $(n_1 + n_2 + \cdots + n_k + 1)p$ or more generally of $(n_1 + n_2 + \cdots + n_k + n)p$ dimensions. Every division of the space into k regions $R_1, \cdots, R_k$ provides a decision rule, by which the ith population is accepted when the points fall in the corresponding region R_i.

The probability of correct classification β_i' for the ith group is the density of the region R_i when the last observation (2.2) arises from the ith group. If the a priori probability that the last observation belongs to the ith group is π_i, then the probability of correct classification is

$$\pi_1 \beta_1' + \cdots + \pi_k \beta_k' . \tag{4.1}$$

This is obviously less than $\pi_1\beta_1 + \cdots + \pi_k\beta_k$, where β_i are the values associated with the solution (3.1) when all the parameters are known a priori and samples do not provide any additional information.

Expression (4.1) is a function of the unknown parameters $(\theta_1, \cdots, \theta_k)$ and of the division $\mathfrak{D}$ of the space of $N = (n_1 + n_2 + \cdots + n_k + 1)p$ dimensions. This function is denoted by $f_N(\mathfrak{D}, \theta_1, \cdots, \theta_k)$ or simply by $f_N(\mathfrak{D}, \theta)$. Let $L_N(\mathfrak{D}_1, \mathfrak{D}_2)$ represent the least upper bound of the difference $f_N(\mathfrak{D}_1, \theta) - f_N(\mathfrak{D}_2, \theta)$ corresponding to two divisions $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

Following a concept due to Wald [20] we define a sequence of divisions $\mathfrak{D}^*$ to be *asymptotically best* if there does not exist any other sequence $\mathfrak{D}$ such that

$$\limsup_{n_i \to \infty} L_N(\mathfrak{D}, \mathfrak{D}^*) > 0. \tag{4.2}$$

If there exists a sequence of divisions $\mathfrak{D}_u$ such that

$$f_N(\mathfrak{D}_u, \theta) \to \pi_1\beta_1 + \cdots + \pi_k\beta_k \qquad \text{as } n_i \to \infty \tag{4.3}$$

uniformly in the parameters as the sample sizes individually tend to infinity, then such a sequence automatically satisfies the criterion (4.2) for being best asymptotically. Fix and Hodges [4] have shown that for the solution

$$R_i: \quad \pi_i p_i(x \mid \hat{\theta}_i) \geqq \pi_j p_j(x \mid \hat{\theta}_j), \qquad j = 1, \cdots, k, \tag{4.4}$$

where $\hat{\theta}_1, \cdots, \hat{\theta}_k$ are uniformly consistent estimates of the parameters the probability of correct classifications, tends uniformly to $\pi_1\beta_1 + \cdots + \pi_k\beta_k$ as each sample size tends to infinity, provided the probability densities satisfy some mild regularity conditions. This result, together with property of uniform consistency of maximum likelihood estimates (true under some general conditions stated by Wald, [21], provides a method of constructing an asymptotically best solution of the type (4.4).

5. Small sample theory. Let us first consider the problem of two alternative groups. There are n_1 observations from the first group, n_2 from the second, and a single observation (each observation means a set of p measurements) on an individual whose group is unknown. If θ_1 and θ_2 are the parameters for the first and second groups, then the parameters applicable to the three sets of observations are

$$H_1: \quad (\theta_1, \theta_2, \theta_1)$$

when the individual belongs to the first group and

$$H_2: \quad (\theta_1, \theta_2, \theta_2)$$

otherwise. The two alternative hypotheses from which one is to be chosen on the basis of observations are, therefore, the vectors $(\theta_1, \theta_2, \theta_1)$ and $(\theta_1, \theta_2, \theta_2)$, whatever θ_1 and θ_2 may be.

5.1. *Test for H_2 against H_1 at a fixed significance level.* Let us choose one of these hypotheses (say H_2) as null and test it against the alternative H_1. For this we need critical regions in the space of $(n_1 + n_2 + 1)p$ observations which are similar with respect to the parameters θ_1 and θ_2 under the hypothesis $(\theta_1, \theta_2, \theta_2)$. Out of these, one which maximises the power with respect to the alternatives $(\theta_1, \theta_2, \theta_1)$ is to be chosen. How far this method yields successful results may be judged by a simple example.

Let $p_1(x \mid \theta_1)$ and $p_2(x \mid \theta_2)$ be univariate normal probability densities with unknown mean values θ_1 and θ_2 and unit standard deviation. From each population n observations are taken; the mean values are found to be $\bar{x}_1$ and $\bar{x}_2$. According to the null hypothesis, the last observation x belongs to the second group. In this case

$$T_1 = \bar{x}_1, \qquad T_2 = (x + n\bar{x}_2) / (1 + n)$$

are sufficient for θ_1 and θ_2. The critical region similar with respect to θ_1 and θ_2 has a conditional size α on the surfaces of constant values of T_1 and T_2.

If to these statistics is added $T_3 = x - \bar{x}_2$, then it is necessary to consider only the conditional distribution of T_3 given T_1, T_2. In fact, T_3 is distributed independently of T_1, T_2 under both hypotheses and has the densities proportional to

$$\exp\left\{\frac{-n}{2(n+1)}\,(T_3 - \overline{\theta_1 - \theta_2})^2\right\}, \qquad \exp\left\{\frac{-n}{2(n+1)}\,T_3^2\right\},$$

whose ratio is independent of the observed values from the first group.

The test derived above is the same as that for testing whether the observation x comes from the second group when the alternatives are unspecified. The situation is somewhat unfortunate in that the test does not utilize the information given by the second sample. Perhaps it is inevitable, if we have to come to decisions independently of any a priori knowledge restricting to a fixed significance level. This, however, suggests an intuitive approach to the problem of classification.

Suppose that it is possible to test the hypothesis that the individual belongs to a specified group, say the ith, (ignoring the fact that the alternatives are confined to a finite number about which we have some information) at any given probability level of rejection, and that all the critical regions corresponding to different probability levels are well ordered, the bigger containing the smaller. We define by ξ_i the least probability level at which the ith hypothesis can be rejected. The k groups supply k values $\xi_1, \cdots, \xi_k$, and it appears to be a reasonable rule to assign the individual to the jth group if ξ_j is the maximum in the set. The optimum properties of this rule will naturally depend on the nature of tests of the above hypotheses, but this is generally applicable in situations where reasonable tests exist.

Consider for example the univariate case where the k samples provide the averages $\bar{x}_1, \cdots, \bar{x}_k$ based on sizes $n_1, \cdots, n_k$ and pooled variance s^2 based on $(\sum n_i - k)$ degrees of freedom. If x is the observation on an individual to be classified, we calculate the probabilities

$$\xi_i = P\{\,|t| > |x - \bar{x}_i|\,/\,s\sqrt{1 + 1/n_i}\},$$

where the variable t has Student's distribution based on $(\sum n_i - k)$ degrees of freedom. The individual is assigned to that group for which ξ_i is a maximum. This rule is immediately applicable since it involves no new technique. Only a reasonable test should exist and the probability integral table should be available. It is, however, not easy to say what optimum properties are implied by this rule, except that errors are less for groups with larger sample sizes.

Another intuitive method which may yield fruitful results is to use fiducial probability distributions if they exist (as defined by Fisher [2]) of the observation x. Corresponding to k groups we can set up the k alternative fiducial distributions, using the samples. These distributions are parameter-free and the problem now reduces to the classical case of assigning the observation x to one of k populations whose distributions are completely defined. It would, however, be somewhat difficult to study the optimum properties of this procedure.

In the following we will lay down a few postulates concerning the nature of the decision rule, and obtain solutions which have optimum properties when the alternative hypotheses are close to one another.

5.2. *A general postulate concerning the decision rule.* Let us denote the probability density of the observations from the ith group by

$$P_i(x^i \mid \theta_i) = p_i(x_1^i \mid \theta_i) \cdots p_i(x_{n_i}^i \mid \theta_i), \qquad i = 1, \cdots, k.$$

For simplicity we shall consider only nonrandomised decision rules which need a division of the sample space of $(n_1 + \cdots + n_k + 1)p$ dimensions into mutually exclusive regions $R_1, \cdots, R_k$. The rule of behaviour is to accept the hypothesis that the individual belongs to the ith population when the sample point falls in R_i.

In developing the arguments we shall choose the case of two alternative populations only, the conclusions being the same for several. In this problem there are two regions R_1 and R_2. The proportion of errors committed when the individual belongs to the first group is

$$\alpha_1(\theta_1, \theta_2) = \int_{R_2} P_1(x^1 \mid \theta_1) \quad P_2(x^2 \mid \theta_2) \quad p_1(x \mid \theta_1)\, dv.$$

Similarly for the other group,

$$\alpha_2(\theta_1, \theta_2) = \int_{R_1} P_1(x^1 \mid \theta_1) \quad P_2(x^2 \mid \theta_2) \quad p_2(x \mid \theta_2)\, dv.$$

Suppose that we need a decision rule for which the linear compound of errors

$$\pi_1\alpha_1(\theta_1, \theta_2) + \pi_2\alpha_2(\theta_1, \theta_2) \tag{5.2.1}$$

is a minimum. The compounding coefficients π_1 and π_2 may be assigned a priori probabilities, or suitable weights may be attached to the errors. If there exists a division of the space which minimises (5.2.1) irrespective of the true values of the parameters, then such a division cannot be improved upon. The minimum value of (5.2.1) for any given values θ_1 and θ_2 of the parameters is attained for the regions

$$R_1: \ \pi_1 p_1(x \mid \theta_1) \geqq \pi_2 p_2(x \mid \theta_2), \qquad R_2: \ \pi_2 p_2(x \mid \theta_2) \geqq \pi_1 p_1(x \mid \theta_1).$$

If the boundary of these regions is independent of the parameters θ_1 and θ_2, then we have a uniformly best division of the space. In such a case the sample observations do not provide any additional information.

If we exclude such special cases, it would appear that whatever may be the set of regions offered it will not be uniformly the best for all values of the unknown parameters and can be good only in some restricted sense. We need then some reasonable postulates governing the choice of a decision rule.

An obvious requirement on the decision rule is that it should not lead to contradictions or give recognizably bad results in particular cases. Let us consider the degenerate case when the two alternative distributions are identical, that is, $\theta_1 = \theta_2 = \theta$. For any division R_1, R_2 of the space, the errors committed for the two groups in this situation are $\alpha_1(\theta, \theta)$ and $\alpha_2(\theta, \theta)$, with the necessary condition $\alpha_1(\theta, \theta) + \alpha_2(\theta, \theta) = 1$. When the population distributions are equal the only rule is to assign individuals at random, subject to a given or a chosen frequency of errors for the two groups. It seems therefore reasonable to postulate that $\alpha_1(\theta, \theta)$ and $\alpha_2(\theta, \theta)$ should be constant independently of the common values of the parameters θ_1 and θ_2.

Further, let us imagine that for a given division of the space the value of $\alpha_1(\theta, \theta)$ at a neighbouring value $(\theta + \delta\theta)$ is more, implying that

$$\delta\theta \left\{ \frac{\partial}{\partial\theta_1} \alpha_1(\theta_1, \theta_2) + \frac{\partial}{\partial\theta_2} \alpha_1(\theta_1, \theta_2) \right\}_{\theta_1=\theta_2=\theta} = \delta\theta(a + b)$$

is positive, or if $\delta\theta$ is positive the expression within the brackets is positive. The value of $\alpha_1(\theta, \theta)$ at the value $\theta - \delta\theta$ is $\alpha_1(\theta, \theta) - \delta\theta(a + b)$, which is smaller than $\alpha_1(\theta, \theta)$.

Since we have assumed continuity of the functions involved, throughout a neighbourhood (over a square) around the point (θ, θ), $\alpha_1(\theta_1, \theta_2)$ lies between $\alpha_1(\theta, \theta) \pm \delta\theta(a + b)/2$. Consequently, throughout this square around (θ, θ), $\alpha_1(\theta_1, \theta_2)$ exceeds the value $\alpha_1(\theta - \delta\theta, \theta - \delta\theta)$ at the neighbouring point. It is clearly undesirable that more errors are committed when the populations are different than when they are equal in any given region including the line of equality (at least as a boundary) in which the possible values of (θ_1, θ_2) are restricted to lie. A necessary condition for this is that $(a + b)$ should vanish at all points on the line of equality, implying that $\alpha_1(\theta, \theta)$ and therefore $\alpha_2(\theta, \theta)$ should be constant independently of the common values.

We are not, at the moment, demanding that the functions $\alpha_1(\theta_1, \theta_2)$ and $\alpha_2(\theta_1, \theta_2)$ should be stationary or that they should be absolutely minimum on the line of equality, although these appear to be desirable properties leading to unbiased divisions of the space. It is, however, necessary that $\alpha_i(\theta, \theta)$ should be constant independently of θ. In our arguments, we have explicitely used one parameter although we said that θ stands for a vector of parameters. This is clearly admissible since we can consider variations in one parameter keeping the others fixed.

The restriction that $\alpha_i(\theta, \theta)$ is constant on the line of equality implies that with respect to the probability density of observations

$$P_1(x^1 \mid \theta)\, P_2(x^2 \mid \theta)\, p(x \mid \theta)$$

that is, when $\theta_1 = \theta_2 = \theta$, the regions R_1 and R_2 are similar to the sample space with respect to the free parameter θ. In such a case we shall say that there exists a *similar division* of the sample space with respect to θ.

Having determined similar divisions, we have to select the best one among them. It is hard to imagine that there exist regions which minimise uniformly any linear compound of the errors $\pi_1\alpha_1(\theta_1, \theta_2) + \pi_2\alpha_2(\theta_1, \theta_2)$ except in some special cases. Some suitable criteria have to be used, as in Section 6, depending on the type of difficulties which the probability densities may present, to obtain reasonable solutions.

We have yet to consider the nature of the error functions on the line of equality where the maximum error for any group cannot be reduced below 50 per cent. It may be reasonable to impose the restriction

$$\alpha_1(\theta, \theta) = \alpha_2(\theta, \theta) = 0.50$$

In some problems the actual specification of the ratio of errors $\alpha_1(\theta, \theta) / \alpha_2(\theta, \theta)$ may be left open, and chosen to satisfy some optimum conditions. We could impose any other restriction specifying the ratio of errors at any value of the set (θ_1, θ_2).

A special case is the choice $\alpha_2(\theta, \theta) = 0.05$, which leads to a test of significance of the null hypothesis H_2, that the observed individual belongs to the second group, against the alternative that he belongs to the first group. This will be useful in further subdividing the regions R_1 and R_2 in such a way that some portions lead to more confident classifications, while other portions permit only provisional decisions. Further theory is developed in the examples considered in the next section.

The arguments of this section can be extended to the case of more than two alternative populations. The division of the space into k regions must be such that the error committed for any group remains constant whenever the populations are identical, whatever may be the common values of the parameters. As in the case of two populations, we may choose this constant to be $1/k$ for each of the alternative populations. Also, any ratio of these constants may be specified, or sometimes suitably determined. Problems of tests of significance may be considered in a similar way.

The general postulate laid down in this section can be used in the solution of a wide variety of problems in classification. For instance, the problem of the greater mean (Bahadur and Robbins, [1]) admits a neat solution once this condition is imposed.

6. Some optimum conditions and derivation of decision rules. It is known (Neyman, [10]) that similar regions can be constructed, when a set of sufficient statistics exist, by considering the relative probability density of the observations, given the set of sufficient statistics. Lehmann and Scheffé [8] have shown recently that when the parameters admit a minimal set of sufficient statistics such that no function depending on them has zero expectation (in which case the set is said to be complete) then all similar regions have Neyman's structure. That sufficient statistics possess this unicity property under some conditions has been formally demonstrated by the author [14]. In the illustrations considered in this paper, these results are used without proof.

If T stands for the complete set of sufficient statistics for θ, then we can write down the joint densities of the observations under H_1 and H_2 as

$$H_1: \quad \mathcal{P}_1(T \mid \eta, \delta)P_1(x^1, x^2, x \mid \eta, \delta, T) = \mathcal{P}_1(\eta, \delta)P_1(\eta, \delta),$$

$$H_2: \quad \mathcal{P}_2(T \mid \eta, \delta)P_2(x^1, x^2, x \mid \eta, \delta, T) = \mathcal{P}_2(\eta, \delta)P_2(\eta, \delta),$$

where P_1 and P_2 are relative probability densities of observations given T, and $\mathcal{P}_1$ and $\mathcal{P}_2$ are the densities of T, while η and δ are the vectors $(\theta_1 + \theta_2)$ and $(\theta_1 - \theta_2)$.

The regions R_{1T} and R_{2T} on the surface of T for which the linear compound of overall errors $a\alpha_1(\theta_1, \theta_2) + b\alpha_2(\theta_1, \theta_2)$ is a minimum subject to the condition

$$\alpha_1(\theta, \theta)/\alpha_2(\theta, \theta) = 1/\rho, \tag{6.1}$$

where ρ is fixed, are given by

$$R_{1T}: \quad a\mathcal{P}_1(\eta, \delta)P_1(\eta, \delta) + \lambda_1 P_1(\eta, 0) \geqq b\mathcal{P}_2(\eta, \delta)P_2(\eta, \delta) + \lambda_2 P_2(\eta, 0), \tag{6.2}$$

with the reverse relationship in R_{2T}. The constants λ_1 and λ_2 are determined to satisfy the condition (6.1). The proof of the result (6.2) and the subsequent ones follow from a lemma proved by the author in ([16], p. 340). The region (6.2) will generally depend upon the unknown quantities η and δ, and is therefore not useful. We therefore need to restrict the regions by imposing some condition on the error functions.

We first note that the errors $\alpha_1(\theta_1, \theta_2)$ and $\alpha_2(\theta_1, \theta_2)$ could be written in terms of η and δ as $\alpha_1(\eta, \delta)$ and $\alpha_2(\eta, \delta)$, using α_1 and α_2 as symbols for error functions. Let

$$\alpha_i'(\eta, \delta) = \frac{\partial}{\partial \delta}\alpha_i(\eta, \delta), \qquad i = 1, 2,$$

denote the derivatives with respect to the parameters δ in any given direction. The values $\alpha_1(\eta, 0)$ and $\alpha_2(\eta, 0)$ are the errors when the populations are identical and the slopes of the error functions in the given direction at $\delta = 0$ are

$$\alpha_1'(\eta, 0), \qquad \alpha_2'(\eta, 0). \tag{6.3}$$

To ensure optimum properties, at least in the neighbourhood of the line of equality of the two populations, we may minimise a linear compound of the slopes (6.3), or minimise them in a given ratio. Observing that minimising the slopes (6.3) is equivalent to minimising the slopes corresponding to the relative errors on the surfaces of T, we find the boundary separating the best regions R_{1T} and R_{2T} on the surfaces of T as

$$a\frac{\partial}{\partial \delta}P_1(\eta, 0) + \lambda_1 P_1(\eta, 0) = b\frac{\partial}{\partial \delta}P_2(\eta, 0) + \lambda_2 P_2(\eta, 0). \tag{6.4}$$

(i) For any a and b and the choice of λ_1 and λ_2 to satisfy the condition (6.1), the linear compound $a\alpha_1'(\eta, 0) + b\alpha_2'(\eta, 0)$ is minimised. The special values $a = b$ may be useful in practice.

(ii) For a suitable choice of a, b, λ_1, and λ_2, the slopes $\alpha_1'(\eta, 0)$ and $\alpha_2'(\eta, 0)$ can be minimised in a given ratio in addition to the condition (6.1) being satisfied. The special case of the equality of the slopes may be of some practical interest.

The solution (6.4) may depend on η when $P_1'(\eta, 0)$ and $P_2'(\eta, 0)$ contain η. In the illustrations considered in Section 7, the $P_i'(\eta, \delta)$ are functions of δ only, so that the solution (6.4) serves the purpose. Otherwise some method has to be devised, such as minimising the average slopes over a set of η or considering regions similar for η with respect to the functions $P_i'(\eta, 0)$.

For the problem of testing the hypothesis H_2 against the alternative H_1 we have to construct a region w on the T surfaces satisfying the four conditions (given $\gamma \leqq 0$)

$$
(6.5) \quad \begin{cases} \text{(a)} & \int_w P_2(\delta = 0)\, dv = 0.05, \\ \text{(b)} & \int_w P_2'(\delta = 0)\, dv = \gamma, \\ \text{(c)} & \int_w P_2(\delta) \quad dv \leqq 0.05, \\ \text{(d)} & \int_w P_1'(\delta = 0)\, dv = \text{a maximum.} \end{cases}
$$

The region satisfying the conditions (a), (b) and (d) is given by

$$
(6.6) \qquad w: \quad aP_1'(0) + \lambda_1 P_1(0) \geqq bP_2'(0) + \lambda_2 P_2(0)
$$

on the T surfaces where a, b, λ_1 and λ_2 are suitably chosen. For this region the slope of the conditional power curve $\beta_1'(\delta)$ at $\delta = 0$ is a function of γ defined in condition (b). We now relax this condition and maximise $\beta_1'(0)$ subject to the condition $\gamma \leqq 0$. With this choice of γ we can set up the region w as in (6.6). If, for this region, condition (c) is satisfied, then we obtain a test of the hypothesis that H_2 is true against the alternative that H_1 is true. This test is most powerful in a given direction for small differences in the parameters of the two populations.

The situations in tests of significance and discriminatory problems are diagrammatically represented in Figure 1.

If the direction used in the above construction with the first derivatives is not justifiable, then we may try to impose further restrictions such as unbiasedness of the error functions on the line of equality

$$
(6.7) \qquad \alpha_1'(\eta, 0) = 0, \qquad \alpha_2'(\eta, 0) = 0.
$$

We will assume that this condition implies that the derivatives of these errors vanish when $\delta = 0$ for all T. The derivatives are calculated from the conditional probability densities

$$\frac{\partial}{\partial \delta} \alpha_1(\eta, \delta, T), \qquad \frac{\partial}{\partial \delta} \alpha_2(\eta, \delta, T), \tag{6.8}$$

where $\int \alpha_i(\eta, \delta, T)\mathcal{P}_i(T \mid \eta, \delta)\, dT = \alpha_i(\eta, \delta)$. In all the illustrations considered in Section 7 this condition is automatically satisfied. Otherwise it may be necessary to impose the conditions (6.8) which may be only sufficient for (6.7).

We consider the second derivatives of the relative probability densities with respect to the elements of the vector of parameters $\delta = (\delta_1, \delta_2, \cdots)$. Defining for $k = 1$ or 2

$$P_k^{ij} = \frac{\partial^2}{\partial \delta_i\, \partial \delta_j} P_k(\delta = 0), \qquad P_k^i = \frac{\partial}{\partial \delta_i} P_k(\delta = 0),$$

let us construct the regions

$$\begin{aligned} (6.9) \quad R_1: \quad & \sum\sum a_{ij}P_1^{ij} + \lambda_{11}P_1^1 + \lambda_{12}P_1^2 + \cdots + \mu_1 P_1 \\ & \geqq \sum\sum b_{ij}P_2^{ij} + \lambda_{21}P_2^1 + \lambda_{22}P_2^2 + \cdots + \mu_2 P_2, \end{aligned}$$

with the reverse relationship in R_2.

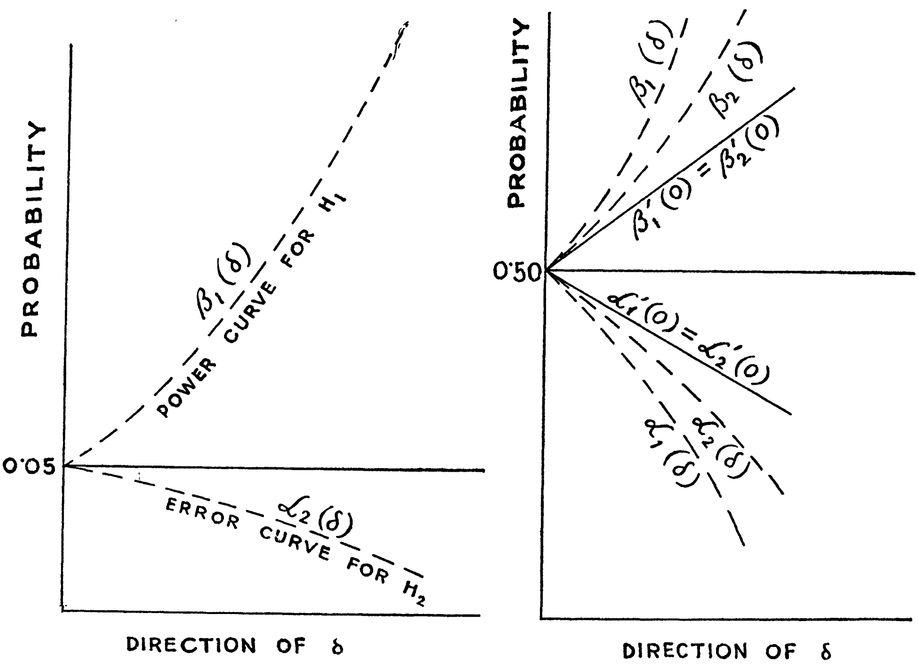

FIG. 1. Power and error curves for tests of significance (left) and for problems of discrimination (right).

(i) The regions R_1 and R_2 minimise

$$\sum\sum a_{ij} \frac{\partial^2}{\partial \delta_i \, \partial \delta_j} \alpha_1(\eta, \delta) + \sum\sum b_{ij} \frac{\partial^2}{\partial \delta_i \, \partial \delta_j} \alpha_2(\eta, \delta)$$

at $\delta = 0$ for given a_{ij} and b_{ij}, provided λ_{ij} and μ_i are chosen to satisfy the condition (6.8) and a given ratio of errors when $\delta = 0$.

(ii) For a suitable choice of a_{ij} and b_{ij}, the local powers of discrimination for the two groups can be made constant on the ellipses

$$\sum\sum \gamma^{ij} \delta_i \delta_j = \text{constant} \tag{6.10}$$

and their sum then maximised. Condition (6.10) implies that the second derivatives are in the ratio γ^{ij}.

(iii) By a suitable choice of a_{ij} and b_{ij} we could also construct a critical region w of a given size such that the first derivatives of $\alpha_1(\eta, \delta)$ and $\alpha_2(\eta, \delta)$ vanish at $\delta = 0$ and that $\int_w \sum\sum \gamma^{ij} P_1^{ij} \, dv$ is maximised subject to the condition $\int_w \sum\sum \beta^{ij} P_2^{ij} \, dv \leqq 0$, where γ^{ij} and β^{ij} are assigned as in (6.10). Such a region can be used in testing the hypothesis H_2 against the alternative H_1, provided the region is so adjusted that its size under H_2 is 5 per cent when $\delta = 0$ and less than or equal to 5 per cent when $\delta \neq 0$.

Another alternative is to restrict to those regions which give the errors as functions of a distance Δ between two populations. (Distance is a suitably defined function of the parameters of two populations. The construction of distance functions is discussed in two papers by the author [12], [14].) Even restricting to this class, it may not be possible to obtain regions for which a given linear compound of the errors $a\alpha_1(\Delta) + b\alpha_2(\Delta)$ is minimised. In such a case, we may try to minimise

$$a \frac{d\alpha_1(0)}{d\Delta} + b \frac{d\alpha_2(0)}{d\Delta} \tag{6.11}$$

to obtain regions for discrimination. For tests of significance we may have to maximise $-d\alpha_1(0)/d\Delta$ subject to the conditions

$$\frac{d\alpha_2(0)}{d\Delta} \leqq 0, \qquad \alpha_2(0) = 0.05.$$

In Section 7, we shall show that such regions can be constructed.

Another possible approach is to consider decision rules which satisfy the principle of invariance [7]. It appears that some of the results obtained here can also be deduced by using this principle.

The necessity of considering the derivatives in (6.11) arises only when no uniformly best regions exist in the class which gives the errors as functions of Δ only.

Besides the parameters θ which are considered to vary from population to population, there may be other unknown parameters ϕ which are the same for all populations. Thus we may consider the class of normal distributions with the same unknown variance but different mean values. In such situations, we may demand that the division of the space be similar for the unknown parameters ϕ also when the populations are identical in the θ parameters.

This introduces fresh complications in the applications of the results (6.4), (6.6), (6.9) and (6.11) for the derivation of optimum regions. Fortunately, in some cases the problems can be reduced in such a way that the above results are directly applicable, as shown in Section 7.2.

One may argue that in laying down the decision rules, undue emphasis is laid on discriminating between populations which are close to one another. In the first place, this is done just to set up decision rules which do not involve the unknown parameters. In the absence of rules which are uniformly best, we can think only of rules which are best at some assigned values of the parameters, or at most for an assigned set of values.

The requirement that the decision rule should possess some optimum properties in the neighbourhood of equality of the populations is not unrealistic since in practice we often meet with alternatives which are closely related; the methods developed are best suited to such situations. It is, however, possible to reduce decision rules which have *optimum properties for a given difference in the parameters* of the two populations. These may be useful in some situations. Of course, whatever may be the rule offered, it is better to examine its performance for all possible differences in the parameters of the two populations and satisfy oneself whether it can be reasonably applied in a given situation.

7. Illustrations.

7.1. *Multivariate populations, dispersion matrix known.* Let us consider p characters and represent the relevant statistics computed from the three samples, and functions based on them, as follows:

Group	1	2	3
Sample size	n_1	n_2	n
Average of ith character	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\bar{x}_i$
Population average	μ_{i1}	μ_{i2}	μ_i

$$\delta_i = \mu_{i1} - \mu_{i2} \qquad Z_i = n\bar{x}_i + n_1\bar{x}_{i1} + n_2\bar{x}_{i2}$$

$$T_i = \bar{x}_{i1} - \bar{x}_{i2}, \qquad U_i = \bar{x}_i - (n_1\bar{x}_{i1} + n_2\bar{x}_{i2})/(n_1 + n_2)$$

$$f_1 = (n_1 + n_2)/n_1n_2, \qquad f_2 = (n + n_1 + n_2)/n(n_1 + n_2)$$

$$g_1 = n_2/(n_1 + n_2), \qquad g_2 = -n_1/(n_1 + n_2)$$

$$q_1 = 1/f_1 + g_1^2/f_1, \qquad q_2 = 1/f_2 + g_2^2/f_2$$

When $\mu_{i1} = \mu_{i2} = \mu_i$ for all i, then Z_i are sufficient statistics for μ_i and we need consider only the relative distribution of T_i, U_i given Z_i. It is easy to see that T_i, U_i, and Z_i are all independently distributed, so Z_i can be dropped from further consideration if errors are restricted to functions of δ only. The joint probability density of T_i, U_i under the hypothesis H_1 is

$$P_1(T, U, \delta) = \text{const} \exp\left[-\tfrac{1}{2}\sum\sum\alpha^{ij}\left\{\frac{(T_i-\delta_i)(T_j-\delta_j)}{f_1}+\frac{(U_i-g_1\delta_i)(U_j-\delta_j g_1)}{f_2}\right\}\right].$$

Under H_2, we replace g_1 by g_2 to obtain $P_2(T, U, \delta)$.

In this problem we consider regions R_1 and R_2 whose size under both hypotheses depends only on the single parameter $\Delta = \sum\alpha^{ij}\delta_i\delta_j$, since the formulae of Section 6, using the first and second derivatives, do not yield fruitful results. With this end in view let us consider the surface integral

$$\int_\Delta P_1(T, U, \delta)G^{-1}\,d\delta_1\cdots d\delta_p$$

where

$$P_1(T, U, \delta) = P_1(T, U, 0)\exp\{-\tfrac{1}{2}q_1\sum\sum\alpha^{ij}[(\delta_i-W_i)(\delta_j-W_j)-W_iW_j]\}, \tag{7.1.1}$$

$$G = \int_\Delta d\delta_1\cdots d\delta_p, \qquad W_i = (T_i/f_1+g_1U_1/f_2)\div q_1.$$

For the above integration, only the first expression in the exponential of (7.1.1) is important. This may be regarded as a p-variant normal distribution of $\delta_1, \cdots, \delta_p$. Then the integration results in a noncentral χ^2 probability density with noncentral parameter M_1, given by

$$\left\{\frac{q_1^{p/2}\,|\alpha^{ij}|^{\frac{1}{2}}}{\pi^{p/2}}\right\}^{-1} e^{-M_1/2-\Delta'/2}(\Delta')^{(p/2)-1}\sum\left(\frac{M_1\Delta'}{4}\right)^r\frac{1}{r!\,\Gamma\left(\frac{p}{2}+r\right)} \tag{7.1.2}$$

$$M_1 = q_1\sum\sum\alpha^{ij}W_iW_j, \qquad \Delta' = q_1\Delta,$$

([16], pp. 51, 57). Observing that

$$\int_\Delta d\delta_1\cdots d\delta_p = \left\{\frac{|\alpha^{ij}|^{\frac{1}{2}}}{(2\pi)^{p/2}}\right\}^{-1}\frac{\Delta^{(p/2)-1}}{2^{p/2}\Gamma\left(\frac{p}{2}\right)}\,d\Delta$$

and changing over to Δ in (7.1.2), we find the total integral in (7.1.1) for the two cases to be

$$G_1(\Delta) = P_1(T, U, 0)e^{-q_1\Delta/2}\sum_0^\infty\frac{\Gamma(\frac{1}{2}p)}{r!\,\Gamma(\frac{1}{2}p+r)}\left(\frac{M_1q_1\Delta}{4}\right)^r,$$

$$G_2(\Delta) = P_2(T, U, 0)e^{-q_2\Delta/2}\sum_0^\infty\frac{\Gamma(\frac{1}{2}p)}{r!\,\Gamma(\frac{1}{2}p+r)}\left(\frac{M_2q_2\Delta}{4}\right)^r.$$

Restricting the minimisation of a linear compound of the errors to the divisions which yield errors as functions of Δ only, the boundary is obtained as

$$aG_1(\Delta) = bG_2(\Delta). \tag{7.1.3}$$

The proof of (7.1.3) is trivial ([16], p. 285). We have to make sure that for the regions based on (7.1.3) the errors are functions of Δ only. This follows from the invariance of the expressions M_1 and M_2. The solution (7.1.3) in general involves Δ and can be used only when Δ is known. We can, however, seek for optimum properties in the neighbourhood of $\Delta = 0$, where

$$\frac{dG_1(\Delta)}{d\Delta} = P_1(T, U, 0)q_1\left(\frac{M_1}{2p} - \frac{1}{2}\right),$$

$$\frac{dG_2(\Delta)}{d\Delta} = P_2(T, U, 0)q_2\left(\frac{M_2}{2p} - \frac{1}{2}\right).$$

Consider the boundary

$$a\frac{dG_1(0)}{d\Delta} + \lambda_1 P_1(T, U, 0) = b\frac{dG_2(0)}{d\Delta} + \lambda_2 P_2(T, U, 0), \tag{7.1.4}$$

or $aq_1M_1 - bq_2M_2 = c$. The choice $a = b$ leads to a minimum value of the sum of the derivatives of the errors. In this case the boundary is

$$\sum\sum\alpha^{ij}\left\{\frac{g_1^2 - g_2^2}{f_2^2} U_iU_j + \frac{2(g_1 - g_2)}{f_1f_2} T_iU_j\right\} = \frac{p(g_1^2 - g_2^2)}{f_2}. \tag{7.1.5}$$

For the case $g_1 = -g_2$, equation (7.1.5) reduces to $\sum\sum\alpha^{ij}T_iU_j = 0$, so that the regions are

$$R_1: \quad \sum\sum\alpha^{ij}T_iU_j \geqq 0, \qquad R_2: \quad \sum\sum\alpha^{ij}T_iU_j \leqq 0,$$

with fifty per cent errors when $\Delta = 0$. In this case the regions are uniformly best for all Δ because $G_1(\Delta) \geqq G_2(\Delta)$ in R_1 and the reverse is true in R_2, irrespective of the value of Δ. The appropriate regions when $n_1 \neq n_2$ have the boundary as in (7.1.5). For these regions the errors may not be fifty per cent when $\Delta = 0$. If this condition is also insisted upon, the boundary is

$$\sum\sum\alpha^{ij}\left\{\frac{g_1^2 - g_2^2}{f_2^2} U_iU_j + \frac{2(g_1 - g_2)}{f_1f_2} T_iU_j\right\} \geqq c, \tag{7.1.6}$$

where c is suitably determined. We can also choose a, b, λ_1, and λ_2 in (7.1.4) subject to the condition that the derivatives of errors are equal when $\Delta = 0$.

For tests of significance the critical region is of the form

$$w: \quad aM_1 - bM_2 \geqq c \tag{7.1.7}$$

where a, b, and c are determined such that

$$\frac{d}{d\Delta}\{P_1(aM_1 - bM_2 \geqq c)\}_{\Delta=0} \tag{7.1.8}$$

is a maximum subject to

$$P_2(aM_1 - bM_2 \geqq c \mid \Delta = 0) = 0.05, \qquad \frac{d}{d\Delta}\{P_2(aM_1 - bM_2 \geqq c)\}_{\Delta=0} \leqq 0.$$

In the above expressions, P_1 stands for the probability according to the first hypothesis and P_2 for the second. The ultimate solution depends on the evaluation of the expressions (7.1.8). The problem needs further investigation.

In the univariate problem, if $(n_1 - n_2)$ is not large compared to $(n_1 + n_2)$, the regions for classification are obtained as special cases of (7.1.5) as

$$R_1: \quad TU \geqq 0, \qquad R_2: \quad TU \leqq 0,$$

$$T = \bar{x}_1 - \bar{x}_2, \qquad U = \bar{x} - (n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2),$$

where $\bar{x}_1$ and $\bar{x}_2$ are the averages of the two samples and $\bar{x}$ that of the sample to be classified. The critical region for testing H_2 against H_1 is of the form $TU \geqq c$, where c is determined to ensure five per cent size when $\delta = 0$. The regions for classification depending on different combinations of T and U are diagrammatically represented in Figure 2.

7.2. *Multivariate populations, dispersion matrix unknown.* In addition to the statistics defined in Section 7.1 we need estimates of the dispersion elements when the populations are identical, that is, when $\delta_1 = 0$ for all i. Let S_{ij} denote the pooled sum of products within the three samples with $(n_1 + n_2 + n - 3)$

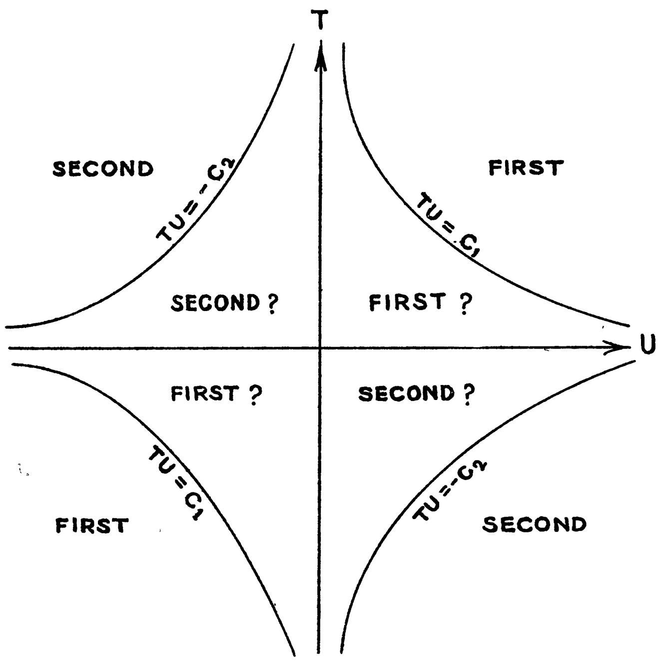

FIG. 2. Division of the space for different decisions

degrees of freedom. When δ_1 is zero, all the observations can be regarded as samples from the same population so that we have estimates of the dispersion elements based on $(n_1 + n_2 + n - 1)$ degrees of freedom. If B_{ij} denotes the corrected sum of products from the combined samples, then

$$B_{ij} = S_{ij} + T_i T_j / f_1 + U_i U_j / f_2 .$$

The statistics Z_i (defined in Sec. 7.1) and B_{ij} are sufficient for the common mean values and the elements α_{ij} of the dispersion matrix. Similar divisions of the space are obtained by considering exclusive regions on the surfaces of constant values of Z_i and B_{ij}, subject to some conditions. The probability density of T_i, U_i, and S_{ij} under the hypothesis H_1 is

$$\text{const } |S_{ij}|^{m/2} \exp \{-\tfrac{1}{2}\textstyle\sum\sum \alpha^{ij}[S_{ij} + (T_i - \delta_i)(T_j - \delta_j)/f_1 + (U_i - g_1\delta_i)(U_j - g_1\delta_j)/f_2]\}$$

where $m = (n_1 + n_2 + n - p - 4)$. Changing over to T_i, U_i, and B_{ij} permits their joint density to be written as the product of

$$P(B_{ij} \mid \delta = 0) = \text{const } |B_{ij}|^{m/2+1}, \exp \{-\tfrac{1}{2}\textstyle\sum\sum \alpha^{ij} B_{ij}\},$$

$$F(B, T, U) = |B_{ij} - T_i T_j/f_1 - U_i U_j / f_2|^{m/2} \div |B_{ij}|^{m/2+1},$$

$$Q_1(\delta) = \exp \{\textstyle\sum (T_i/f_1 + g_1 U_i/f_2)\zeta_i - \tfrac{1}{2}\psi_1^2 ,$$

where $\zeta_1 = \alpha^{1i}\delta_1 + \cdots + \alpha^{pi}\delta_p$ and $\psi_1^2 = q_1\Delta = q_1\sum\sum\alpha_{ij}\zeta_i\zeta_j$. The probability density under the second hypothesis is obtained by replacing g_1 and q_1 by g_2 and q_2 in the above expressions. We shall consider divisions R_1, R_2 for which the errors are a function of the Mahalanobis distance $\Delta = \sum\sum\alpha_{ij}\zeta_i\zeta_j$ only. This means

$$\text{(7.2.1)} \qquad \int P(B_{ij} \mid \delta = 0)\, dB \int_{R_{1B}} e^{\psi_1^2/2} F(B, T, U) Q_1(\delta)\, dT\, dU = \beta_1(\Delta).$$

Following the arguments of Hsu [6] and Simaika [17] in a similar situation, we can show that condition (7.2.1) implies

$$\int_{R_{1B}} e^{\psi_1^2/2} F(B, T, U) Q_1(\delta)\, dT\, dU = G_1(K),$$

where $K = \sum\sum B_{ij}\zeta_i\zeta_j$. If we are minimising a linear compound of errors it is enough to maximise $ae^{-\psi_1^2/2}G_1(K) + be^{-\psi_2^2/2}G_2(K)$ on the surfaces of B_{ij}, since the expected value of this linear compound integrated over B_{ij} with density $P(B_{ij} \mid \delta = 0)$ gives the linear compound of correct classifications to be maximised. There is no hope of obtaining a solution without involving Δ, except perhaps when $n_1 = n_2$. We shall therefore minimise a linear compound of the derivatives with respect to Δ at the value zero, or maximise

$$\text{(7.2.2.)} \qquad a\frac{d}{d\Delta}\{e^{-\psi_1^2/2}\beta_1(\Delta)\} + b\frac{d}{d\Delta}\{e^{-\psi_2^2/2}\beta_2(\Delta)\}$$

$+ \Delta = 0$. Evaluation of the three terms at $\Delta = 0$ yields

$$e^{-\psi_1^2/2}\beta_1(\Delta) = \int P(B_{ij} \mid \delta = 0)\, dB \int_{R_{1B}} F(B, T, U) Q_1(\delta)\, dT\, dU,$$

$$= e^{-\psi_1^2} \int P(B_{ij} \mid \delta = 0) G_1(K)\, dB;$$

$$\alpha_{ij} \frac{d\beta_1(0)}{d\Delta} = \int B_{ij} P(B_{ij} \mid \delta = 0) \frac{dG_1(0)}{dK}\, dB,$$

$$= \frac{dG_1(0)}{dK} \int B_{ij} P(B_{ij} \mid \delta = 0)\, dB = (m + p + 3)\alpha_{ij} \frac{dG_1(0)}{dK};$$

$$\frac{d}{d\Delta} e^{-\psi_1^2/2}\beta_1(\Delta) = -\frac{q_1}{2}\beta_1(0) + \beta_1'(0).$$

Consequently the value of (7.2.2) at $\Delta = 0$ is

$$\left\{a \frac{dG_1(0)}{dK} + b \frac{dG_2(0)}{dK}\right\}(m + p + 3) - \left\{\frac{aq_1}{2} G_1(0) + \frac{bq_2}{2} G_2(0)\right\}.$$

Since the latter depends only on the errors when $\Delta = 0$, we need only maximise the former or the expression $a dG_1(0)/dK + b dG_2(0)/dK$ if possible, subject to given magnitudes of errors when $\Delta = 0$. Observing that

$$G_1(K) = \int_{R_{1B}} F(B, T, U) \exp \{\textstyle\sum (T_i/f_1 + g_1 U_i/f_2)\zeta_i\}\, dT\, dU,$$

let us consider the surface integral over the surface $S = \sum\sum B_{ij}\zeta_i\zeta_j$

$$\text{(7.2.3)} \qquad \int_S F(B, T, U) \exp \{\textstyle\sum (T_i/f_1 + g_1 U_i/f_2)\zeta_i\} G^{-1}\, d\zeta_1 \cdots d\zeta_p, \qquad G = \int_S d\zeta_1 \cdots d\zeta_p.$$

As in (7.1.1), the value of (7.2.3) is

$$\text{(7.2.4)} \qquad \text{const}\, F(B, T, U) \sum \frac{\Gamma(\frac{1}{2}p)}{r!\Gamma(\frac{1}{2}p + r)} \left(\frac{M_1 K}{4}\right)^r,$$

where

$$M_1 = \sum\sum B^{ij}(T_i/f_1 + g_1 U_i/f_2)(T_j/f_1 + g_1 U_j/f_2).$$

The derivative of (7.2.4) with respect to K at $K = 0$ is const M_1, and the derivative corresponding to the second hypothesis is const M_2 where the two constants are the same. We can now define the boundary $aM_1 - bM_2 = c$ over the surfaces of B_{ij}. The constants a, b, and c may be suitably chosen. For discrimination we might choose $a = b$ and $c = 0$, in which case the sum of the derivatives of the errors is minimised. We can choose a, b, and c differently as in other cases considered in Section 7.1.

For tests of significance we need to determine a, b, and c such that over the surfaces of B_{ij}

$$\int_w F(B, T, U)Q_2(0)\, dT\, dU = 0.05 \qquad \frac{d}{dK}\int_w F(B, T, U)Q_1(\delta)\, dT\, dU$$

is a maximum at $K = 0$ subject to

$$\frac{d}{dK}\int_w F(B, T, U)Q_2(\delta)\, dT\, dv \leqq 0, \qquad k = 0.$$

Here w is the region on the surfaces of B_{ij} where $aM_1 - bM_2 \geqq c$. It is easy to see that the distribution of the statistic $aM_1 - bM_2$ under any hypothesis is dependent on Δ only, thus ensuring the validity of the arguments used in the derivation of the regions.

The distribution problems connected with the test criteria developed here have yet to be tackled. Some results obtanied by Wald [22], Harter [5] and Sitgreaves [18] in the reduction of distribution of the discriminant function and difference of two quadratic forms will be extremely useful in the study of these problems.

REFERENCES

[1] R. R. Bahadur and H. Robbins, "The problem of the greater mean," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 469–487.

[2] R. A. Fisher, "The fiducial argument in statistical inference," *Ann. Eugenics*, Vol. 6 (1935), pp. 391–398.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, Vol. 7 (1936), pp. 179–188.

[4] E. Fix and J. L. Hodges, "Discriminatory analysis: nonparametric discrimination: consistency problems", School of Aviation Medicine. Project number 21-49-004, (1951).

[5] H. L. Harter, "On the distribution of Wald's classification statistic," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 58–67.

[6] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), pp. 62–69.

[7] E. L. Lehmann, "Theory of testing hypotheses," (Lecture notes, 1948).

[8] E. L. Lehmann and H. Scheffe, "Completeness, similar regions and unbiased estimation," *Sankhya*, Vol. 10 (1950), pp. 305–340.

[9] R. v. Mises, "On the classification of observation data into distinct groups," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 68–73.

[10] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 236 (1937), pp. 333–380.

[11] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1-161.

[12] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[13] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *J. Roy. Stat. Soc. Suppl.*, Vol. 10 (1948), pp. 159–193.

[14] C. R. Rao, "Sufficient statistics and minimum variance estimates," *Proc. Cambridge Phil. Soc.*, Vol. 45 (1949), pp. 213–218.

[15] C. R. Rao, "Statistical inference applied to classificatory problems," *Sankhya*, Vol. 10 (1950), pp. 229–256, Vol. 11 (1951), pp. 107–116, 257–272, Vol. 12 (1952–53), pp. 229–246.

[16] C. R. Rao, "*Advanced Statistical Methods in Biometric Research*," John Wiley and Sons, 1952, New York.

[17] J. B. Simaika, "On the optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), pp. 70–80.

[18] R. Sitgreaves, "On the distribution of two random matrices used in classification procedures," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 263–270.

[19] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 299–326.

[20] A. Wald, "Asymptotically most powerful tests of statistical hypotheses," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 1–19.

[21] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[22] A. Wald, "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[23] A. Wald, "Statistical decision functions," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 165–205.

[24] B. L. Welch, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218–220.