CHOICE OF ONE AMONG SEVERAL STATISTICAL HYPOTHESES

By Ralph J. Brookner[1]

New York City

1. Introduction. Statistical decision is a term which we will apply to that phase of statistical inference which deals with the following question. Consider one or several variates whose distribution function depends on one or several unknown parameters; suppose there be given a finite number of mutually exclusive hypotheses regarding the parameters, whose totality completely exhausts every possibility. If a sample of observations on the variates is made, the choice of one of the given hypotheses on the basis of that sample is called a statistical decision. In other words, to make a statistical decision is to give a procedure which will divide the sample space into as many regions as there are given hypotheses, and to set up a one-to-one correspondence between these regions and the hypotheses so that if the sample point lies in any particular region, the corresponding hypothesis is chosen.

This notion is quite closely connected with both of the fields of statistical inference that have engaged most of the modern statistical theorists. On the one hand, it may be considered a generalization of the notion of testing hypotheses, for in this theory, one gives a procedure which divides the sample space into a region of rejection and a region of non-rejection of a given null hypothesis. Then one makes either of two decisions depending upon which of the regions contains the sample point. On the other hand, the theory of estimation is a generalization of the notion of statistical decision in which the number of alternatives is not restricted to be finite

As in any phase of statistical inference, our primary aim is to define broad principles upon which "good" or "best" procedures for making statistical decisions may be based. The general problem of statistical decisions has been formulated by A. Wald, who has also proposed a principle on which the solution can be based. We are interested, however, in several of the simpler but important particular problems in which quite serious calculation difficulties are encountered in actually finding Wald's solution. Hence, we will propose in its stead another principle which quite closely resembles Wald's for selecting a solution of the problem of statistical decision.

It may be pointed out immediately that, from a purely logical point of view, the substitute principle we shall offer will probably be considered to be less acceptable than its predecessor. We will find, however, by considering its application to some of the well known problems of testing hypotheses, that the principle is at least reasonable in leading to certain well accepted results.

[1] Research under a grant-in-aid of the Carnegie Corporation of New York.

2. Principle determining the "best" procedure. We will first discuss briefly Wald's principle and the definition of the criterion that we will employ will be accomplished by pointing out the differences. A much more general formulation is possible [1], [2], but we will discuss the principle as it will be directly applied to the problems of statistical decisions when the number of hypotheses is finite

Consider the variates $x_1, x_2, \cdots, x_p$ whose probability density function $f(x_1, x_2, \cdots, x_p \mid \theta_1, \theta_2, \cdots, \theta_k)$ is known except for the unknown values of the parameters $\theta_1, \theta_2, \cdots, \theta_k$. We denote by θ a point in k-dimensional space whose coordinates are $(\theta_1, \theta_2, \cdots, \theta_k)$ and shall speak of this parameter space as Ω. Suppose that ω is any subset of Ω and that S represents a system of finitely many such sets which are mutually disjunct and which cover Ω. Each element, ω_0, of S corresponds to a hypothesis H_{ω_0}, which is the hypothesis that θ is a point of ω_0, and the system of all such hypotheses corresponding to S we denote by H_S.

A sample of N observations on $x_1, x_2, \cdots, x_p$ is drawn and the sample may be considered as a point, E, in the pN dimensional sample space; denote the sample space by M. We want to decide on the basis of the point E which of the hypotheses of H_S should be accepted. That is, we seek a procedure by which the sample space may be divided into a system of mutually exclusive regions M_ω which are the same in number as the number of elements of S, and by which a correspondence is set up so that the falling of the sample point into a particular M_{ω_0} shall cause us to accept a particular hypothesis H_{ω_0} as the true one. If the totality of regions M_ω be denoted M_S, it is necessary to give a principle by which we may prefer a particular system M_S over any other system M'_S.

Wald introduces the notion of a weight function of errors, a function of the parameters and of the decision made, which might well be defined as the loss incurred if θ be the true parameter point and the sample point falls in M_ω which causes us to accept the hypothesis H_ω. Denote the weight function by $W(\theta, \omega_E)$ where ω_E stands for that hypothesis which we choose if E is the sample point; then we require that $W(\theta, \omega_E)$ be non-negative, and if θ lies in ω_E, $W(\theta, \omega_E) = 0$ for then the correct decision has been made and there is no loss.

Perhaps the notion of a weight function can be most clearly understood, and its importance appreciated, if we consider the place of statistics in the business world, where possible losses are often computable in terms of money. The weight function may be taken to be equal to this loss. Suppose a manufacturing plant has a process which manufactures a product whose efficiency is a measurable quantity that we will denote by x. Suppose x is a random variable whose distribution depends only upon its mean value θ, and the company contemplates renewing its machinery if the mean value of the efficiency falls short significantly from a particular value θ_0. Then on the basis of a sample of N observations on x, one of two decisions must be reached: the rejection of the hypothesis $\theta \geqq \theta_0$ (the decision to renew the machinery), or the non-rejection of $\theta \geqq \theta_0$ (the decision not to renew it). Suppose the region M_ω is the region of the sample space such that if E falls into M_ω, we reject $\theta \geqq \theta_0$ and $M_{\bar{\omega}}$ is the complementary region. Then we may say that the weight function can be defined by

$$
\begin{aligned}
W(\theta, \bar{\omega}) &= 0 && \text{fcr } \theta \geqq \theta_0 \\
W(\theta, \bar{\omega}) &= g(\theta) && \text{for } \theta < \theta_0 \\
W(\theta, \omega) &= 0 && \text{for } \theta < \theta_0 \\
W(\theta, \omega) &= h(\theta) && \text{for } \theta \geqq \theta_0
\end{aligned}
$$

where $h(\theta)$ is the company's monetary loss in needlessly changing its machinery and $g(\theta)$ is a function which expresses the company's loss in not changing its process even though the true value of the parameter is $\theta < \theta_0$. The function $g(\theta)$ may be of almost any form, but it is only reasonable that it should be a monotonic non-decreasing function of $|\theta_0 - \theta|$, since the loss should, it seems, increase as the true value of θ is farther from θ_0.

Wald then defines the risk as the expected value of the loss; since θ is an unknown, the risk will be a function of θ, and it will also be a function of the system M_S:

$$r(\theta, M_S) = \int_M W(\theta, \omega_E) \cdot f(E \mid \theta) \, dE.$$

According to Wald, the "best" system of regions, M_S, is that system for which the maximum of the risk function with respect to the parameter θ is a minimum with respect to all possible systems, M'_S, of regions. Several important properties are enjoyed by the system of regions defined in this way, though other reasonable definitions are possible. Perhaps the criterion of minimizing an average with respect to θ of $r(\theta, M_S)$ rather than the maximum may be considered more plausible, but such definitions would raise the question of which average should be used, and the result obtained by using any particular average would not be invariant with respect to transformations of the parameter space.

Using the notations as introduced above, and introducing the notation $W(\theta, \omega_i)$ to be the weight function if the ith hypothesis is chosen, the principle which we will use to solve some of the problems of statistical decisions can be given as follows: In place of the risk function, we consider the s functions

$$R_i(\theta, E) = W(\theta, \omega_i) \cdot f(E \mid \theta) \qquad (i = 1, 2, \cdots, s)$$

where $f(E \mid \theta)$ is a notation for the probability density, and s is the number of given hypotheses. If we denote by $\bar{R}_i(E)$ the least upper bound of $R_i(\theta, E)$ with respect to θ, then we choose the system of "best" regions of acceptance by *including each sample point E in a region M_i determined such that for all E_0 in M_i, $\bar{R}_i(E_0) \leqq \bar{R}_j(E_0)$ for all $j \neq i$.*

It is interesting to note that a rather general case exists in which the principle is exactly equivalent with the test of a hypothesis based upon the likelihood ratio principle. Consider the distribution function $f(x_1, x_2, \cdots, x_p \mid \theta_1, \theta_2, \cdots, \theta_k)$ which is a bounded function of the x's and θ's. Suppose we are interested in the test of the hypothesis $(\theta_1, \theta_2, \cdots, \theta_k) \in \omega$ where ω is a closed set of points of the parameter space which does not contain any open subset of the parameter space. Furthermore assume that for each set of x's the distribution function is continuous in $\theta_1, \cdots, \theta_k$ on an open subset of Ω containing ω.

We will show that the principle will lead to the test based on the likelihood ratio if the following is the weight function:

I. If ω is accepted, the loss is zero if the true parameter point is in ω, and the loss is a constant c_1 if the true parameter point is not in ω.

II. If ω is rejected (i.e. $\bar{\omega}$ is chosen), the loss is zero if the true parameter point is in $\bar{\omega}$ and is a constant c_2 if the true parameter point is in ω.

Consider then the region of the sample space for which ω is rejected according to the principle. This region is that for which

$$\text{l.u.b. w.r.t. } \theta \text{ in } \omega \text{ of } [c_2 f(x \mid \theta)] < \text{l.u.b. w.r.t. } \theta \text{ in } \bar{\omega} \text{ of } [c_1 f(x \mid \theta)]$$

where we have set $f(x \mid \theta) = f(x_1, x_2, \cdots, x_p \mid \theta_1, \theta_2, \cdots, \theta_k)$, and where l.u.b. w.r.t. means "least upper bound with respect to." But the left-hand member of this inequality is equal to

$$c_2[\text{l.u.b. w.r.t. } \theta \text{ in } \omega \text{ of } f(x \mid \theta)]$$

and because of the restriction on ω and the continuity of f, we can see that the l.u.b. of $f(x \mid \theta)$ with respect to all θ in $\bar{\omega}$ must coincide with the l.u.b. of the function with respect to all θ in Ω, which is the total parameter space. Thus we have that the hypothesis ω is rejected when

$$c_2[\text{l.u.b. w.r.t. } \theta \text{ in } \omega \text{ of } f(x \mid \theta)] < c_1[\text{l.u.b. w.r.t. } \theta \text{ in } \Omega \text{ of } f(x \mid \theta)]$$

or when

$$\frac{\text{l.u.b. w.r.t. } \theta \text{ in } \omega \text{ of } f(x \mid \theta)}{\text{l.u.b. w.r.t. } \theta \text{ in } \Omega \text{ of } f(x \mid \theta)} < \frac{c_1}{c_2}.$$

The left hand member of this inequality is the likelihood ratio statistic introduced by Neyman and Pearson [3]; hence our test is exactly equivalent with the likelihood ratio test where the size of the critical region is determined by c_1 and c_2.

We pose the following quite hypothetical example to show circumstances under which the principle proposed is reasonable. The principle does not exactly apply as it was stated in terms of probability densities and the example involves discrete probabilities, but the logic seems somewhat applicable. Suppose a game is played which consists of the player's guessing the number of white balls in an urn known to contain 10 balls, each of which is either white or black, on the basis of a sample of four drawings with replacements from the urn. Let us assume that there are eleven mutually exclusive hypotheses (as to the number of white balls in the urn) to choose among, and the player must make a choice of one of them after observing the drawing which can give 16 different results. Assume that the one who plays the game pays a banker a varying sum of money if he makes a wrong decision and that the banker has the privilege of choosing the population (i.e. the number of white and black balls originally in the urn). Now on the basis of the assumption that the banker knows the player's decision function and will attempt to fix the population so as to make the player's expected loss a maximum, it is clear that Wald's principle, which minimizes the maximum loss, leads to the best way to play the game.

Now suppose that instead of one player making the choice among the decisions, we have 16 players participating in the game and the first player is to make the choice if, and only if, the drawing is $WWWW$, the second player if the drawing is $WWWB$, and so on, where W stands for the drawing of a white ball and B for the drawing of a black one. In this case, if player x assumes that the banker will try to choose the population most unfavorable to him, then his decision function based on the new principle is the best method of play.

Although the example indicates that in the usual case which would come up in practice, Wald's principle would lead to the better procedure, since the statistician is usually faced with the necessity of giving a decision no matter what the sample point is, the new principle is useful since one may hope that in many practical cases the two principles will not lead to widely varying results, especially if the sample is large.

3. Application of the criterion to the case of testing the mean of a normal distribution. Now we will show that the criterion will lead to the widely used test of "Student's hypothesis." Suppose x is known to be distributed normally with unknown mean μ and unknown variance σ^2. On the basis of a sample of N independent observations $x_1, x_2, \cdots, x_N$, "Student's t" is used to test the hypothesis $\mu = 0$. If $\bar{x}$ is the arithmetic mean of the N observations and s^2 the usual sample estimate of the variance, then with $t = \sqrt{N}\,\bar{x}/s$, the hypothesis is to be rejected if $|t| \geqq t_0$ where t_0 is a critical value at some chosen level of significance α obtained from the distribution of t under the null hypothesis. We will use the notation ω_1 for the set of points $\mu \neq 0$ and ω_2 for the set of points $\mu = 0$.

We will consider the problem in reference to the particular weight function defined as follows:

$$\begin{aligned} W(\mu, \sigma; \omega_2) &= (\mu/\sigma)^k \qquad \text{for } \mu \neq 0 \\ W(0, \sigma; \omega_1) &= W \\ W(\mu, \sigma; \omega_1) &= 0 \qquad \text{for } \mu \neq 0 \\ W(0, \sigma; \omega_2) &= 0 \end{aligned}$$

where as a matter of convenience, we will take k an even positive integer in order to avoid the introduction of the absolute value of μ/σ which is necessary if k is an odd integer. We also take $k \leqq N$.

The density function of the sample of N observations is

$$\frac{C}{\sigma^N} \cdot e^{-(1/2\sigma^2)S(x_\alpha - \mu)^2}$$

where C is a constant. Then the two functions $R_i(\theta, E)$ are

$$R_1(\theta, E) = \frac{WC}{\sigma^N} \cdot e^{-(1/2\sigma^2)Sx_\alpha^2} \qquad \text{if } \mu = 0$$

$$R_1(\theta, E) = 0 \qquad \text{if } \mu \neq 0$$

$$R_2(\theta, E) = \frac{C\mu^k}{\sigma^{N+k}} \cdot e^{-(1/2\sigma^2)S(x_\alpha - \mu)^2} \qquad \text{if } \mu \neq 0$$

$$R_2(\theta, E) = 0 \qquad \text{if } \mu = 0.$$

To maximize $R_1(\theta, E)$, we set

$$\frac{\partial R_1(\theta, E)}{\partial \sigma} = \left[\frac{-NW}{\sigma^{N+1}} + \frac{WSx_\alpha^2}{\sigma^{N+3}}\right] Ce^{-(1/2\sigma^2)Sx_\alpha^2} = 0$$

which gives

$$\sigma^2 = \frac{Sx_\alpha^2}{N}$$

hence

$$\bar{R}_1(E) = \frac{CWN^{\frac{1}{2}N}}{(Sx_\alpha^2)^{\frac{1}{2}N}} \cdot e^{-\frac{1}{2}N}.$$

To maximize $R_2(\theta, E)$, we set

$$\frac{\partial R_2(\theta, E)}{\partial \mu} = \left[k + \frac{\mu}{\sigma^2} S(x_\alpha - \mu)\right] \frac{C\mu^{k-1}}{\sigma^{N+k}} \cdot e^{-(1/2\sigma^2)S(x_\alpha - \mu)^2} = 0$$

and

$$\frac{\partial R_2(\theta, E)}{\partial \sigma} = \left[-N - k + \frac{S(x_\alpha - \mu)^2}{\sigma^2}\right] \frac{C\mu^k}{\sigma^{N+k+1}} e^{-(1/2\sigma^2)S(x_\alpha - \mu)^2} = 0$$

which give the two relations

$$\sigma^2 = -\frac{\mu}{k} S(x_\alpha - \mu)$$

and

$$\sigma^2 = \frac{1}{N + k} S(x_\alpha - \mu)^2.$$

Then

$$-\mu(N + k)S(x_\alpha - \mu) = kS(x_\alpha - \mu)^2$$

or

$$\mu^2 - \mu\bar{x}(1 - k/N) - (k/N^2)Sx_\alpha^2 = 0$$

which gives the maximizing value of

$$\mu^* = \frac{\bar{x}(1 - k/N) \pm \sqrt{\bar{x}^2(1 - k/N)^2 + (4k/N^2)Sx_\alpha^2}}{2}$$

and it can easily be shown that the maximum is reached for the value of μ^* using the $+$ sign when $\bar{x}$ is positive and the $-$ sign when $\bar{x}$ is negative. We will carry through the case $\bar{x} > 0$ only as the case $\bar{x} < 0$ follows in a similar manner. We have

$$\bar{R}_2(E) = \frac{(\mu^*)^k k^{\frac{1}{2}(N+k)} C}{(\mu^*)^{\frac{1}{2}(N+k)}[-S(x_\alpha - \mu^*)]^{\frac{1}{2}(N+k)}} \cdot e^{-\frac{1}{2}(N+k)}.$$

To find the region of the sample space for which we should accept the hypothesis $\mu \neq 0$ (i.e. the critical region for rejection of the hypothesis $\mu = 0$), we seek those points E for which $\bar{R}_1(E) \leqq \bar{R}_2(E)$, i.e. those for which

$$\frac{WN^{\frac{1}{2}N}}{(Sx_\alpha^2)^{\frac{1}{2}N}} \leqq \frac{(\mu^*)^k k^{\frac{1}{2}(N+k)}}{(\mu^*)^{\frac{1}{2}(N+k)}[-S(x_\alpha - \mu^*)]^{\frac{1}{2}(N+k)}} \cdot e^{-\frac{1}{2}k}$$

or for which

$$\frac{(\mu^*)^{\frac{1}{2}(N-k)}[-S(x_\alpha - \mu^*)]^{\frac{1}{2}(N+k)}}{(Sx_\alpha^2)^{\frac{1}{2}N}} \leqq c$$

where c is a positive constant. Since both sides of the inequality are positive, this inequality is equivalent to

$$(1) \qquad \frac{(\mu^*)^{N-k}(\mu^* - \bar{x})^{N+k}}{(Sx_\alpha^2)^N} \leqq c_1$$

where c_1 is another positive constant.

Now we consider the statistic

$$T^2 = \frac{t^2}{N-1} = \frac{N\bar{x}^2}{Sx_\alpha^2 - N\bar{x}^2} = \frac{N}{Sx_\alpha^2/\bar{x}^2 - N}$$

from which we have

$$Sx_\alpha^2/\bar{x}^2 = (N/T^2) + N.$$

Also note that

$$2(\mu^*/\bar{x}) = (1 - k/N) + \sqrt{(1 - k/N)^2 + (4k/N^2)(Sx_\alpha^2/\bar{x}^2)}$$

(and this is true whether $\bar{x}$ is positive or negative). Now we can write the critical region (1) as

$$\frac{(\mu^*/\bar{x})^{N-k}(\mu^*/\bar{x} - 1)^{N+k}}{(Sx_\alpha^2/\bar{x}^2)^N} \leqq c_1$$

or

$$[1 - k/N + \sqrt{(1 - k/N)^2 + (4k/N)(1 + 1/T^2)}]^{N-k}[1 + 1/T^2]^{-N}$$
$$\cdot[-1 - k/N + \sqrt{(1 - k/N)^2 + (4k/N)(1 + 1/T^2)}]^{N+k} \leqq c_2$$

where c_2 is another positive constant. We denote the left side of this inequality by $\Phi(T^2)$, and it can be shown that $\Phi(T^2)$ is a monotone decreasing function of T^2.

Thus since the critical region is defined by the relation $\Phi(T^2) \leqq$ constant and the critical region using "Student's t" is $T^2 \geqq$ constant, these procedures are exactly equivalent.

4. A Problem in statistical decisions. The question which aroused the interest of the writer in statistical decisions is the following one of multivariate statistical analysis. Suppose $x_1, x_2, \cdots, x_p$ are known to be normally distributed with unknown means and unknown variances and covariances, and on the basis of a set of N independent observations, a test is to be made of the hypothesis $E(x_1) = E(x_2) = \cdots = E(x_p) = 0$. Such a test may be carried out by using the generalized Student Ratio [4], and the hypothesis is either to be rejected or accepted as a whole. But consider the case in which the null hypothesis is rejected; it seems quite natural to ask for a more enlightening statement. Is it not possible to say that on the basis of the sample, the hypothesis should be rejected for $x_{i_1}, x_{i_2}, \cdots, x_{i_k}$ but not rejected for $x_{i_{k+1}}, x_{i_{k+2}}, \cdots, x_{i_p}$? Thus we seek a division of the sample space into 2^p mutually exclusive regions, each of which will lead us to reject the hypothesis of zero expected values for a particular set of the x_i's and to accept it for the remaining set.

We will consider a solution of the problem in the case that the covariance matrix of the joint normal distribution is known, and will motivate that solution by considering first the case of two variables.

Suppose that X and Y are normally and independently distributed with unknown means, α and β, and with unit variances. The joint probability density function is then of the form

$$f(X, Y) = (1/2\pi)\cdot e^{-\frac{1}{2}[(X-\alpha)^2+(Y-\beta)^2]}.$$

The set of hypotheses is given as follows:

H_1 is the hypothesis that $\alpha = 0$ and $\beta = 0$
H_2 is the hypothesis that $\alpha \neq 0$ and $\beta = 0$
H_3 is the hypothesis that $\alpha = 0$ and $\beta \neq 0$
H_4 is the hypothesis that $\alpha \neq 0$ and $\beta \neq 0$.

We have a sample of N independent pairs of observations (X_σ, Y_σ) where $\sigma = 1, 2, \cdots, N$; then the density function in the $2N$-dimensional sample space is

$$(1/2\pi)^N\cdot e^{-\frac{1}{2}S[(X_\sigma-\alpha)^2+(Y_\sigma-\beta)^2]}.$$

We seek the set of regions M_1, M_2, M_3, M_4 in the sample space which are chosen such that if the sample point E falls in M_i, we accept the hypothesis H_i. We take the following as the values of the losses if the wrong decision is reached:

I. If H_1 is accepted,

i) for any parameter point (α, β), the loss is a continuous function of $(\alpha^2 + \beta^2)$, say $W(\alpha^2 + \beta^2)$, which is zero for $\alpha = \beta = 0$, is differentiable, strictly monotonically increasing, and possesses a finite maximum when multiplied by the normal density function.

II. If H_2 is accepted,
  i) for any parameter point (α, β) except $(0, 0)$, the loss is $W(\beta^2)$ where W is the same function as above,
  ii) the loss is W_1 if the true parameter point is $(0, 0)$.

III. If H_3 is accepted,
  i) for any parameter point (α, β) except $(0, 0)$, the loss is $W(\alpha^2)$ where W is the same function as above,
  ii) the loss is W_1 if the true parameter point is $(0, 0)$.

IV. If H_4 is accepted,
  i) the loss is W_2 if the true parameter point is either $(\alpha, 0)$ for $\alpha \neq 0$, or $(0, \beta)$ for $\beta \neq 0$
  ii) the loss is W_3 if the true parameter point is $(0, 0)$

where W_1, W_2, and W_3 are constants subject to some slight restrictions which will be pointed out later.

The functions $R_i(\theta, E)$ are then the following:

$$
\begin{aligned}
R_1(\theta, E) &= W(\alpha^2 + \beta^2)G(\alpha, \beta) && \text{for } \alpha^2 + \beta^2 \neq 0 \\
&= 0 && \text{for } \alpha = \beta = 0 \\
R_2(\theta, E) &= W(\beta^2)G(\alpha, \beta) && \text{for } \beta \neq 0 \\
&= W_1G(0, 0) && \text{for } \alpha = \beta = 0 \\
&= 0 && \text{for } \alpha \neq 0, \beta = 0 \\
R_3(\theta, E) &= W(\alpha^2)G(\alpha, \beta) && \text{for } \alpha \neq 0 \\
&= W_1G(0, 0) && \text{for } \alpha = \beta = 0 \\
&= 0 && \text{for } \alpha = 0, \beta \neq 0 \\
R_4(\theta, E) &= W_2G(\alpha, 0) && \text{for } \alpha \neq 0, \beta = 0 \\
&= W_2G(0, \beta) && \text{for } \alpha = 0, \beta \neq 0 \\
&= W_3G(0, 0) && \text{for } \alpha = \beta = 0 \\
&= 0 && \text{for } \alpha\beta \neq 0
\end{aligned}
$$

where $G(\alpha, \beta)$ is the normal distribution function

$$C \cdot e^{-\frac{1}{2}N[(x-\alpha)^2 + (y-\beta)^2]}$$

x and y being the sample means. It should be pointed out that the use of the distribution of the sample means instead of the joint distribution of the observations is justified since the sample means are sufficient statistics for the parameters α and β.

We will use the notation $\bar{R}_2(E)$ to denote the maximum of $R_2(\theta, E)$ with respect to α and β, and it can easily be seen to be the maximum of two expressions which we will denote by II(1) and II(2) where II(1) is the maximum of $W(\beta^2)G(\alpha, \beta)$ and II(2) is the maximum of $W_1G(0, 0)$. Similarly, $\bar{R}_3(E)$ is the maximum of III(1) and III(2), and $\bar{R}_4(E)$ is the maximum of IV(1), IV(2), and IV(3), where these are the maxima of the two expressions involved in $R_3(\theta, E)$, and the three expressions in $R_4(\theta, E)$, respectively.

We will first show that the function $\bar{R}_1(E)$ is a monotonic increasing function of $(x^2 + y^2)$. We know that the maximum of $R_1(\theta, E)$ is reached for values of

α and β for which the partial derivatives of $R_1(\theta, E)$ with respect to α and β are zero, i.e., for which

$$[N(x - \alpha)W(\alpha^2 + \beta^2) + 2\alpha W'(\alpha^2 + \beta^2)]G(\alpha, \beta) = 0$$

and

$$[N(y - \beta)W(\alpha^2 + \beta^2) + 2\beta W'(\alpha^2 + \beta^2)]G(\alpha, \beta) = 0$$

where $W'(\alpha^2 + \beta^2)$ is the derivative of $W(\alpha^2 + \beta^2)$ with respect to $(\alpha^2 + \beta^2)$. Since $G(\alpha, \beta) \neq 0$, and $W'(\alpha^2 + \beta^2) \neq 0$, these relations imply

$$\begin{vmatrix} x - \alpha & \alpha \\ y - \beta & \beta \end{vmatrix} = 0$$

or $\beta x = \alpha y$. Thus the maximum of the function $R_1(\theta, E)$ occurs for values of α and β which satisfy the relation $\alpha = (x/y)\beta$.

Consider any two straight lines $\alpha = (x'/y')\beta$ and $\alpha = (x''/y'')\beta$, and the values of the function $R_1(\theta, E)$ along these two lines. Obviously the values of the first factor $W(\alpha^2 + \beta^2)$ are equal for points along the lines equidistant from the origin. Also, if the values of x', y', x'', and y'' are such that $x'^2 + y'^2 = x''^2 + y''^2$, the values of the function $G(\alpha, \beta)$ along both lines are equal for points equidistant from the origin, and it follows that $\bar{R}_1(x', y') = \bar{R}_1(x'', y'')$. Thus we have that $\bar{R}_1(E)$ is a function of $(x^2 + y^2)$.

Note that if the value of $x''^2 + y''^2$ is greater than the value of $x'^2 + y'^2$, the curve representing the function $G(\alpha, \beta)$ along $\alpha = (x''/y'')\beta$ is the same as that along the line $\alpha = (x'/y')\beta$, but it is shifted further from the origin. The values of $W(\alpha^2 + \beta^2)$ are independent of x and y and the function is monotonic in $\alpha^2 + \beta^2$. Thus, the value of $G(\alpha, \beta)$ for which $R_1(\theta, E)$ is a maximum on $\alpha = (x'/y')\beta$ multiplies a larger value of $W(\alpha^2 + \beta^2)$ than on $\alpha = (x''/y'')\beta$, so the maximum when $x''^2 + y''^2$ exceeds x'^2 is the greater. But this proves that $\bar{R}_1(E)$ is monotonically increasing in $(x^2 + y^2)$.

In a similar manner, we now proceed to show that II(1) is a monotonically increasing function of y^2. We know that a necessary condition for a maximum of II(1) is that

$$\frac{\partial \text{II}(1)}{\partial \alpha} = \frac{\partial \text{II}(1)}{\partial \beta} = 0.$$

The first of these two relations is

$$W(\beta^2)N(x - \alpha)G(\alpha, \beta) = 0$$

which has the solutions $W(\beta^2) = 0$ and $\alpha = x$. But $W(\beta^2) = 0$ only for $\beta = 0$ and this value is a minimum of II(1), hence we have that the maximum is reached for $\alpha = x$, so

$$\text{II}(1) = \max_{\beta}. \text{ of } W(\beta^2)Ce^{-\frac{1}{2}N(y-\beta)^2}.$$

But along any two lines $\alpha =$ constant in the (α, β)-plane, the function $W(\beta^2)$ has identical monotonically increasing values in β^2 and the normal density function is identical along two such lines for a fixed value of y^2. An increase in the value of y^2 displaces the normal function from the origin but does not affect its shape, hence the value of the normal density function at which II(1) takes on its maximum is multiplied by a greater value of $W(\beta^2)$ when y^2 is increased, so II(1) is monotonically increasing in y^2. In exactly the same manner, we find that III(1) is a monotonically increasing function of x^2.

Because the remaining functions are identical with the functions considered in the special case above, we have that

$$\begin{aligned} \text{II}(2) &= W_1 C e^{-\frac{1}{2}N(x^2+y^2)} \\ \text{III}(2) &= W_1 C e^{-\frac{1}{2}N(x^2+y^2)} \\ \text{IV}(1) &= W_2 C e^{-\frac{1}{2}Ny^2} \\ \text{IV}(2) &= W_2 C e^{-\frac{1}{2}Nx^2} \\ \text{IV}(3) &= W_3 C e^{-\frac{1}{2}N(x^2+y^2)}. \end{aligned}$$

Now it is apparent that $\bar{R}_1(E)$ is never less than II(1) since

$$W(\alpha^2 + \beta^2)G(\alpha, \beta) \geqq W(\beta^2)G(\alpha, \beta)$$

(the equality holds only for $\alpha = 0$) and since a function which is never less than a second function cannot have a maximum less than the maximum of the second function. Also $\bar{R}_1(E)$ for the same reason is never less than III(1). Thus $\bar{R}_1(E)$ can be the minimum of the four functions $\bar{R}_i(E)$ at most when $\bar{R}_2(E)$ is defined by II(2) and $\bar{R}_3(E)$ is defined by III(2).

Since II(2) and III(2) are the same monotonic decreasing function of $(x^2 + y^2)$ and since $\bar{R}_1(E)$ is a monotonic increasing function of $(x^2 + y^2)$, there is a value r_0^2 of $(x^2 + y^2)$ such that $\bar{R}_1(E) <$ II(2) when and only when $x^2 + y^2 < r_0^2$. But for all values (x, y) we have that $\bar{R}_1(E) \geqq$ II(1) and $\bar{R}_1(E) \geqq$ III(1), hence for all values within the circle $x^2 + y^2 = r_0^2$ we have that

$$\text{II}(1) \leqq \bar{R}_1(E) < \text{II}(2) \tag{2}$$

and

$$\text{III}(1) \leqq \bar{R}_1(E) < \text{III}(2) \tag{3}$$

so it follows that $\bar{R}_2(E)$ is defined by II(2) and $\bar{R}_3(E)$ is defined by III(2) within the circle.

We restrict the values of W_1, W_2, and W_3 used in the definitions of the weight functions to be $W_1 \leqq W_2 \leqq W_3$, hence for all values of (x, y)

$$W_1 C e^{-\frac{1}{2}N(x^2+y^2)} \leqq W_2 C e^{-\frac{1}{2}Ny^2}$$

$$W_1 C e^{-\frac{1}{2}N(x^2+y^2)} \leqq W_2 C e^{-\frac{1}{2}Nx^2}$$

and

$$W_1 C e^{-\frac{1}{2}N(x^2+y^2)} \leqq W_3 C e^{-\frac{1}{2}N(x^2+y^2)}$$

so $\bar{R}_4(E)$ is at least as great as II(2) over the whole plane; hence, in light of relation (2), $\bar{R}_4(E)$ is at least as great as $\bar{R}_1(E)$ for $x^2 + y^2 \leqq r_0^2$. Therefore, since (2) shows that $\bar{R}_1(E) < \bar{R}_2(E)$ within the circle; (3) shows that $\bar{R}_1(E) <$

$\bar{R}_3(E)$ within the circle; and since quite obviously the relations do not hold outside the circle, we have that M_1 is the set of points

$$x^2 + y^2 < r_0^2 .$$

To determine the region M_2, we must determine those points outside M_1 for which $\bar{R}_2(E) < \bar{R}_3(E)$ and $\bar{R}_2(E) < \bar{R}_4(E)$. Consider first the part of the plane outside M_1 for which $\bar{R}_2(E)$ is defined by II(2). This is the region for which II(2) > II(1). Consider the curve in the plane defined by II(2) = II(1), that is,

$$W_1Ce^{-\frac{1}{2}N(x^2+y^2)} = \mathrm{II}(1).$$

We take differentials and have

$$-N(x^2 + y^2)W_1Ce^{-\frac{1}{2}N(x^2+y^2)}[xdx + ydy] = 2y[d\mathrm{II}(1)/d(y^2)]dy$$

but this shows that dy/dx has the opposite sign from y/x since $d\mathrm{II}(1)/d(y^2)$ is always positive. Also note that for $x = 0$, the equation $\bar{R}_1(E) = \mathrm{II}(2)$ is identical with the equation II(1) = II(2), so for $x = 0$, we have II(1) > II(2) when $|y| > r_0$ and II(1) < II(2) when $|y| < r_0$. Furthermore, the curve II(1) = II(2) crosses the x axis at a finite value of x, since for $y = 0$, II(1) is a constant while II(2) is a decreasing function of x.

We will refer to the various regions in the first quadrant of the (x, y)-plane shown in Figure I as follows: A is the part of the quadrant which is M_1; A, B, B', and C are the regions in which $\bar{R}_2(E)$ is defined by II(2), that is, in which II(2) > II(1); and in the same manner, A, B, B', and C' are the regions in which $\bar{R}_3(E)$ is defined by III(2).

Since II(2) and III(2) are identical, we see that within the regions B and B', $\bar{R}_2(E) = \bar{R}_3(E)$ since in these regions $\bar{R}_2(E)$ is defined by II(2) and $\bar{R}_3(E)$ is defined by III(2). We have previously pointed out that II(2) is never greater than $\bar{R}_4(E)$, hence it is clear that B and B' should belong to either M_2 or M_3, and we will arbitrarily decide that B is part of M_2 and B' part of M_3.

Consider then the region C; here $\bar{R}_2(E)$ is defined by II(2) and $\bar{R}_3(E)$ by III(1), so within C

$$\mathrm{II}(2) = \mathrm{III}(2) < \mathrm{III}(1) = \bar{R}_3(E)$$

and again $\mathrm{II}(2) \leqq \bar{R}_4(E)$, so the region C is part of M_2. By the same argument we have that C' is a part of M_3 since within C'

$$\mathrm{III}(2) = \mathrm{II}(2) < \mathrm{II}(1) = \bar{R}_2(E)$$

and $\mathrm{III}(2) \leqq \bar{R}_4(E)$.

Now consider the remainder of the quadrant outside A, B, B', C, and C'. Here $\bar{R}_2(E)$ is defined by II(1) and $\bar{R}_3(E)$ is defined by III(1). Since II(1) is the same monotone increasing function of y^2 as III(1) is of x^2, we have II(1) > III(1) for $|y| > |x|$ and II(1) < III(1) for $|x| > |y|$. Thus we see that in the region under discussion, $\bar{R}_2(E)$ is a minimum at most in the regions D and E and $\bar{R}_3(E)$ a minimum at most in D' and E'.

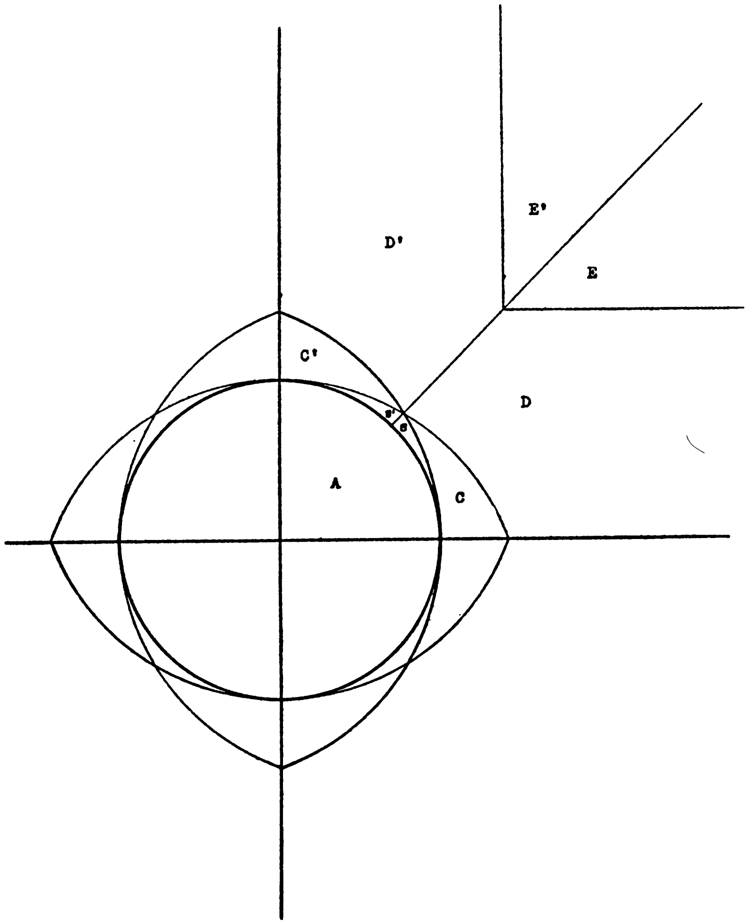

FIG. 1

In order to determine then, that part of D and E which belongs to M_2, we seek the region for which

$$\begin{aligned} &II(1) < IV(1) \text{ when } \bar{R}_4(E) \text{ is defined by } IV(1) \\ &II(1) < IV(2) \text{ when } \bar{R}_4(E) \text{ is defined by } IV(2) \\ &II(1) < IV(3) \text{ when } \bar{R}_4(E) \text{ is defined by } IV(3). \end{aligned}$$

But within D and E we have that $y^2 < x^2$, so it follows that $IV(1) > IV(2)$ so $\bar{R}_4(E)$ is never defined by IV(2) in D or E. Hence we need determine the points which satisfy the first and third of these relations. Now it is clear that the relation $II(1) < IV(1)$ is equivalent to the relation $|y| < y_0$ for some value y_0 since II(1) is monotonically increasing in y^2 and IV(1) is monotonically decreasing in y^2. Let $y = y_0$ be the line dividing D and E.

We impose a restriction on W_3 such that D is part of M_2 and E is part of M_4.

This restriction is that within E, IV(3) $\leqq$ IV(1); note that since we are concerned only with $|y| < |x|$, this imposes the greatest restriction on W_3 when $x = y = y_0$, so we are requiring that

$$W_3 Ce^{-\frac{1}{2}N(y_0^2+y_0^2)} \leqq W_2 Ce^{-\frac{1}{2}Ny_0^2}$$

or

$$W_3 \leqq W_2 e^{+\frac{1}{2}Ny_0^2}.$$

It is simple to see that because of symmetry with respect to both axes and the origin, M_2 is defined by $x^2 + y^2 > r_0^2$ and $|y| < |x|$ and $|y| < y_0$; M_3 by $x^2 + y^2 > r_0^2$ and $|x| < |y|$ and $|x| < x_0$; and M_4 by $x^2 + y^2 > r_0^2$ and $|y| > y_0$ and $|x| > x_0$. It should be pointed out that $x_0 = y_0$.

We now consider the general case with a known covariance matrix. Consider the joint normally distributed variates $X_1^*, X_2^*, \cdots, X_p^*$ whose covariance matrix is $\| \sigma_{ij}^* \|$ $(i, j = 1, 2, \cdots, p)$, where the σ_{ij}^*'s are all known and where $\| \sigma_{ij}^* \|$ is positive definite. The mean values of the X_i^*'s are $\beta_1, \beta_2, \cdots, \beta_p$ which are unknown. It is simple to see that we can consider new variates $X_i = X_i^*/\sqrt{\sigma_{ii}^*}$ whose mean values are $\alpha_i = \beta_i/\sqrt{\sigma_{ii}^*}$ and whose covariance matrix is $\| \sigma_{ij} \|$ where $\sigma_{ii} = 1$. If a sample of N independent observations on the X_i^*'s are given, we have immediately the observations on the X_i's, and we denote the sample means of the X_i's by $x_1, x_2, \cdots, x_p$, respectively.

There are 2^p hypotheses among which we wish to choose; as notation, we let

H_0 be $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$
H_1 be $\alpha_1 \neq 0, \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$
H_2 be $\alpha_2 \neq 0, \alpha_1 = \alpha_3 = \cdots = \alpha_p = 0$
H_{12} be $\alpha_1\alpha_2 \neq 0, \alpha_3 = \alpha_4 = \cdots = \alpha_p = 0$

etc. As a further abbreviation, let H^1 denote any one of the p hypotheses $H_1, H_2, \cdots, H_p$; let H^2 denote any of the $\binom{p}{2}$ hypotheses $H_{12}, H_{13}, \cdots$; H^3 denote any of the $\binom{p}{3}$ hypotheses $H_{123}, H_{124}, \cdots$; etc. Also let $M_{i_1 i_2 \cdots i_k}$ be the region of the sample space for which we accept the hypothesis $H_{i_1 i_2 \cdots i_k}$, and let $R_{i_1 i_2 \cdots i_k}(\theta, E) = W(\theta, H_{i_1 i_2 \cdots i_k})f(E \mid \theta)$ be the risk density function if the hypothesis $H_{i_1 i_1 \cdots i_k}$ is chosen, where we have used the notation θ to represent the parameter point $\alpha_1, \alpha_2, \cdots, \alpha_p$.

We will also adopt the following notations: in referring to the parameter point $(\alpha_1, \alpha_2, \cdots, \alpha_p)$, we will write $(i_1, i_2, \cdots, i_k) = 0$ to mean all points for which $\alpha_{i_1} = \alpha_{i_2} = \cdots = \alpha_{i_k} = 0$ and $(\alpha_{j_1})(\alpha_{j_2}) \cdots (\alpha_{j_s}) \neq 0$ where $i_1, i_2, \cdots i_k, j_1, j_2, \cdots, j_s$ are a permutation of the integers $1, 2, \cdots, p$. Furthermore, we will write $[j_1, j_2, \cdots, j_s] \neq 0$ to mean $(i_1, i_2, \cdots, i_k) = 0$.

By Q we denote the covariance matrix of the X_i's and by L its inverse; we will denote the elements of L by λ_{ij}. By $Q^{i_1 i_2 \cdots i_k}$ we denote the matrix obtained by striking out rows $i_1, i_2, \cdots, i_k$ and columns $i_1, i_2, \cdots, i_k$ from Q; by $L^{i_1 i_2 \cdots i_k}$ we denote the inverse of the matrix $Q^{i_1 i_2 \cdots i_k}$, and we will write the elements of

$L^{i_1 i_2 \cdots i_k}$ as $\lambda_{ij}^{i_1 i_2 \cdots i_k}$. Thus we can write the joint distribution of the set of sample means $x_1, x_2, \cdots, x_p$ as

$$Ce^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}(x_i-\alpha_i)(x_j-\alpha_j)}. \tag{4}$$

Concerning the definition of the weight function, we will assume the following:

I. If H_0 is accepted,
   i) the loss is $W(\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j)$ if the true parameter point is $(\alpha_1, \alpha_2, \cdots, \alpha_p)$, where W is a continuous, strictly monotonic increasing function whose value is zero if $(1, 2, \cdots, p) = 0$. The function is restricted to increase slowly enough that the product of it and the density function (4) has a finite maximum with respect to the α_i's

II. If H^1 is accepted,
   i) consider in particular H_a, then for all parameter points except $(1, 2, \cdots, p) = 0$, the value of the loss is $W(\Sigma\Sigma\lambda_{ij}^a\alpha_i\alpha_j)$, where W is the function defined above.
   ii) the loss is W_0^1 if the true parameter point is $(1, 2, \cdots, p) = 0$.

III. If H^2 is accepted,
   i) consider in particular H_{ab}, then for all parameter points except $(1, 2, \cdots, p) = 0$ and $[a] \neq 0$ and $[b] \neq 0$, the loss is $W(\Sigma\Sigma\lambda_{ij}^{ab}\alpha_i\alpha_j)$, where W is the function defined above,
   ii) the loss is W_1^2 if the true parameter point is either $[a] \neq 0$ or $[b] \neq 0$, where $W_0^1 \leqq W_1^2$,
   iii) the loss is W_0^2 if the true parameter point is $(1, 2, \cdots, p) = 0$ where $W_0^2 \geqq W_1^2$.

In general; if H^k is accepted,
   i) consider in particular $H_{i_1 i_2 \cdots i_k}$, then for all parameter points except $(1, 2, \cdots, p) = 0$, $[i_1] \neq 0$, $[i_2] \neq 0, \cdots, [i_1, i_2] \neq 0$, $[i_1, i_3] \neq 0, \cdots$, etc., the loss is $W(\Sigma\Sigma\lambda_{ij}^{i_1 i_2 \cdots i_k}\alpha_i\alpha_j)$,
   ii) the loss is W_r^k $(r = 1, 2, \cdots, k - 1)$ if $[i_{j_1}, i_{j_2}, \cdots, i_{j_r}] \neq 0$, where $j_1, j_2, \cdots, j_r$ are r different positive integers less than or equal to k. Also $W_{k-1}^k \leqq W_{k-2}^k \leqq \cdots \leqq W_1^k$, $W_{k-2}^{k-1} \leqq W_{k-2}^k$, $W_{k-3}^{k-2} \leqq W_{k-3}^{k-1} \leqq W_{k-3}^k$, etc.
   iii) the loss is W_0^k if $(1, 2, \cdots, p) = 0$, where $W_1^k \leqq W_0^k$,

where the W_s^t are constants subject to some further slight restrictions which we will impose later. The $\Sigma\Sigma$ has been used throughout to denote summation over all values which i and j take on in $L^{i_1 i_2 \cdots i_k}$.

We consider first the risk density function corresponding to H_0, that is

$$R_0(\theta, E) = W(\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j)Ce^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}(x_i-\alpha_i)(x_j-\alpha_j)}.$$

To maximize $R_0(\theta, E)$, we have the set of p equations obtained by setting the p partials of $R_0(\theta, E)$ with respect to the α_i equal to zero, which are necessary conditions. We have

$$\frac{\partial R_0(\theta, E)}{\partial \alpha_i} = \left\{\frac{\partial W}{\partial \alpha_i} + [N\Sigma\lambda_{ij}(x_j - \alpha_j)]W\right\} Ce^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}(x_i-\alpha_i)(x_j-\alpha_j)}$$

so the necessary conditions are

$$\frac{\partial W}{\partial \alpha_i} + [N\Sigma\lambda_{ij}(x_j - \alpha_j)]W = 0 \qquad (i = 1, 2, \cdots, p).$$

This can also be written

$$(2\Sigma\lambda_{ij}\alpha_j)D_z W(z) + W(z)N\Sigma\lambda_{ij}(x_j - \alpha_j) = 0$$

where we have set $z = \Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j$ and where we use the notation D_z to indicate differentiation with respect to z. Fix i at two particular values, say a and b; then two of the equations of this system can be written

$$(2\Sigma\lambda_{aj}\alpha_j)D_z W(z) + W(z)N\Sigma\lambda_{aj}(x_j - \alpha_j) = 0$$

$$(2\Sigma\lambda_{bj}\alpha_j)D_z W(z) + W(z)N\Sigma\lambda_{bj}(x_j - \alpha_j) = 0$$

that is

$$(\Sigma\lambda_{aj}\alpha_j)[\Sigma\lambda_{bj}(x_j - \alpha_j)] = (\Sigma\lambda_{bj}\alpha_j)[\Sigma\lambda_{aj}(x_j - \alpha_j)]$$

or

$$(\Sigma\lambda_{aj}\alpha_j)(\Sigma\lambda_{bj}x_j) = (\Sigma\lambda_{bj}\alpha_j)(\Sigma\lambda_{aj}x_j).$$

This we can write as

$$\Sigma\Sigma\lambda_{aj}\lambda_{bk}\alpha_j x_k = \Sigma\Sigma\lambda_{bk}\lambda_{aj}\alpha_k x_j$$

or

$$\Sigma\Sigma\lambda_{aj}\lambda_{bk}(\alpha_j x_k - \alpha_k x_j) = 0.$$

Giving a and b the p^2 combinations of values which are possible, this is a set of p^2 linear homogeneous equations in the p^2 unknowns $(\alpha_j x_k - \alpha_k x_j)$ which has the obvious solution $\alpha_j x_k - \alpha_k x_j = 0$ or $\alpha_j x_k = \alpha_k x_j$.

Thus we have that the maximum of the function $R_0(\theta, E)$ is reached for a set of values of the α_i's which lie on the straight line

$$\alpha_i = (x_i/x_1)\alpha_1 . \tag{5}$$

The function $\bar{R}_0(E)$, which is the maximum of $R_0(\theta, E)$ with respect to the α_i's is a monotonically increasing function of $(\Sigma\Sigma\lambda_{ij}x_i x_j)$, which we show in the following manner. Because of (5), we see that

$$\begin{aligned}\Sigma\Sigma\lambda_{ij}(x_i - \alpha_i)(x_j - \alpha_j) &= \Sigma\Sigma\lambda_{ij}[x_i - (x_i/x_1)\alpha_1][x_j - (x_j/x_1)\alpha_1] \\ &= \Sigma\Sigma\lambda_{ij}x_i x_j[1 - (\alpha_1/x_1)]^2.\end{aligned}$$

Also,

$$\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j = \Sigma\Sigma\lambda_{ij}x_i x_j(\alpha_1/x_1)^2.$$

Hence we see that $\bar{R}_0(E)$ is the maximum with respect to ω of

$$W(\omega^2\Sigma\Sigma\lambda_{ij}x_i x_j)Ce^{-\frac{1}{2}N(1-\omega)^2\Sigma\Sigma\lambda_{ij}x_i x_j}$$

so for two sample points $E' = (x_1', x_2', \cdots, x_p')$ and $E'' = (x_1'', x_2'', \cdots, x_p'')$ such that $\Sigma\Sigma\lambda_{ij}x_i'x_j' = \Sigma\Sigma\lambda_{ij}x_i''x_j''$, it is clear that $\bar{R}_0(E') = \bar{R}_0(E'')$; thus $\bar{R}_0(E)$ is a function of $\Sigma\Sigma\lambda_{ij}x_ix_j$.

But then without loss of generality, we can consider $\bar{R}_0(E)$ along the x_1 axis, i.e. for $x_2 = x_3 = \cdots = x_p = 0$. Using relation (5), we see that this implies that the maximizing parameter values are $\alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$. But then

$$\bar{R}_0(E) = \underset{\alpha_1}{\text{max. of}}\ W(\lambda_{11}\alpha_1^2)Ce^{-\frac{1}{2}N\lambda_{11}(x_1-\alpha_1)^2}$$

which we have previously shown is a monotonic increasing function of x_1^2. Therefore $\bar{R}_0(E)$ is a monotonic increasing function of $\Sigma\Sigma\lambda_{ij}x_ix_j$.

We will furthermore show that the maximum of each risk density function corresponding to parts i) as given in the weight functions are monotonically increasing functions of certain quadratic forms in the x_i. Consider for example the function corresponding to part i) of $R_1(\theta, E)$, that is

$$W(\Sigma\Sigma\lambda_{ij}^1\alpha_i\alpha_j)Ce^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}(x_i-\alpha_i)(x_j-\alpha_j)}. \tag{6}$$

We will write the maximum of this function with respect to the α_i's as $\bar{R}_1(i)$. Note that the weight function is not a function of α_1, hence the partial derivative of (6) with respect to α_1 set equal to zero is equivalent to

$$\Sigma\lambda_{1j}(x_j - \alpha_j) = 0.$$

Squaring this relation and multiplying by $N/2\lambda_{11}$ gives

$$(N/2\lambda_{11})\Sigma\Sigma\lambda_{1i}\lambda_{1j}(x_i - \alpha_i)(x_j - \alpha_j) = 0$$

so we can write the exponent in (6)

$$\text{Exp.} = -(N/2\lambda_{11})\Sigma\Sigma(\lambda_{11}\lambda_{ij} - \lambda_{1i}\lambda_{1j})(x_i - \alpha_i)(x_j - \alpha_j).$$

Because of the definition of λ_{ij}, if we write ω_{ij} for the cofactor of σ_{ij} in $|\sigma_{ij}|$, we have

$$\text{Exp.} = -[N/2\lambda_{11}(|\sigma_{ij}|)^2]\Sigma\Sigma(\omega_{11}\omega_{ij} - \omega_{1i}\omega_{1j})(x_i - \alpha_i)(x_j - \alpha_j).$$

But by a well known algebraic identity[2],

$$\begin{aligned}\omega_{11}\omega_{ij} - \omega_{1i}\omega_{1j} &= |\sigma_{ij}| \cdot [\text{cofactor of } (\sigma_{11}\sigma_{ij} - \sigma_{1i}\sigma_{1j}) \text{ in } |\sigma_{ij}|] \\ &= |\sigma_{ij}| \cdot \omega_{ij}^1\end{aligned}$$

where we have written ω_{ij}^1 to be the cofactor of σ_{ij} in $|\sigma_{ij}^1|$, so

$$\text{Exp.} = -(N/2\lambda_{11}|\sigma_{ij}|)\Sigma\Sigma\omega_{ij}^1(x_i - \alpha_i)(x_j - \alpha_j).$$

But $\lambda_{11}|\sigma_{ij}| = \omega_{11} = |\sigma_{ij}^1|$, hence

$$\text{Exp.} = -\frac{N}{2}\Sigma\Sigma\lambda_{ij}^1(x_i - \alpha_i)(x_j - \alpha_j).$$

Therefore

$$\bar{R}_1(i) = \underset{\text{all } \alpha_i\text{'s}}{\text{max. of}}\ W(\Sigma\Sigma\lambda_{ij}^1\alpha_i\alpha_j)Ce^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}^1(x_i-\alpha_i)(x_j-\alpha_j)}.$$

---

[2] See M. Bocher, *Introduction to Higher Algebra.*

But then it follows in exactly the same way as with $\bar{R}_0(E)$ that $\bar{R}_1(i)$ is a monotonically increasing function of $\Sigma\Sigma\lambda^1_{ij}x_ix_j$. For the other functions $\bar{R}_k(i)$ corresponding to other hypotheses H^1, the argument is identical, and for risk density functions corresponding to hypotheses with more than one $\alpha_i \neq 0$, the same argument is repeated two or more times in succession to give the result.

We will show that for any value of the parameters $\alpha_1, \alpha_2, \cdots, \alpha_p$ the relation

$$\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j \geqq \Sigma\Sigma\lambda^1_{ij}\alpha_i\alpha_j$$

holds. This relation is true if the relation

$$\Sigma\Sigma[(\omega_{ij}/\mid\sigma_{ij}\mid) - (\omega^1_{ij}/\mid\sigma^1_{ij}\mid)]\alpha_i\alpha_j \geqq 0 \tag{7}$$

is true where we define $\omega^1_{1j} = 0$. That is, if

$$(1/\mid\sigma_{ij}\mid\mid\sigma^1_{ij}\mid)\Sigma\Sigma\omega_{ij}\omega_{11} - \omega^1_{ij}\mid\sigma_{ij}\mid)\alpha_i\alpha_j \geqq 0$$

where we have substituted ω_{11} for its equal $\mid\sigma^1_{ij}\mid$. But note that

$$\omega^1_{ij} = \text{cofactor of } (\sigma_{11}\sigma_{ij} - \sigma_{1i}\sigma_{1j}) \text{ in } \mid\sigma_{ij}\mid$$

hence by the identity quoted (see footnote 2)

$$\mid\sigma_{ij}\mid\omega^1_{ij} = \omega_{11}\omega_{ij} - \omega_{1i}\omega_{1j}$$

so the left hand member of relation (7) is

$$\begin{aligned}(1/\mid\sigma_{ij}\mid\mid\sigma^1_{ij}\mid)&\Sigma\Sigma(\omega_{ij}\omega_{11} - \omega_{11}\omega_{ij} + \omega_{1i}\omega_{1j})\alpha_i\alpha_j\\ &= (1/\mid\sigma_{ij}\mid\mid\sigma^1_{ij}\mid)\Sigma\Sigma\omega_{1i}\omega_{1j}\alpha_i\alpha_j\\ &= [\Sigma\omega_{1i}\alpha_i]^2/(\mid\sigma_{ij}\mid\mid\sigma^1_{ij}\mid)\\ &\geqq 0\end{aligned}$$

since all matrices here are symmetric and positive definite. Note that the argument can be repeated one or more times to show

$$W(\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j) \geqq W(\Sigma\Sigma\lambda^{i_1i_2\cdots i_k}_{ij}\alpha_i\alpha_j)$$

or

$$W(\Sigma\Sigma\lambda^{j_1j_2\cdots j_s}_{ij}\alpha_i\alpha_j) \geqq W(\Sigma\Sigma\lambda^{i_1i_2\cdots i_k}_{ij}\alpha_i\alpha_j)$$

where $i_1i_2, \cdots, i_k$ are any set of k different integers less than or equal to p, and $j_1j_2\cdots, j_s$ are any subset of $i_1i_2\cdots, i_k$.

Consider the maximum of the expressions

$$W^k_rCe^{-\frac{1}{2}N\Sigma\Sigma\lambda'_{ij}(x_i-\alpha_i)(x_j-\alpha_j)}.$$

We know that $(p - r)$ of the α_i's in these expressions are zero and by an argument similar to that given above[3], it is clear that if the r α_i's not equal to zero are $\alpha_{i_1}, \alpha_{i_2}, \cdots, \alpha_{i_r}$, then the maximum of the expressions is given by

$$W^k_rCe^{-\frac{1}{2}N\Sigma\Sigma\lambda^{i_1i_2\cdots i_r}_{ij}x_ix_j}.$$

[3] See p. 36.

Also for $r = 0$, the maximum is obviously

$$W_0^k C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j}.$$

Recall that we have restricted the W_r^k's so that

$$W_0^1 \leqq W_0^2 \leqq \cdots \leqq W_0^p \quad \text{and} \quad W_{p-1}^p \leqq \cdots \leqq W_0^p. \tag{8}$$

From a previous calculation, it follows that

$$\Sigma\Sigma\lambda_{ij}x_i x_j \geqq \Sigma\Sigma\lambda_{ij}^{i_1}x_i x \geqq \Sigma\Sigma\lambda_{ij}^{i_1 i_2}x_i x_j \geqq \cdots. \tag{9}$$

We can then quite easily calculate the region M_0, that is, the region of the sample space for which $\bar{R}_0(E)$ is the minimum of all the $\bar{R}_{i_1 i_2 \cdots i_k}(E)$'s. We have pointed out that

$$W(\Sigma\Sigma\lambda_{ij}\alpha_i\alpha_j) \geqq W(\Sigma\Sigma\lambda_{ij}^{i_1 i_2 \cdots i_k}\alpha_i\alpha_j)$$

so it follows that

$$\bar{R}_0(E) \geqq \bar{R}_{i_1 i_2 \cdots i_k}(i) \tag{10}$$

that is

$$\bar{R}_0(E) \geqq \bar{R}_{i_1 i_2 \cdots i_k}(E)$$

so long as $\bar{R}_{i_1 i_2 \cdots i_k}(E)$ is defined by $\bar{R}_{i_1 i_2 \cdots i_k}(i)$.

From the relations (8) and (9), we have that

$$W_0^1 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j} \leqq W_0^k C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j} \tag{11}$$

for $k = 2, 3, \cdots, p$. Now because

$$W_0^1 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j}$$

is a monotonic decreasing function of $\Sigma\Sigma\lambda_{ij}x_i x_j$, and because $\bar{R}_0(E)$ is a monotonically increasing function of $\Sigma\Sigma\lambda_{ij}x_i x_j$, there is a value r_0^2 such that within the ellipse $\Sigma\Sigma\lambda_{ij}x_i x_j = r_0^2$, the relation

$$\bar{R}_0(E) < W_0^1 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j} \tag{12}$$

holds, and outside it the opposite inequality holds. But from relations (10) and (12), it follows that within this ellipse, no $\bar{R}_{i_1 i_2 \cdots i_k}(E)$ except $\bar{R}_0(E)$ can be defined by $\bar{R}_{i_1 i_2 \cdots i_k}(i)$. Then in view of relation (11) and since a quantity is certainly less than the maximum of several quantities if it is less than one of those several quantities, the region M_0 is the set of points $\Sigma\Sigma\lambda_{ij}x_i x_j < r_0^2$.

Now consider the functions $\bar{R}_a(E)$ in the region outside M_0. We know that $\bar{R}_a(E) = \bar{R}_a(i)$ when

$$\max_{\alpha_i\text{'s}} \text{ of } W(\Sigma\Sigma\lambda_{ij}^a\alpha_i\alpha_j)e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}^a(x_i-\alpha_i)(x_j-\alpha_j)} \geqq W_0^1 e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}x_i x_j}$$

and we will write $\bar{R}_i(E) = \bar{R}_i(ii)$ when the opposite inequality holds. Consider a part of the sample space outside M_0 in which

$$\begin{aligned} \bar{R}_{i_1}(E) &= \bar{R}_{i_1}(ii) \\ \bar{R}_{i_2}(E) &= \bar{R}_{i_2}(ii) \\ &\cdots\cdots\cdots\cdots \\ \bar{R}_{i_k}(E) &= \bar{R}_{i_k}(ii) \end{aligned}$$

where $k \geqq 1$, and where $\bar{R}_j(E) \neq \bar{R}_j(ii)$ for $j \neq i_1, i_2, \cdots, i_k$. We see in this case that $\bar{R}_{i_1}(E) = \bar{R}_{i_2}(E) = \cdots = \bar{R}_{i_k}(E) < \bar{R}_j(E)$, where again $j \neq i_1, i_2, \cdots, i_k$. Furthermore, in this case, because of the relation (11), we have that E should be a point of either $M_{i_1}, M_{i_2}, \cdots$ or, M_{i_k}. We will arbitrarily decide in this case that E should be a point of M_{i_s} (s an integer $\leqq k$) where i_s is determined so that

$$\Sigma\Sigma\lambda_{ij}^{i_s} x_i x_j \leqq \Sigma\Sigma\lambda_{ij}^{i_t} x_i x_j \quad \text{for any } t = 1, 2, \cdots, k.$$

Now consider the region in which $\bar{R}_r(E) = \bar{R}_r(i)$ for all $r = 1, 2, \cdots, p$. We see that each $\bar{R}_r(i)$ is the same monotonically increasing function of a quadratic form of the type $\Sigma\Sigma\lambda_{ij}^r x_i x_j$. Hence in order that E be a point of a particular M_r, it is necessary that

$$\Sigma\Sigma\lambda_{ij}^r x_i x_j \leqq \Sigma\Sigma\lambda_{ij}^s x_i x_j \qquad \text{for all } s \neq r. \tag{13}$$

Now let us consider a fixed r and compare $\bar{R}_r(i)$ with all $R_{ri_1i_2\cdots i_k}(E)$'s for $k \geqq 1$. We have pointed out that

$$\Sigma\Sigma\lambda_{ij}^r x_i x_j \geqq \Sigma\Sigma\lambda_{ij}^{ri_1i_2\cdots i_k} x_i x_j \tag{14}$$

so $\bar{R}_r(i) \geqq \bar{R}_{ri_1i_2\cdots i_k}(i)$ and hence $\bar{R}_r(i)$ can be a minimum at most when all $\bar{R}_{ri_1i_2\cdots i_k}(E)$'s are defined by other than $\bar{R}_{ri_1i_2\cdots i_k}(i)$.

Consider then, any $\bar{R}_{ri_1}(E)$ when defined by other than $\bar{R}_{ri_1}(i)$, that is when $\bar{R}_{ri_1}(E)$ is equal to one of

$$\begin{aligned} W_1^2 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}^r x_i x_j} &= \bar{R}_{ri_1}(ii) \quad \text{(say)} \\ W_1^2 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij}^{i_1} x_i x_j} &= \bar{R}_{ri_1}(iii) \quad \text{(say)} \\ W_0^2 C e^{-\frac{1}{2}N\Sigma\Sigma\lambda_{ij} x_i x_j} &= \bar{R}_{ri_1}(iv) \quad \text{(say)}. \end{aligned}$$

Because of the relations (8) and (14), we have that

$$\bar{R}_{ri_1}(E) \leqq \bar{R}_{ri_1i_2\cdots i_k}(E)$$

whenever these are defined by other than $\bar{R}_{ri_1}(i)$ and $\bar{R}_{ri_1i_2\cdots i_k}(i)$. Furthermore in the region defined by (13), we see that $\bar{R}_{ri_1}(ii) \geqq \bar{R}_{ri_1}(iii)$, hence $\bar{R}_{ri_1}(E)$ is never defined by $\bar{R}_{ri_1}(iii)$ in this region.

Now the relation $\bar{R}_r(i) < \bar{R}_{ri_1}(ii)$ is easily seen to be equivalent to the relation

$$\Sigma\Sigma\lambda_{ij}^r x_i x_j < r_1^2 \tag{15}$$

for some value r_1. With the restriction on W_0^2 that it be not so much larger than W_1^2 that when (12) does not hold, $\bar{R}_{ri_1}(E)$ is not defined by $\bar{R}_{ri_1}(iv)$, we have that the region for which $\bar{R}_r(i) < \bar{R}_{ri_1 i_2 \cdots i_k}(E)$ is the region defined by (13) and (15).

We then restrict the relationship between the constants W_0^1 and W_0^2 to be such that for all points outside of M_0 but within the region defined by (13) and (15), the relation $\Sigma\Sigma\lambda_{ij}^{j_1 j_2 \cdots j_k} x_i x_j \geqq \Sigma\Sigma\lambda_{ij}^{r} x_i x_j$ holds for $j_1, j_2, \cdots, j_k$ each different from r. Note that this is not an unreasonable restriction since the right hand side of the relation is bounded above by r_1^2, $\Sigma\Sigma\lambda_{ij} x_i x_j$ is bounded below by r_0^2, and therefore, $\Sigma\Sigma\lambda_{ij}^{j_1 j_2 \cdots j_k} x_i x_j$ is bounded below by some positive value r^2 where r^2 is a monotonically increasing function of r_0^2.

Using a similar method, the region $M_{i_1 i_2 \cdots i_k}$ can be obtained after all regions $M_{i_1 i_2 \cdots i_m}$ for all $m < k$ have been derived. If some further restrictions are imposed on the constants in the weight functions similar to those formulated in deriving the region M_r, it can be shown that the region $M_{i_1 i_2 \cdots i_k} (k \geqq 1)$ will be given by the inequalities

$$\Sigma\Sigma\lambda_{ij} x_i x_j \geqq r_0^2$$

$$\Sigma\Sigma\lambda_{ij}^{j_1 j_2 \cdots j_m} x_i x_j \geqq r_m^2 \qquad \text{for all } m < k \text{ and all } j_1, \cdots, j_m$$

$$\Sigma\Sigma\lambda_{ij}^{i_1 i_2 \cdots i_k} x_i x_j \leqq \Sigma\Sigma\lambda_{ij}^{j_1 j_2 \cdots j_k} x_i x_j \qquad \text{for all } j_1, \cdots, j_k$$

and

$$\Sigma\Sigma\lambda_{ij}^{i_1 i_2 \cdots i_k} x_i x_j < r_k^2.$$

Thus we have rationalized the following solution of the question posed at the beginning of section 4. We test the hypothesis $E(x_1) = E(x_2) = \cdots = E(x_p) = 0$ using the generalized Student ratio replacing the sample covariance matrix by the population covariance matrix since the latter is assumed to be known, at some chosen level of significance. If the hypothesis is not rejected, we make the decision corresponding to H_0. If the ratio is significant, we compute the ratios $T^1, T^2, \cdots, T^p$ where by definition $T^{i_1 i_2 \cdots i_k}$ is the generalized Student ratio computed for $x_{j_1}, x_{j_2}, \cdots, x_{j_s}$ ($i_1, i_2, \cdots, i_k, j_1, j_2, \cdots, j_s$ is a permutation of the integers $1, 2, \cdots, p$), the variates $x_{i_1}, x_{i_2}, \cdots, x_{i_k}$ being ignored.

We consider the smallest of the ratios computed on the basis of $(p - 1)$ of the x_i's; say it is T^r. Then if T^r is not significant at some level of significance (which need not be the same level as considered before), we make the decision corresponding to H_r; if T^r is significant, we compute all the ratios based on $(p - 2)$ of the x's. If T^{rs} is the smallest of these, we make the decision corresponding to H_{rs} if T^{rs} is not significant but proceed to calculate the ratios based on $(p - 3)$ of the x_i's if it is significant, and so on.

5. Concluding remarks. It should be pointed out that while the derivation of the explicit inequalities defining the various regions of acceptance may be rather involved, for any given sample point E, it is relatively simple to determine the region of acceptance to which this point E belongs. That is, we calculate the various values $\bar{R}_{i_1 i_2 \cdots i_k}(E)$ and choose the decision $H_{j_1 \cdots j_s}$ if $\bar{R}_{j_1 j_2 \cdots j_s}(E)$ is the minimum of the values of $\bar{R}_{i_1 i_2 \cdots i_k}(E)$ for all values of $i_1, i_2, \cdots, i_k$. For making a decision on the basis of a given sample point E, it is not necessary to find explicit analytic formulas defining the shapes of the various regions of acceptance.

Since the principle used here is proposed merely as a substitute for Wald's principle for the sake of mathematical simplification, it is felt that in certain problems Wald's principle may be used as a check on the results. For example, it is felt that the new principle is apt to lead to decision regions of the proper shape though the exact sizes of these regions may not be correct. In cases where the decision regions cannot be determined by Wald's principle, it seems possible that a determination may be made in Wald's sense among the various decision regions having the same shapes as those given by the new principle. In the case considered here, for example, it may be possible to determine new values of $r_0^2, r_1^2, \cdots, r_{p-1}^2$.

I should like to express my very great appreciation to Professor H. Hotelling for many suggestions during the preparation of this paper and to Professor A. Wald for constant guidance. I should also like to credit Professor Helen Walker with originally posing the question that led to this research.

REFERENCES

[1] A. Wald, *Annals of Math. Stat.*, Vol. 10 (1939), pp. 299–326.
[2] A. Wald, *On the Principles of Statistical Inference*, Notre Dame, Ind., 1942.
[3] J. Neyman and E. Pearson, *Transactions of the Royal Society*, A., Vol. 231 (1933), p. 295.
[4] H. Hotelling, *Annals of Math. Stat.*, Vol. 2 (1931), pp. 360–378.