STATISTICAL INFERENCE APPLIED TO CLASSIFICATORY PROBLEMS

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

PART I : NULL HYPOTHESIS, DISCRIMINATORY PROBLEMS AND DISTANCE POWER TESTS.

			Page
Section	1.	Introduction ...	229
	2.	Null Hypothesis ...	230
	3.	Problem of discrimination	
		(a) The general problem ...	234
		(b) Uncertainty of the apriori information that one of the alternatives is correct ...	235
		(c) The doubtful region ...	236
		(d) Allocation of a number of individuals to two or more groups in assigned ratios ...	240
	4.	Distance Power tests	
		(a) Distance functions ...	244
		(b) A lemma on power functions ...	247
		(c) Tests for a finite number of alternatives ...	248
		(d) Tests when alternatives are continuous ...	252
		(e) Neyman and Pearson's region of type C with a suitable ellipsoid of equidetectability ...	253
	5.	References. ...	256

Introduction

The discriminant function introduced by R. A. Fisher has proved a valuable tool in biological research. It has also opened up a variety of problems which are of interest in the general theory of statistical inference. The aim of this paper is to present some of the logical problems in the theory of discrimination and suggest suitable methods of solving them.

To start with, it is useful to distinguish problems of discrimination from those of testing of hypothesis. Recently, there has been a tendency to treat both these problems on an equal footing and this has, no doubt, caused a good deal of confusion. In testing of hypothesis we have a clearly stated null hypothesis and a comparatively undefined set of alternatives. The emphasis is more on the null hypothesis which may be rejected or provisionally accepted. When a null hypothesis is rejected no decision is made about the actual alternative hypohthesis. But in problems of discrimination we have a class of alternative hypotheses out of which one has to be chosen. While it is a question of rejecting the null hypothesis at a given risk in the former problem, it is a question of balancing between wrong and correct decisions in the latter problem. While the apriori probabilities have no place, even conceptually, in problems of testing a null hypothesis, they are essential for a satisfactory solution of the problems of discrimination. In all scientific investigations both the problems are important.

The first part of this paper deals with the concept of null hypothesis and the various situations where tests of hypotheses and discriminant functions have to be used. The concept of distance power test has been introduced as an alternative to the average power test. This introduces some uniqueness in the test criteria being invariant for transformations of the parameters occuring in a given probability density function.

Some examples and other complicated problems leading to the sequential theory of discrimination (with closer limits than those proposed by Wald, 1945) when there is an upper limit to the number of observations (Rao: 1950) will be discussed in Part II of this paper.

2. NULL HYPOTHESIS

Consider the following problem which frequently crops up in biological research.

A specimen is observed and it is desired to know, on the basis of some morphological measurements, whether it belongs to a previously classified group whose characteristics are either known or estimated from a sample of individuals from that group. In such a problem there are only two possibilities, either the new find belongs to a known group or to an unknown group. The alternative to the specified one is clearly an undefinied one. This might be a new group whose existence has yet to be established.

Thus, when a fossil is discovered the paleontologist enquires whether it is a specimen from a known collection. Such an enquiry is, often, made with the hope of obtaining a negative answer in which case the fossil could be taken as a new specimen.

The investigator may not be successful in distinguishing a new specimen from a previous collection because the answer depends on the evidence supplied by the observed one. The only safeguard offered by a statistical test in such a case is to check the investigator from rushing to a hasty conclusion unless the evidence is strong enough. If specimens, such as the observed or those differing to a greater extent than the observed from the characteristics of a known group, form a reasonable proportion in the group then the evidence for rejecting the null hypothesis cannot be considered conclusive. Only when this proportion is small, one could take a risk in asserting that the observed specimen belongs to a new group. How small this proportion or *level of significance* should be depends on the risk involved in asserting that the null hypothesis is wrong when in fact it is true. The choice of level of significance is arbitrary in this sense[1] but once it is fixed the rule of procedure is determined exactly. Thus it may be possible to refute any statement made about the observed specimen. Such an inference is possible only when some risk is allowed.

On the other hand it is almost impossible to assert that the new find belongs to a specified group. To make this latter statement one must ascertain whether the chance of the observed specimen arising from any other group is small. This is clearly not possible when the alternative groups are undefined ones.

There is clearly no scope for the introduction of apriori probabilities in this case. However perfect our past knowledge may be about the species that have been already studied and their relative numbers nothing can be said about the new species to be discovered. When these latter ones are considered as alternatives to a null hypothesis tested there is no method of attaching apriori probabilities to the alternatives.

Sometimes the apriori probabilities are introduced not as objective quantities measured by observed frequencies but as measuring merely 'psychological tendencies'. If this is so we need further rules of procedure for choosing the apriori probabilities themselves. One can recall the efforts made by Jeffreys (1948) in this connexion. To remove some apparent contradictions in Bayes' postulate of equal ignorance Jeffreys advocates the use of certain invariant functions of the parameters occuring in a probability distribution as apriori weights. Even here no argument is put forward for using particular invarint functions of the parameters. In fact, different choices lead to different results so that no objective theory could be built up on the lines of inverse probability.

To take another example a geneticist enquires on the basis of observed data whether two factors are segregating independently. If he can disprove this with some confidence then he acquires some basis to plan the future experiments to estimate the intensity of their linkage and to study the relationship of the two factors under consideration with others. If data are insufficiently numerous loose linkages go undetected and it is only by repeated experimentation and accumulation of evidence supplied by other factors linked with the former that some definite conclusion can be arrived at.

[1] It is not arbitrary in the sense that we are assuming one value when in fact it should be something else. It is one which is chosen by the investigator. Thus if the consequences of rejecting a true hypothesis involve a great loss it is reasonable to keep it as low as possible.

The alternative to the hypothesis of independence in the above problem is linkage with all possible values of the recombination fraction[2] (lying between 0 to 1). To the experimenter it is a definite knowledge if he can disprove the hypothesis of independence. Then only, he will proceed to enquire what the value of the recombination fraction is and try to obtain an estimate. To ask for apriori probabilities of the alternative recombination fractions before attempting to answer the problem posed is to believe that from previous experience the frequencies with which various recombination fractions occur could be deduced. But there may not be sufficient reason to believe that the frequencies so derived correspond to the total frequencies obtainable from all possible factors known and unknown.

In the previous problem of the paleontologist the alternatives are completely undefined while in the second of the geneticist the possible alternatives are known, viz., that the recombination fraction lies between (0 to 1). But in both these types of problems there is no scope for the introduction of apriori probabilities. Experience has shown that for an advance in knowledge an investigator need not depend on the apriori probabilities. "Whatever the reasons are which give experimenters confidence that they can draw valid conclusions from their results, they seem to act just as powerfully whether the experimenter has heard of the theory of inverse probability or not". (Fisher, 1947, p.7).

The null hypothesis is one which is chosen by the experimenter appropriate to his enquiry. When sufficient evidence gathers against it during the experimental work he rejects it. He is not trying to balance between the evidences supplied by the data on the various alternatives.

Whether a particular null hypothesis is rejected or not there is a class of null hypótheses which are not contradicted by the data at a given level of significance. Any hypothesis outside this class is rejected. The class of null hypotheses acceptable to the data supplies us with what may be called a *fiducial set*. When the hypotheses refer to the values of a parameter the fiducial set will be in the nature of an interval called the fiducial interval (Fisher, 1947). The fiducial set of hypotheses may be asserted to contain the true hypothesis because the chance of it being left out is small (equal to the percentage level of significance chosen). Thus, although it is not possible to *accept* any single hypothesis it is possible to restrict the scope of enquiry to only a subset of all possible alternatives. Any further discrimination among the alternatives in the fiducial set has necessarily to be based on insufficient evidence. No statement of confidence can be made about a single hypothesis chosen by any rule of procedure as the most appropriate for the data and consequently such a procucure does not possess a scientific basis of inference.

If the problem needs the choice of a single hypothesis then what should be the nature of the answer? We might try to formulate a rule of procedure which selects

[2] Formerly it was believed that the range of this fraction is (0, ½). Fisher, Lyon and Owen (1947) discussed the possibility of the recombination fraction exceeding ½.

a hypothesis which is as near as possible to the true hypothesis and which in large samples differs very little from the true one with probability approaching certainty. The procedure of choosing that hypothesis which maximises the likelihood, advocated by Fisher, conforms to the above requirement to a large extent. Thus the two methodological problems, testing of hypotheses and estimation admit neat solutions independent of the probabilities *apriori*. Much of the confusion in the current discussions of these topics can be avoided once the logic of null hypothesis as stated by R. A. Fisher is admitted.

"The two classes of results which are distinguished by our test of significance are, on one hand, those which show a significant discrepancy from a certain hypothesis; and on the other, results which show no significant discrepancy from this hypothesis. This hypothesis, which may or may not be impugned by the result of an experiment, is again characteristic of all experimentation. Much confusion would often be avoided if it were explicitly formulated when the experiment is designed. In relation to any experiment we may speak of this as the *Null Hypothesis*; and it should be noted that the null hypothesis is never proved or established, but is possibly disproved in the course of the experimentation. Every experiment may be said to exist only in order to give the *facts a chance of disproving the null hypothesis*" (Fisher, 1947, p.16).

Although the emphasis in tests of significance is on the null hypothesis sufficient care should be taken to see that the facts are given a *fair* chance of disproving the null hypothesis. The following views of R. A. Fisher are again relevant.

"The interpretation of the experiment consisted in dividing these results in two classes, one which is to be judged as opposed to, and the other as conformable with the null hypothesis. If these classes of results are chosen, such that the first will occur when the null hypothesis is true with a known degree of rarity, for example, 5% or 1% of trials, then we have a test by which to judge, at a known level of significance whether or not the data contradict the hypothesis to be tested.

"We may now observe that the same data may contradict the hypothesis in any one of a number of different ways. For example in the psycho-physical experiment[3], it is not only possible for the subject to designate the cups correctly more often than would be expected by chance, but is also possible that she may do so less often. Instead of using a test of significance which separates from the remainder a group of possible occurences, known to have certain small probability when the null hypothesis is true and characterised by showing an excess of correct classifications, we might have chosen a test separating an equally infrequent group of occurences of the opposite kind. The reason for not using this later test is obvious, since the object of the experiment was to demonstrate, if it existed, the sensory discrimination of a subject claiming to be able to distinguish correctly two classes of objects. For this purpose the new test proposed would be entirely inappropriate, and no experimenter would be tempted to employ it. Mathematically, however, it is as valid as any other, in that

[3] This reference is to an experiment designed to test the claim of a lady that by tasting a cup of tea she can discriminate whether milk or tea infusion was first added to the cup.

with proper randomisation it is demonstrable that it would give a significant result with known probability, if the null hypothesis were true'' (Fisher, 1947, p.183).

No definite rules can be given in the choice of a suitable test criteria. R. A. Fisher objects to any formal theory which explicitly makes use of the alternative hypotheses and which in practice does not always lead to tests of significance independent of the alternative hypotheses assumed. Commenting on the error of the second kind introduced by Neyman and Pearson (1933), Fisher says

"It (the notion of the second kind of error) has no meaning with respect to simple tests of significance in which the only available expectations are those which flow from the null hypothesis being true" (Fisher 1947, p.17).

Neyman and Pearson (1933) have no doubt brought in the alternative hypothesis only as an intermediate step. Ultimately the aim is not to make any assumption about the alternative hypothesis. This is automatically provided when a *uniformly most powerful test* exists. But such tests are very rare so that they had to introduce tests depending on locally powerful unbiased regions. But these regions ensure maximum power only in the neighbourhood of the null hypothesis and the 'final verdict of the practical value of these regions will depend therefore on the properties of the power function throughout the whole range of admissable values of a parameter'. If the power is small over a wide range except in the neighbourhood of the null hypothesis the test would not be useful. Any consideration of the power function beyond the neighbourhood of the null hypothesis will introduce the unknown parameters, or their apriori distribution in the test criteria.

Although the theory of locally powerful tests cannot be accepted as a general theory, the methods suggested can be used to obtain test criteria which may be compared with any other test.

General theories of testing of hypothesis have been useful, in as much as they provide methods for comparison of various test criteria for a given set of alternative hypotheses. But it may not be maintained that all tests of significance applicable for any situation can be derived from suitably defined set of axioms by a purely deductive process.

3. Problems of Discrimination

(a) *The general problem*

We now come to a group of problems where apriori probabilities are needed for a satisfactory solution, and the null hypothesis does not play a prominent role but is sometimes posed to arrive at a decision subject to a small risk.

Thus when a question is asked whether a skull or a jaw bone belonged to a male or a female there are evidently two alternative hypotheses and one has to be chosen.

Here a rule of procedure is needed by which the individual specimen could be assigned to one or other of the groups. In any such rule of procedure errors are nevitable. If α_1 is the proportion of wrong classifications for individuals from the first group and α_2 for the second group then the proportion of wrong classifications on the total is $\pi_1\alpha_1+\pi_2\alpha_2$, when the individuals are regarded as randomly observed from a poulation containing the members of the two groups in the ratio $\pi_1:\pi_2$. $(\pi_1+\pi_2=1)$ Evidently that procedure is the best which gives the least possible value to the expression $\pi_1\alpha_1+\pi_2\alpha_2$.

If $f_1(x|\theta_1)$[4] and $f_2(x|\theta_2)$ are probability densities of some measurable characters in the two groups then the best solution which minimises ($\pi_1\alpha_1+\pi_1\alpha_2$) is that of assigning an individual to the first group if

$$\pi_1 f_1(x|\theta_1) \geqslant \pi_2 f_2(x|\theta_2) \qquad \dots \quad (3a.1)$$

and to the second if

$$\pi_1 f_1(x|\theta_0) \leqslant \pi_2 f_2(x|\theta_2) \qquad \dots \quad (3a.2)$$

The case where the equality holds can be decided by a toss of the coin.

If r_1 is the loss resulting in assigning an individual of the first group to the second and r_2 for the second to the first then the best solution is one which minimises the expected loss $\pi_1\alpha_1r_1+\pi_2\alpha_2r_2$.

The best solution which minimises the expected loss is that of assigning an individual to the first group if

$$\pi_1 r_1 f_1(x|\theta_1) \geqslant \pi_2 r_2 f_2(x|\theta_2) \qquad \dots \quad (3a.3)$$

and to the second if

$$\pi_1 r_1 f_1(x|\theta_1) \leqslant \pi_2 r_2 f_2(x|\theta_2) \qquad \dots \quad (3a.4)$$

These solutions are derived in an earlier paper by using a general lemma given in appendix A of Rao (1948).

In any given problem the quantities r_1 and r_2 appearing in the best solution (3a.3 & 4) are easily ascertainable but the apriori probabilities are either unknown or estimated from scrappy data. In sexing jaw bones one might use the sex ratio derived from other properly sexed skeletal material recovered from the same area as the jaw bone. As for instance if a number of pelvic bones could be recovered they can be sexed almost correctly. With other bones sexing by anatomical appreciation is subject to certain amount of error. It is claimed by anthropologists that a skull could be sexed with a fair chance of success. In any case even a rough idea of the apriori probabilities will be useful in such problems of classification.

(b) *Uncertainty of the apriori information that one of the alternatives is correct*

For the use of the discriminant function, it must be known that an individual to be classified belongs to one or other of a given number of groups. This knowledge

[4] In the representation of the probability density x stands for all the available measurements and θ for all the parameters.

is, very often, available from external evidence (*apriori* information). In some problems, as in the case of two sexes, there are only two possible groups. But in cases where the external evidence is meagre, the classification of an individual as a member of one of the given groups may be subject to another kind of error, viz., the wrong assumption that he belongs to one or other of the given groups when, in fact, he comes from another unknown group. In the absence of any definite knowledge it may be necessary to examine by the internal evidence supplied by the measurements whether the individual could be considered to belong to one among the given groups.

In solving such a problem whether the Highdown skull belongs to the Bronze Age or the Iron Age comprising the Romano-British the author (Rao, 1948) has tested separately the two null hypotheses, (1) it belongs to the Bronze Age and (2) it belongs to the other group. On 5% level each of the two hypotheses could not be rejected. This gave a sufficient justification for setting up a discriminant function between the two groups and then deciding the issue.

It must be noted that in testing two null hypotheses separately we are not testing the combined null hypothesis that the specimen belongs to one or other of the groups at the 5% level. This is because out of the 5% of the rejected cases under one hypothesis some of them are accepted under the second hypothesis so that when the 5% level is used for the two hypotheses separately we will be judging the combined null hypothesis at a lower level.

Some adjustments[1] could be made for this but in practice the procedure indicated above can be safely followed because the error involved is on the right side.

If the two populations are multinormal with the same dispersion matrix $\Lambda = (\lambda_{ij})$ with its inverse $\Lambda^{-1} = (\lambda^{ij})$ and mean values $m_1, m_2, \ldots m_p$ and $m'_1, m'_2, \ldots m'_p$ then the critical region, for testing the hypothesis that an observed specimen with measurements $x_1, \ldots, x_p$ belongs to one or other of the two populations, can be defined by

$$\Sigma\,\Sigma\,\lambda^{ij}(x_i-m_i)(x_j-m_j) > \mu \text{ and } \Sigma\,\Sigma\,\lambda^{ij}(x_i-m'_i)(x_j-m'_j) \geqslant \mu$$

where μ is chosen such that the size of the region is 5% which ever of the two hypotheses is true.

(c) *The doubtful region*

In the problem of discrimination considered in section (3a) a working rule is provided by which the number of wrong classifications could be kept at a minimum level. Suppose a doctor wants to discriminate between two types of neurotics, psychopaths and obsessionists, on the basis of some tests. If the test scores of neurotics properly diagnosed are available, then assuming that the ratio of the two types of pati-

[1] In general if the null hypothesis consists of a set of distinct hypotheses the critical region may be chosen, if possible, to have the same size for any distinct hypothesis in the set comprising the null hypothesis. The solution satisfying the above requirement may not exist in which case the critical region may be chosen such that its size does not exceed a given value for any distinct hypothesis in the null set. Some results obtained in this connexion will be published in a subsequent communication.

ents admitted into the hospital in the past, represents the ratio in the general population, the doctor can set up the criteria (3a. 1 & 2). By following this procedure he can minimise the number of cases of wrong diagnosis. This method was suggested in a paper by Rao and Slater (1949) in the discrimination of five types of neurotic patients using three indicators.

But in problems like this the groups overlap to a large extent so that even by following the best procedure the percentage of wrong classifications remain quite high. By increasing the number of characters or indicators in the above problem (ref. Rao, 1948, section 6) this percentage could be made smaller and smaller but not always below an irreducible minimum because of the correlations between the characters.

Further a stage may be reached when the cost involved in further examination will not be commensurate to the reduction in the number of wrong classifications. But subject to a given cost, the indicators can be chosen so as to minimise the number of wrong classifications. Thus one has to balance between the errors committed and the time or money available.

But by following this procedure it may be difficult to assert that an individual belongs to one group or the other unless the groups are well separated in which case the proportion of wrong classifications will be low. On the other hand one may take the view that, whatever may be the basis of judgement, in some cases it should be possible to give a decisive answer (subject to a small risk) while in others no decision or only provisional decisions could be made. The latter constitute the doubtful cases which need further examination.

Cases also arise where the question asked is whether a selected individual can be asserted to belong to one particular group out of a given number of possibilities. Consider the problem of the Highdown skull referred to in a previous paper (Rao: 1948). The grave findings associated with the skull excavated from the 'invasion horizon' do not give any conclusive evidence as to whether the skull belonged to a Bronze Age 'defender' or an Iron Age 'invader'. It may or may not be possible to give a definite answer in such a problem. The case has to be judged on its individual merits considering the probability of its having come from one group or the other.

Sitting on the fence is a scientific attitude if it means looking for further evidence and better methods of judgement to be able to give a definite answer.

Take the case of doctor who has a routine method of diagnosing a disease or discriminating among a number of diseases[*]. Although by following this procedure he commits the least possible errors, he likes to be more confident about his diagnosis in some selected cases. If the routine method does not give him *sufficient assurance* in any such case he may supplement it by further tests.

[*] For instance, there are two types of Jaundice which are difficult to distinguish. One calls for a surgical treatment, the other for medical treatment. A discriminant function based on two bio-chemical tests is used in practice to ensure a greater certainty of diagnosis for far less laboratory work.

For any specially chosen case like this or an individual find as the Highdown skull, the rule of procedure suggested should necessarily be independent of the apriori probabilities used in the general problem of discrimination. Firstly, such apriori probabilities may not be available, as in the case of the Highdown skull it is not possible to know the proportions of Bronze and Iron Age cranial population. Secondly, even if such knowledge is available from previous experience, this is not strictly applicable in a case *not chosen at random* from a mixed population. For instance the proportions applicable to the Highdown skull may depend on the numbers of Bronze and Iron Age warriors who went down fighting and not the general proportion.

Thus a problem where one individual is in question has to be distinguished from the problem where a number of individuals have to be classified into a given number of groups by using suitable criteria. The latter, however, supplies a provisional answer to the former but for definite answers suitable criteria have to be developed.

Let us consider the case of two alternative groups with probability densities $f_1(x|\theta_1)$ and $f_2(x|\theta_2)$. The risk minimising solution referred to in section 3a leads to assigning an individual to the first group if

$$r_1\pi_1 f_1(x\theta_1) \geqslant r_2\pi_2 f_2(x|\theta_2)$$

and to the second if

$$r_1\pi_1 f_1(x\theta_1) \leqslant r_2\pi_2 f_2(x|\theta_2)$$

Let x_1 and x_2 be the expected proportions of wrongly classified individuals of the first and second groups by following the above rule. If x_1 and x_2 are small then we could assert in any individual case that he is rightly classified. Otherwise we may follow the procedure of assigning an individual to the first group if

$$f_1(x|\theta_1) > Af_2(x\theta_2)$$

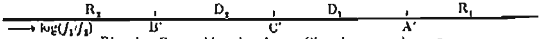
to the second if

$$f_1(x|\theta_1) \leqslant Bf_2(x\theta_2) \qquad \ldots \quad (3c.1)$$

and remain in doubt if

$$Af_2(x\theta_2) > f_1(x\theta_1) > Bf_2(x\theta_2)$$

The quantities A and B are chosen such that the probabilities of wrong decisions are at assigned levels. The diagram below (fig. 1) shows the nature of decisions that could be made after ascertaining the value of the ratio f_1/f_2 or its logarithm.



$B' = \log_e B, \quad A' = \log_e A, \quad C' = \log_e r_2\pi_2 - \log_e r_1\pi_1$

Fig. 1

In the region R_2 the individual can be asserted (at a given risk) to belong to the second group while in D_2 he can be provisionally assigned to the second group and similarly for R_1 and D_1. In such cases it may be possible to measure more characters and thus bring in further evidence to decide the issue. The procedure is analogous to Wald's theory of sequential tests.

Let there be three alternative groups with probability densities $f_1(x,\theta_1)$, $f_2(x,\theta_2)$, and $f_3(x,\theta_3)$. If π_1, π_2, π_3, are apriori probabilities and r_{ij} represents the risk in wrongly classifying an individual from the i-th group in the j-th, then the minimum risk solution leads to assigning the individual to the first group if

$$\begin{aligned} \pi_2 r_{21} f_2 + \pi_3 r_{31} f_3 &< \pi_1 r_{12} f_1 + \pi_3 r_{32} f_3 \qquad \ldots \quad (3c.2) \\ &< \pi_1 r_{13} f_1 + \pi_2 r_{23} f_2 \end{aligned}$$

to the second if

$$\begin{aligned} \pi_1 r_{12} f_1 + \pi_3 r_{32} f_3 &< \pi_2 r_{21} f_2 + \pi_3 r_{31} f_3 \qquad \ldots \quad (3c.3) \\ &< \pi_1 r_{13} f_1 + \pi_2 r_{23} f_2 \end{aligned}$$

and to the third if

$$\begin{aligned} \pi_1 r_{13} f_1 + \pi_2 r_{23} f_2 &< \pi_2 r_{21} f_2 + \pi_3 r_{31} f_3 \qquad \ldots \quad (3c.4) \\ &< \pi_1 r_{12} f_1 + \pi_3 r_{32} f_3 \end{aligned}$$

If the probabilities associated with the wrong classifications for the various groups are not small then a procedure similar to the one stated above in the case of two groups has to be followed. This leads to assigning an individual to the first group if

$$f_1 > A_1 f_2 + B_1 f_3$$

where A_1 and B_1 are chosen such that the probalbilities of individuals from the second and third groups being wrongly classified by this rule are *less than some assigned values*[1]. If these values are chosen to be small then an individual can be classified into the first group with some confidence. Similar criteria can be set up for the second and third groups. This procedure makes provision for some doubtful cases which cannot be classified into one group or the other with some confidence.

The boundaries meeting at the point O in figure 2 are defined by the relationships in (3c.2), (3c.3) and (3c.4). An individual falling in the region R_1 will be classified into the fitst group with some confidence while in D_1 he will be provisionally classified into the first group. One important thing to be noted is that the regions R_1, R_2, and R_3 are independent of the probabilities apriori.

There is some suggestion in literature (the author is also given to understand during seminars and symposiums held in Calcutta and Poona) that the determination of the region of doubt is not a statistical problem but would be made on practical considerations concerning the consequences of a wrong decision. It is also said that

[1] The reason for introducing an inequality relationship is discussed in (Rao, 1948)

indeterminableness of the risk function in certain regions of the parametric space leads to a doubtful region. In the method indicated above the doubtful region is derived merely on probability considerations. Neither apriori probabilities nor risk functions are used.

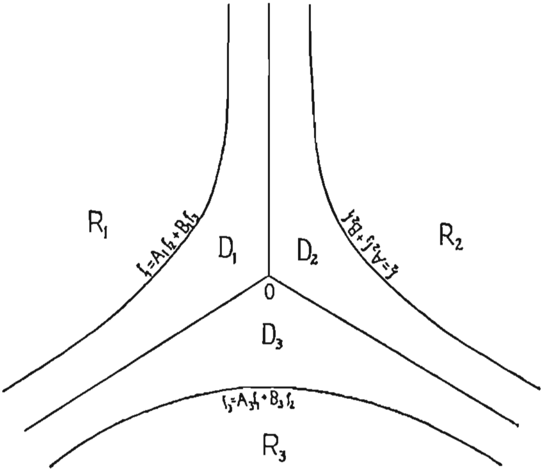


Fig. 2 Division of the space for six possible judgements.

(d) *Allocation of a number of individuals to two or more groups.*

Suppose n_1 and n_2 posts have to be filled in the Navy and Airforce, a candidate being chosen on the basis of his performance in a test. Assuming that the distribution of test scores for those who are fit for the Navy and Airforce are available from past experience how could this knowledge be used for the most efficient selection?

In the actual population the relative proportions of candidates suitable for the Navy may be different from $n_1{:}n_2$. The procedure for selection must be such that *whatever may be the actual proportion in the population* the division of a sample of individuals in the assigned ratio $n_1{:}n_2$ should involve the least possible errors. A similar problem is the allocation of a given number of skulls into two sexes in a given ratio which is determined from some apriori considerations. This may be only an estimated proportion and hence may not represent the true sex ratio. Whatever criterion is chosen some male skulls will be classified as female and vice versa. A procedure which gives the least value to the expected number of wrong classifications in either group may be regarded as the best one.

Let $f_1(x/\theta_1)$ and $f_2(x/\theta_2)$ be the probability densities in the two groups having apriori probabilities π_1 and π_2 so that the probability density in the mixed population is $\pi_1 f_1(x/\theta_1)+\pi_2 f_2(x/\theta_2)$. For the each individual supplying observations on the characters under consideration one can calculate, what may be called the discriminant score

$$S = \log_e f_1(x/\theta_1) - \log_e f_2(x/\theta_2)$$

If there are $n = n_1+n_2$ individuals one has n discriminating scores which can be arranged in descending order of magnitude. The n_1 individuals obtaining the first n_1 ranks are assigned to the first group and the rest to the second.

This procedure minimises the expected loss whatever may be the relative risk involved in wrong classification and whatever may be the apriori probabilities π_1 and π_2 applicable to the population. To prove this it is enough to show that the expected number of wrong classifications in either group is greater for any other procedure.

Consider all samples of n_1+n_2 individuals for whom the discriminant score of (n_1+1)th rank is constant say S.

If π denotes the probability

$$\pi = \int_C (\pi_1 f_1+\pi_2 f_2)\, dv$$

where C denotes the region $f_1(x/\theta_1) > e^s f_2(x/\theta_2)$, then the individuals with discriminant scores greater than S arise from a mixed population with individuals of the first and second groups in the proportions

$$a = \frac{\pi_1}{\pi}\int_C f_1\, dv,\quad 1-a = \frac{\pi_2}{\pi}\int_C f_2\, dv$$

If the procedure of assigning all individuals with discriminant scores greater than S to the first group is followed, then in samples which have S as the (n_1+1)-th rank discriminant score, the expected number of wrong classifications in the first group is

$$\frac{n_1\pi_2}{\pi}\int_C f_2\, dv$$

Any other method would choose some individuals from the first n_1 ranks and others from the rest and allocate them to the first group. If the average proportions chosen from the first n_1 ranks and the rest are denoted by γ and δ, then

$$n_1\gamma + n_2\delta = n_1$$

For any procedure which chooses a proportion γ on the average from the sub-population defined by the density

$$\frac{\pi_1 f_1+\pi_2 f_2}{\pi},\ f_1>e^{s} f_2$$

the minimum expected proportion of wrong classifications is

$$\frac{\pi_2}{\pi}\int_{C_1} f_2\, dv \qquad \dots \quad (3d.0)$$

where C_1 is defined by $f_1>e^{s_1} f_2$, S_1 being chosen to satisfy the relationship

$$\frac{1}{\pi}\int_{C_1}(\pi_1 f_1+\pi_2 f_2)\, dv = \gamma$$

The minimum expected proportion of wrong classifications for the other sub-population is

$$\frac{\pi_2}{1-\pi}\int_{C_2} f_2\, dv$$

where C_2 is defined by $e^{s_2} f_2<f_1<e^{s} f_2$, S_2 being chosen such that

$$\frac{1}{1-\pi}\int_{C_2}(\pi_1 f_1+\pi_2 f_2)\, dv = \delta$$

The minimum expected number of wrong classifications under the second procedure is not less than

$$\frac{n_1\pi_2}{\pi}\int_{C_1} f_2\, dv+\frac{n_2\pi_2}{1-\pi}\int_{C_2} f_2\, dv \qquad \dots \quad (3d.1)$$

Because of the identity

$$\gamma n_1+\delta n_2 = n_1$$

we have

$$\frac{n_1}{\pi}\int_{C-C_1}(\pi_1 f_1+\pi_2 f_2)\, dv = \frac{n_2}{1-\pi}\int_{C_2}(\pi_1 f_1+\pi_2 f_2)\, dv$$

Now $$\frac{n_2}{1-\pi}\int_{C_2}(\pi_1 f_1+\pi_2 f_2)\, dv < \frac{n_2}{1-\pi}\int_{C_2}\pi_2 f_2\left(1+\frac{\pi_1 e^{s}}{\pi_2}\right) dv$$

and $$\frac{n_1}{\pi}\int_{C-C_1}(\pi_1 f_1+\pi_2 f_2)\, dv > \frac{n_1}{\pi}\int_{C-C_1}\pi_2 f_2\left(1+\frac{\pi_1 e^{s}}{\pi_2}\right) dv$$

Since the left hand sides are equal it follows that

$$\frac{n_2}{1-\pi}\int_{C_2}\pi_2 f_2\, dv > \frac{n_1}{\pi}\int_{C-C_1}\pi_2 f_2\, dv$$

The expected loss (3d.1) is then greater than or equal to

$$\frac{\pi_1}{\pi}\int_{C_1}\pi_2 f_2\,dv+\frac{\pi_1}{\pi}\int_{C-C_1}\pi_2 f_2\,dv$$

or

$$n_1(1-\alpha)$$

which is tho expected loss for the procedure suggested earlier. For any value of tho (n_1+1)th rank discriminant score the above procedure is the best possible and hence has the least possible number of wrong classifications on the average. The loss for the other group is simultaneously at a minimum. This proves the required result.

The problem of allocation increases in complexity with increase in the number of groups. Consider only three groups with apriori probabilities π_1, π_2, π_3 and probability densities f_1, f_2, f_3. Suppose a selection of n_1 individuals has to be made for the first group. For each individual we calculate the score

$$S=\frac{\pi_1 f_1+\pi_2 f_2+\pi_3 f_3}{r_{21}\pi_2 f_2+r_{31}\pi_3 f_3}$$

where r_{21} and r_{31} denote the relative losses incurred in accepting individuals of the second and third groups for the first. The expected loss will be a minimum if the n_1 individuals corresponding to the highest values of S are chosen. It is necessary to know the apriori probabilities for a satisfactory solution of this problem while in the previous case of two groups this information is not needed. The proof is similar to that given above.

Suppose we want to set up a criterion by which, on the average, individuals are assigned to the three groups in the ratios $\rho_1:\rho_2:\rho_3$, $(\Sigma\rho=1)$. If R_1, R_2, R_3 are the regions which determine the groups of the individuals then

$$\int_{R_i}(\pi_1 f_1+\pi_2 f_2+\pi_3 f_3)\,dv=\rho_i,\ i=1,2,3. \qquad \ldots \quad (3d.2)$$

The expected loss is

$$\int_{R_1}(r_{21}\pi_2 f_2+r_{31}\pi_3 f_3)\,dv+\int_{R_2}(r_{12}\pi_1 f_1+r_{32}\pi_3 f_3)\,dv+\int_{R_3}(r_{13}\pi_1 f_1+r_{23}\pi_2 f_2)\,dv \quad (3d.3)$$

where r_{ij} denotes the loss in accepting an individual of tho ith group for the j-th group. The best regions which minimise the expected loss are defined by

$$R_1 \cap F_1 \leqslant F_2,\ F_1 \leqslant F_3$$
$$R_2 \cap F_2 \leqslant F_1,\ F_2 \leqslant F_3$$
$$R_3 \cap F_3 \leqslant F_1,\ F_3 \leqslant F_2$$

where the symbol $\cap$ stands for 'defined by' and

$$F_1=r_{21}\pi_2 f_2+r_{31}\pi_3 f_3+\lambda_1(\pi_1 f_1+\pi_2 f_2+\pi_3 f_3)$$
$$F_2=r_{32}\pi_3 f_3+r_{12}\pi_1 f_1+\lambda_2(\pi_1 f_1+\pi_2 f_2+\pi_3 f_3)$$
$$F_3=r_{13}\pi_1 f_1+r_{23}\pi_2 f_2+\lambda_3(\pi_1 f_1+\pi_2 f_2+\pi_3 f_3)$$

where λ_1, λ_2 and λ_3 are determined so as to satisfy the conditions (3d.2). This solution follows from a lemma in Appendix 1 in (Rao, 1948).

The solution given above can be used in obtaining the most efficient allocation of a number of individuals into three groups in an assigned ratio. Suppose N individuals have to be distributed into groups of n_1, n_2 and $n_3.(\Sigma n = N)$ for Navy, Airforce and Army. To start with we can try the regions defined above using the proportions $\rho_1 : \rho_2 : \rho_3 :: n_1 : n_2 : n_3$. But this will not give an exact division of the N individuals into groups of $n_1 . n_2$ and n_3 in any single sample but will do so on the average. In such a case we can try some alternative values of ρ_1, ρ_2, ρ_3 till the required numbers fall into the regions R_1, R_2, R_3. There may be many sets of ρ's for which this condition is satisfied in which case that set for which the loss (3d.3) is a minimum is chosen and the desired grouping arrived at. In practice there may be some difficulty in following this procedure. Some simpler methods will be discussed in a subsequent communication.

4. Distance Power Tests

(a) *Distance functions*

The concept of distance between two statistical populations was first developed by Mahalanobis (1930) and the generalized distance D^2 defined by him has become a useful tool in biological and anthropological researches (Rao, 1948; Mahalanobis, Majumdar & Rao, 1949; Rao & Slater 1949). Since functions defined to be distance between two hypotheses concerning a population are being used in the theory of the testing of hypotheses it is of interest to examine the logical basis of the distance functions.

The overlap function: Let $f_1(x|\theta_1)$ and $f_2(x|\theta_2)$ denote the probability densities corresponding to two populations or two hypotheses. Suppose an individual is chosen from one of these two populations and it is desired to know whether he belongs to the first or to the second. If an individual could be assigned to the proper group with certainty then we have a case of no overlap which may be considered as the maximum possible distance between any two populations. On the other hand if the two populations are identical there is no means of discriminating between the individuals of the two groups in which case the only method of classification is by a toss of the coin leading to 50% errors. This may be considered as the complete overlap leading to the minimum distance between any two populations,

The frequency of errors in any other situation lies between 0 to 50%. Let α, the probability of an individual from any group being misclassified by following the *best possible procedure*, be defined as the extent of overlap between two populations (Rao, 1947). A distance function can be defined to be any decreasing function of x. One such function is $(1-x)$ which is shown to satisfy some mathematical and emperical requirements of a distance function (Rao, 1948).

The quantity α considered above is the common value of the probabilities

$$\alpha = \int_{f_2 > \lambda f_1} f_1(x|\theta_1)\, dv = \int_{f_2 < \lambda f_1} f_2(x|\theta_2)\, dv$$

where λ is properly chosen.

The quadratic differential metric: Let us consider a population characterised by the probability density

$$\phi(x, \theta_1, \ldots, \theta_q)$$

By varying θ's we generate different populations. We shall assume that the space, in which the parameters leading to valid probability densities lie, is continuous.

If $(\theta_1, \ldots, \theta_q)$ and $(\theta_1+d\theta_1, \ldots, \theta_q+d\theta_q)$ represent two contiguous points in the parameter space the best method of discriminating between the individuals from the populations defined by them is to make use of the likelihood ratio

$$\frac{\phi(x\theta+d\theta)}{\phi(x,\theta)} = 1+\frac{d\phi}{\phi}$$

where

$$d\phi = \frac{\partial\phi}{\partial\theta_1}\, d\theta_1 + \ldots + \frac{\partial\phi}{\partial\theta_q}\, d\theta_q$$

All individuals for whom $d\phi > \lambda\phi$ will be put into one group while the others in the second. It is difficult to find the probability of wrong classifications by following the above procedure. A fair idea of this is given by the square of the difference in mean values of $d\phi/\phi$ for the two populations divided by its variance[*] which to the order of infinitesimals considered is same for both the populations.

$$V\left(\frac{d\phi}{\phi}\right) = \Sigma\Sigma g_{ij} d\theta_i d\theta_j \qquad \ldots \quad (4a.1)$$

where $g_{ij} = E\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta_i}\right)\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta_j}\right)$ is an element of the information matrix defined by R. A. Fisher.

The square of the difference in mean values is

$$(\Sigma\Sigma g_{ij}\, d\theta_i d\theta_j)^2 \qquad \ldots \quad (4a.2)$$

so that the required ratio of (4a.2)/(4a.1) is equal to

$$\Sigma\Sigma g_{ij}\, d\theta_i d\theta_j \qquad \ldots \quad (4a.3)$$

[*] In fact all the higher moments are equal to the order of infinitesimals considered so that the entire distance is contributed by the difference in a single quantity viz., the mean so that division by the variance is formal. In an earlier paper the variance of $d\phi/\phi$ was considered to obtain the distance. But the proper justification is that the metric denotes the difference in mean values of the likelihood ratio.

This quadratic differential form with its *fundamental tensor* as the elements of the Information matrix may be used as a suitable measure of the distance between two populations. If two given populations have the parameters as $\theta_1, ..., \theta_q$ and $\theta'_1, ..., \theta'_q$ then the distance between them can be obtained by integration along a geodesic using the element of length as defined by the metric (4a.3) (ref. Rao, 1945).

Angular distance: A perfectly general measure has been developed by Bhattacharya (1943) who defines the distance between two populations as the angular distance between the points representing the populations on a unit sphere. If $\pi_1, ..., \pi_k$ are the proportions in a population consisting of k classes then the population is represented by a point with co-ordinates $\sqrt{\pi_1}, ..., \sqrt{\pi_k}$ on a unit sphere of k dimensions. If two populations have the co-ordinates $\sqrt{\pi_1}, ..., \sqrt{\pi_k}$ and $\sqrt{\pi_1'}, ..., \sqrt{\pi_k'}$ then the distance between them is

$$\text{Cos}^{-1}(\sqrt{\pi_1 \pi_{1'}} + \sqrt{\pi_2 \pi_{2'}} + .. + \sqrt{\pi_k \pi_{k'}})$$

If the populations are continuous with probability densities $\phi(x)$ and $\psi(x)$ the distance is given by

$$\text{Cos}^{-1} \int \sqrt{\phi(x) \psi(x)}\, dv$$

If $\phi(x)$ and $\psi(x)$ are two contiguous populations as defined above then the actual length ds along the sphere on which $\phi(x)$ and $\psi(x)$ are represented is proportional to

$$\sqrt{\Sigma \Sigma g_{ij} d\theta_i d\theta_j}$$

which involves the metric defined above. This establishes an interesting correspodence between the angular distance and the quadratic differential metric.

It must be observed that in problems of classification of groups which involve the identification of clusters or constellations in a given configuration of groups the use of the overlap distance function has a greater appeal in that it is, to some extent, based on the organic or physical resemblance or difference between the individuals in the two groups as revealed by some observable characters.

A necessary requirement of this tool is that the variables considered should be continuous. Analogous distance functions could be defined when the variables are discrete but these are not attempted here.

The quadratic differential metric though mathematically elegant is useful only when the populations considered are of the same type and differ only in the values of the parameters which can take continuous values. Although the angular distance is free from this limitation it does not explicitly make use of the parameters in the distribution so that it is difficult to judge how far the angular distance reflects the changes in the distribution function consequent on changes in the parameters.

In tests of significance which will be discussed in sections (4c, d, & e) the distance function can be conveniently chosen from the alternatives suggested above.

(b) *A lemma on power functions*

Lemma: Let $f_1, f_2, \dots$ be a finite number of probability densities alternative to f_0 which is specified by the Null hypothesis. Let ω be any region satisfying the conditions

$$\int_\omega f_0\, dv = \alpha \qquad \dots \quad (4b.1)$$

and

$$\frac{1}{a_1}\int_\omega f_1 dv = \frac{1}{a_2}\int_\omega f_2 dv = \dots \qquad \dots \quad (4b.2)$$

where $a_1, a_2, \dots$ are positive assigned quantities.

Out of all regions satisfying the conditions (4b.1) and (4b.2), the region ω_0

inside which, $\quad f_0 < \lambda_1 f_1 + \lambda_2 f_2 + \dots$

outside which, $\quad f_0 > \lambda_1 f_1 + \lambda_2 f_2 + \dots$

where $\lambda_1, \lambda_2, \dots$ are determined such that the above conditions are satisfied, gives the highest common value to the quantities in (4b.2).

Proof: Let β and β_0 be the common values (4b.2) associated with the regions ω and ω_0 and denote by $\omega\omega_0$ the region common to ω and ω_0. Then we have

$$\begin{aligned}(\lambda_1 a_1 + \lambda_2 a_2 + \dots)\beta_0 &= \int_{\omega_0} (\lambda_1 f_1 + \lambda_2 f_2 + \dots)\, dv \\ &\geqslant \int_{\omega_0 - \omega\omega_0} f_0\, dv + \int_{\omega\omega_0} (\lambda_1 f_1 + \lambda_2 f_2 + \dots)\, dv \\ &= \int_{\omega - \omega\omega_0} f_0\, dv + \int_{\omega\omega_0} (\lambda_1 f_1 + \lambda_2 f_2 + \dots)\, dv \\ &> \int_{\omega - \omega\omega_0} (\lambda_1 f_1 + \lambda_1 f_2 + ..)\, dv + \int_{\omega\omega_0} (\lambda_1 f_1 + \lambda f_2 + ..)\, dv \\ &= \int_\omega (\lambda_1 f_1 + \lambda_2 f_2 + ..)\, dv = (\lambda_1 a_1 + \lambda_2 a_2 + ..)\beta\end{aligned}$$

If $(\lambda_1 a_1 + \lambda_2 a_2 ..)$ is positive then $\beta_0 > \beta$. To prove that $(\lambda_1 a_1 + \lambda_2 a_2 + ..)$ is positive we observe that

$$\int_{\omega_0} (\lambda_1 f_1 + \lambda_2 f_2 + ..)\, dv \geqslant \int_{\omega_0} f_0\, dv$$

i.e. $\qquad (\lambda_1 a_1 + \lambda_2 a_2 + ..)\beta_0 > \alpha$

Since β_0 and α are positive it follows that $(\lambda_1 a_1 + \lambda_2 a_2 ..)$ is necessarily positive. The lemma is proved.

This lemma gives us a method of determining a region with respect to which the powers of the various alternative hypotheses are in an assigned ratio and subject to this condition every alternative hypothesis has the maximum power.

c. *Test for a finite number of alternatives.*

Consider a null hypothesis H_0 and a series of alternatives $H_1, H_2, \ldots$ Let the power of the best possible test for H_0 when H_i is the *only alternative* be denoted by $\gamma_i(\alpha)$ where α denotes the level of significance. Any region w suggested as the critical region for testing H_0 will have

$$\int_w f_i \, dv = \beta_i(\alpha)$$

as the power for the alternative H_i. In no case can β exceed γ but there may exist a single region w_0 such that $\beta_i(\alpha) = \gamma_i(\alpha)$ for all i in which case this region is uniformly the best and no criticism can be levelled against it.

If this is not so, various alternatives have been suggested. One is to choose a region which maximises the minimum β (Neyman and Pearson, 1933b; Jackson, 1936; Wald, 1939). A procedure like this may give undue preference to the hypotheses nearer to the null hypothesis. It may be felt that a method which effectively controls the errors of not accepting a nearer hypothesis when it is true will be good enough for distant hypotheses. But robbing Peter to pay Paul is not necessarily the best solution.

On the other hand one may take the view that in the course of experimentation it is necessary to detect a distant hypothesis as early as one can. If in fact a distant hypothesis were true and the critical region had been so chosen as to give this hypothesis the maximum possible power then it could be discovered with the minimum possible number of observations. If, in fact, a nearer hypothesis were true a larger experiment would be necessary to detect it. In such a case the experimenter might consider himself unlucky on the choice of his subjct or might regard the consequences of accepting H_0 when in fact an alternative close to it is true as less serious than when the alternative is distant.

A compromise solution may be suggested if the experimenter could assign apriori probabilities for the various alternatives. This means that he has a knowledge of a series of similar experiments and frequencies of various types of alternatives he has to deal. When such a knowledge is imperfect or the experimenter is not sure that the particular experiment he is conducting belongs to the same group of experiments that have been conducted before, no unique solution is possible. In the absence of any information about the apriori probabilities, as a compromise between the two views of maximising the minimum power or giving more weight to distant hypotheses, the following solution is suggested.

The critical region w is chosen such that the common ratio

$$\frac{\beta_1(\alpha)}{\gamma_1(\alpha)} = \frac{\beta_2(\alpha)}{\gamma_2(\alpha)} = \ldots$$

is a maximum where the β's and γ's are as defined above. This method supplies a system of weights to be attached to the powers due to various alternatives, the weights being the individual maximum powers. This region has the following two properties.

(a) The distant hypotheses have necessarily more power than the nearer hypotheses

(b) The individual maximum powers are now reduced by the same proportion with the provision that this proportion is as small as possible.

If $f_0, f_1, f_2 \ldots$ denote the probability densities for the hypotheses $H_0, H_1, H_2 \ldots$ then the region satisfying the above requirements is deducible from the lemma proved in section 4b. The boundary of this region ω is defined by

$$f_0 = \lambda_1 f_1 + \lambda_2 f_2 +$$

where $\lambda_1, \lambda_2, \ldots$ are determined from the relations

$$\int_{\omega} f_0 \, dv = \alpha \qquad \ldots \quad (4c.2)$$

and

$$\frac{1}{\gamma_1(\alpha)} \int_{\omega} f_1 \, dv = \frac{1}{\gamma_2(\alpha)} \int_{\omega} f_2 \, dv = \ldots \qquad \ldots \quad (4c.3)$$

The solution deduced above is not useful in practice because of the difficulty in evaluating the constants. It may be convenient to consider the region complimentary to

$$f_0 > \mu_i f_i, \quad i = 1, 2, \ldots \qquad \ldots \quad (4c.4)$$

as the critical region, the quantities $\mu_1, \mu_2, \ldots$ being determined to satisfy the relations (4c.2) and (4c.3).

Example: Consider two independent normal variates x and y with unit variances. The hypothesis to be tested is $E(x) = E(y) = 0$. Two possible alternatives are

$$H_1, \quad E(x) = E(y) = m$$
$$H_2, \quad E(x) = -E(y) = m$$

where $m > 0$, the exact value being unknown.

For any given m, the maximum powers of the individual hypotheses are the same. The inside of the best region ω for an assigned m is determined by

$$f_0 < \lambda_1 f_1 + \lambda_2 f_2 \qquad \ldots \quad (4c.4)$$

where λ_1 and λ_2 are chosen such that

$$\int_{\omega} \exp -\tfrac{1}{2}(x^2+y^2) \, dx dy = 2\pi\alpha \qquad \ldots \quad (4c.5)$$

$$\int_{\omega} \exp -\tfrac{1}{2}[(x-m)^2+(y-m)^2] \, dx dy = \int_{\omega} \exp -\tfrac{1}{2}[(x-m)^2+(y+m)^2] \, dx dy$$

Two things are to be noted, (1) the region defined in (4c.4) is not independent of m so that no valid test is possible and (2) it is difficult to determine the value of λ_1 and λ_2 satisfying the conditions (4c.5). On the other hand, if the simpler type of regions suggested above is followed we obtain the region a complimentary to the critical region as

$$\exp-\tfrac{1}{2}(x^2+y^2) > \mu_1 \exp-\tfrac{1}{2}[(x-m)^2+(y-m)^2]$$
$$> \mu_2 \exp-\tfrac{1}{2}[(x-m)^2+(y+m)^2]$$

which on taking logarithms and simplifying reduces to

$$\begin{aligned} x+y &\leqslant v_1 \\ x-y &\leqslant v_2 \end{aligned} \qquad \ldots \quad (4c.6)$$

where v_1 and v_2 are to be determined such that in the region complimentary to that defined by (4c.6) the conditions (4c.5) hold good. We thus obtain a region independent of the unknown parameters. By symmetry it can be seen that $v_1 = v_2$ so that the critical region is the shaded portion beyond the boundary lines (4c.6) shown in figure 3.

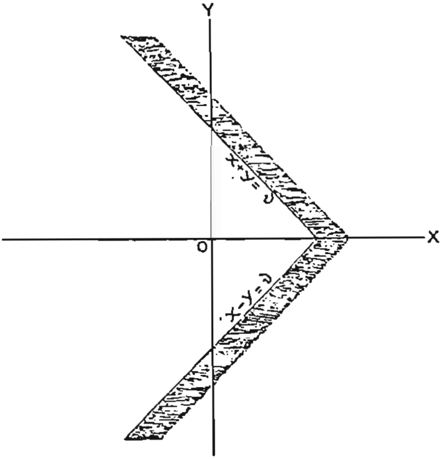


Fig. 3 The region of rejection

The value of v has to be determined such that the integral of the probability density under the null hypothesis in the shaded region is equal to α an assigned level of significance. In the above case the variables $(x+y)$ and $(x-y)$ are independent so that

$$1-P[(x+y)<\nu, (x-y)<\nu] = 1-\{P(x+y<\nu)\}^2$$

If this is equated to .05 then

$$P(x+y<\nu)=\sqrt{.95}$$

Observing that $(x+y)$ is a normal variate with variance 2, we find from normal tables the value of $\nu = 1.954$ for which the above relation is true.

If x and y have different variances say σ_1^2 and σ_2^2 then the region ω is defined by

$$\frac{x}{\sigma_1^2}+\frac{y}{\sigma_2^2}<v_1, \quad \frac{x}{\sigma_1^2}-\frac{y}{\sigma_2^2}<v_2$$

Again by symmetry it can be seen that $v_1 = v_2$. Now if

$$z_1 = \frac{x}{\sigma_1^2}+\frac{y}{\sigma_2^2}, \quad z_2 = \frac{x}{\sigma_1^2}-\frac{y}{\sigma_2^2}$$

then

$$V(z_1) = \frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2} = V(z_2)$$

$$\mathrm{Cov}\,(z_1z_2) = \frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}$$

so that the correlation between z_1 and z_2 is $(\sigma_2^2-\sigma_1^2)/(\sigma_1^2+\sigma_2^2)$. For any given correlation one can find the value of λ such that

$$P(z_1<\lambda\sqrt{\sigma_1^2+\sigma_2^2}/\sigma_1\sigma_2, z_2<\lambda\sqrt{\sigma_1^2+\sigma_2^2}/\sigma_1\sigma_2) = .05$$

The values of λ for various values of the correlation coefficients are given in table 1.

TABLE 1. 5% SIGNIFICANT VALUES OF λ FOR POSITIVE AND NEGATIVE CORRELATIONS

ρ	λ	ρ	λ	ρ	λ	ρ	λ
.05	1.953	.55	1.909	−.05	1.956	−.55	1.960
.10	1.951	.60	1.900	−.10	1.957	−.60	1.960
.15	1.949	.65	1.889	−.15	1.958	−.65	1.960
.20	1.946	.70	1.877	−.20	1.959	−.70	1.960
.25	1.943	.75	1.863	−.25	1.959	−.75	1.960
.30	1.939	.80	1.846	−.30	1.959	−.80	1.960
.35	1.934	.85	1.825	−.35	1.960	−.85	1.960
.40	1.929	.90	1.798	−.40	1.960	−.90	1.960
.45	1.923	.95	1.758	−.45	1.960	−.95	1.960
.50	1.916	1.00	1.645	−.50	1.960	−1.00	1.960

$\rho=0, \quad \lambda=1.954$

If z_1 and/or z_2 exceeds $h\sqrt{\sigma_1^2+\sigma_2^2}/\sigma_1\sigma_2$ then the null hypothesis H_0 is rejected.

This test which has a genetical application was suggested to me by R.A. Fisher.

(d) *Tests when the alternatives are continuous*

The foregoing theory could be extended to the case where the alternatives can be specified by parameters with continuous variaton. The following definitions will be useful.

A region ω which gives equal power to all hypotheses *equidistant* (in the sense in which distance is defined in secton 4a) from the null hypothesis is called the *distance power* region. A test based on a distance power region ω_0 is said to be uniformly the best distance power test[*] if

(i) the size of the region ω_0 with respect to the null hypothesis is α (an assigned value)

(ii) ω_0 is a distance power region, and

(iii) for any specified alternative hypothesis the power associated with the region ω_0 is not less than the power for any other region satisfying the requirements (i) and (ii).

Let Δ denote the distance of a hypothesis H from H_0 the null hypothesis. Then a distance power region satisfies the condition

$$\int_\omega f_H\, dv = \phi(\Delta), \text{ a function of } \Delta \text{ only}$$

If the parameters entering in the alternative hypothesis be denoted symbolically by θ and in the null hypothesis by θ_0, then

$$\int_\omega f(\theta)\, dv = \phi(\Delta) \text{ and } \int_\omega f(\theta_0)\, dv=\alpha \qquad \dots \quad (4d.1)$$

Let us define the inside of the region ω_0 by

$$f(\theta_0) \leqslant \int_{\Delta=\text{const.}} \lambda(\theta) f(\theta)\; dS \qquad \dots \quad (4d.2)$$

where the integral is taken over the surface Δ=constant. Let there *exist* a function $\lambda(\theta)$ such that the con litions (4:l.1) are satisfied. The region ω_0, if it exists, is the best distance power region for alternatives on the surface Δ=constant. This follows from the lemma of section 4b extended to an infinite set of alternatives. If the relationship (4d.2) is independent of the alternative used then we obtain a uniformly best distance power test. It is seen that the region (4d.2) is same as the region which has the best average power for alternatives on the surface Δ=constant and for an *assigned apriori* probability density $\lambda(\theta)$ of the parameters (Nandi, 1947). While in the theory of average power tests there is no justification for choosing a particular type of

[*] An example of such a test which has wide practical applications is discussed in a note following this article. (Sankhya, Vol. 10, Pt. 3, p. 257)

the density function $\lambda(\theta)$ on which the test generally depends, the function $\lambda(\theta)$ is suitably determined in constructing distance power tests. The determination of such a function, even if its existence is known, may be a difficult task. When once it is determined by trial or otherwise the optimum property of the test is immediately established.

The region defined in (4d.2) may depend on other parameters which are not directly involved in the null hypothesis but occur in the specification of the probability density. If the true values of such parameters are not known, then the best region must satisfy the further conditions that it is similar to the sample space with respect to these parameters. Thus the theory of composite hypotheses could be brought in to construct a valid region. The classical and studentised D^2 statistics considered by Hsu (1941) and Nandi (1947) can be shown to be uniformly best distance power tests because the regions associated with them can be expressed as

$$f(\theta_0) < \int_{\Delta=\text{const.}} \lambda(\theta) f(\theta)\, dS$$

It is of interest to examine the critical region obtained by extending the results in (4c.4) to the case of alternative hypotheses specified by parameters with continuous variation. The outside of such a critical region is defined by

$$f(\theta_0) \geqslant \lambda(\theta, \Delta) f(\theta) .$$

for all θ on the surface $\Delta(\theta) = \Delta$ where Δ is the specified distance of the alternative from the null hypothesis. If due to considerations of symmetry the function $\lambda(\theta, \Delta)$ could be replaced by a function of Δ only then the critical region is the outside of the envelope of the surfaces

$$f(\theta_0)/f(\theta) = \text{const.}$$

for variations in θ on the surface $\Delta(\theta) = \Delta$. This is the likelihood ratio test developed by Neyman and Pearson (1928).

e. *Neyman and Pearson's region of type C with a suitable ellipsoid of equidetectability*

Let

$$\frac{\partial f(\theta)}{\partial \theta_i} = f^i(\theta) \text{ and } \frac{\partial^2 f(\theta)}{\partial \theta_i \partial \theta_j} = f^{ij}(\theta)$$

If w is any region such that

$$\int_w f(\theta_0)\, dv = \alpha \qquad \ldots \quad (4e.1)$$

$$\int_w f^i(\theta)\, dv = 0,\ i = 1, 2, \ldots \qquad \ldots \quad (4e.2)$$

then the power of the test associated with w at the alternative hypothesis $(\theta_1^* + d\theta_1,$

$\theta_2^0+d\theta_2,\ldots$) in the neighbourhood of the null hypothesis (θ_1^0, $\theta_2^0,\ldots$) is given by

$$\Sigma\Sigma\beta_{ij}(\theta_0)d\theta_i d\theta_j \qquad \ldots \quad (4c.3)$$

where

$$\beta_{ij} = \int_w f^{ij}(\theta_0)\ dv$$

If w can be determined such that the expression (4c.3) is proportional to the quadratic differential metric

$$\Sigma\Sigma g_{ij}d\theta_i d\theta_j$$

defined in section (4a), then the test associated with w has equal power for all neighbouring hypotheses equally distant from the null hypothesis. This means

$$\beta_{ij}/g_{ij} = \text{constant for all } i \text{ and } j \qquad \ldots \quad (4c.4)$$

By using the lemma in section 4b it may now be shown that the inside of the region w is given by

$$f(\theta_0) < \Sigma\Sigma\lambda_{ij}f^{ij}(\theta_0)+\Sigma\lambda_i f^i(\theta_0)$$

and the outside by $\qquad \ldots \quad (4c.5)$

$$f(\theta_0) > \Sigma\Sigma\lambda_{ij}f^{ij}(\theta_0)+\Sigma\lambda_i f^i(\theta_0)$$

where λ_i and λ_j are determined to satisfy the conditions (4c.1), (4c.2) and (4c.4).

It is easy to see that the region determined by (4c.5) is invariant under transformations of the parameters because of the invariance of the quadratic differential metric. This is an important property of the distance power region such as the one deduced above. This is not true of the general unbiased critical region of the type C given by Neyman and Pearson (1938). They say (on p.44), "..when deciding to apply an unbiased critical region of type C to test a hypothesis H_0 it is necessary to be quite clear about the system of parameters which it is most appropriate to adopt in any practical problem. The choice of such a system lies clearly beyond the bounds of the theory of statistics and must be made in accordance with the practical importance of errors which it is desired to avoid when testing a particular hypothesis". In the method suggested above the subjective element in the choice of parameters is eliminated.

I wish to express my thanks to Prof V. M. Dandekar for going through the manuscript and making some suggestions.

APPENDIX

Suppose that individuals are being drawn from a mixed population with probablity density

$$\pi_1 f_1 + \pi_2 f_2$$

If the method of selection is such that on an average a proportion γ of the individuals are identified as belonging to the first group what is the minimum error committed.

The most general selection procedure is one which chooses a proportion $p(x_1, \ldots\ldots, x_p) = p(x)$ of individuals with the same measurements $(x_1, \ldots\ldots x_p)$ for assigning to the first group. By definition

$$\int_w (\pi_1 f_1 + \pi_2 f_2)\, p(x)\, dv = \gamma$$

where w represents the whole space. The proportion of correct classifications is

$$\int_w \pi_1 f_1\, p(x)\, dv$$

How can $p(x)$ be chosen such that the above integral is a maximum?

Let c be a region inside which $\pi_1 f_1 > (\pi_2 f_1 + \pi_2 f_2)$ where λ is chosen such that

$$\int_c (\pi_1 f_1 + \pi_2 f_2)\, dv = \gamma$$

For any general $p(x)$

$$\int_w \pi_1 f_1\, p\,(x)\, dv$$

$$= \int_c \pi_1 f_1\, p(x) dv + \int_{w-c} \pi_1 f_1\, p(x)\, dv$$

$$< \int_c \pi_1 f_1\, p(x) dv + \int_w \lambda(\pi_1 f_1 + \pi_2 f_2)\, p(x) dv$$

$$- \int_c \lambda\,(\pi_1 f_1 + \pi_2 f_2)\, p(x) dv$$

$$= \int_c \pi_1 f_1\, p(x)\, dv + \int_c \lambda\,(\pi_1 f_1 + \pi_2 f_2)\, [1 - p(x)] dv$$

$$< \int_c \pi_1 f_1\, p(x) dv + \int_c [1 - p(x)]\, \pi_1 f_1\, dv$$

$$= \int_c \pi_1 f_1\, dv$$

This shows that the function

$$p(x)=1 \text{ when } \pi_1 f_1 > \lambda(\pi_1 f_1 + \pi_2 f_2)$$
$$p(x)=0 \text{ elsewhere}$$

leads to the maximum number of correct classifications. This result is used in deriving the equation (3d. 0) in section 3d. The proof given above refers to a most general case and includes the important lemma of Neyman and Pearson on critical regions.

References

1. A. Bhattacharyya. (1943): On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Cal. Math. Soc.*, 35, 99.

2. Fisher, R. A. (1947): Design of experiments. Fourth edition. Oliver and Boyd. Edinburgh.

3. Fisher, R. A., Lyon, M. F. & Owen, A. R. G. (1947): Sex chromosome in house mouse. *Heredity*, 2, 355.

4. Hsu, P. L. (1941): Analysis of variance from the power function stand point. *Biometrika*, 32, 62.

5. Jackson, R. W. B. (1936): Tests of statistical hypotheses in the case when the set of alternatives is discontinuous illustrated on some genetical problems. *Stat. Res. Mem.*, 1, 138.

6. Jeffreys, H. (1948): Theory of Probability. Oxford University Press. Oxford.

7. Mahalanobis, P. C., (1930): On tests and measures of group divergence. *Jour. Proc. Asiatic Soc. Bengal*, 26, 541.

8. Mahalanobis, P. C. Majumdar, D. N. & Rao, C. R. (1949): Anthropometric survey of United Provinces, 1941: a statistcal study. *Sankhya*, 9, 90.

9. Nandi, H. K. (1947): On the average power of test criteria. *Sankhya*, 8, 67.

10. Neyman, J. & Pearson, E. S. (1928): On the use and interpretation of certain test criteria for purpose of statistical inference. *Biometrika*, 20A, 175 and 263.

11. ——— (1933a): On the problem of most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc.* A, CCXXXI, 289.

12. ——— (1933b): The testing of statistical hypotheses in relation to the probabilities apriori. *Proc. Cam. Phil. Soc.*, 29, 492.

13. ——— (1936): Contributions to the theory of testing statistical hypotheses. *Stat. Res Mem.* 1, 1 & 25.

14. Rao, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Calcutta Mathematical Bulletin*, 37, 81.

15. ——— (1947)) The problem of classification and the distance between two populations. *Nature*, 159, 30.

16 ——— (1948): The Utilization of multiple measurements in problems of biological classification. *Jour. Roy. Stat. Soc.*Series B, 10, 159.

17. ——— (1950): Symposium on time series, Abstract. *Journal Indian Math. Soc.* (in Press).

18. Rao, C. R.& Slater, P. (1949): Multivariate analysis applied to differences between neurotic groups. *British Jour. of Psychology* (Statistical Section) 2, 17.

19. Wald, A. (1939): Contributions to the theory of statistical estimation and testing of hypotheses. *Ann. Math. Stat.*, 10, 299.

20. Wald, A. (1945): Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16, 1.

Paper received July, 1950.