AN APPLICATION OF INFORMATION THEORY TO MULTIVARIATE ANALYSIS, II

By S. Kullback

The George Washington University

0. Summary. Certain results of information theory are applied to some problems of multivariate analysis, including the multivariate linear hypothesis and the hypothesis of homogeneity of covariance matrices. A discussion of certain related linear discriminant functions is also included. Some asymptotic distributions on the null hypothesis are derived. Related problems, still under investigation, are mentioned. The procedures are based on the principle of maximizing information. For the cases considered, the estimates of $I(1:2)$ and $J(1,2)$ turn out to be those obtained by replacing the parameters by unbiased estimates, appropriate to the hypotheses under consideration.

1. Introduction. In a previous paper [20], the author considered certain results of information theory as applied to multivariate normal populations. In particular there was examined the problem of finding the "best" linear function for discriminating between two normal populations, assuming equal means but different population covariance matrices. The multivariate analysis techniques of discriminant analysis, principal components, and canonical correlations were seen to be closely related concepts. (Greenhouse [12], using the information-theory approach, has examined the problem of finding the "best" linear function for discriminating between two multivariate normal populations, with no restrictive assumptions as to means or covariance matrices.)

In [20], the discussion was in terms of population parameters, and questions of estimation and distribution were omitted. In addition to discussing some of the problems of estimation and distribution herein, we also want to consider further application of information theory, and the relation with previous developments, by studying the following four multivariate problems (cf. Roy [28], Section 5.1):

(a) The hypothesis that a k-variate normal population has the covariance matrix σ;

(b) The hypothesis of equality of r means for each of k variates for r k-variate normal populations with different covariance matrices, and with the same covariance matrix;

(c) The multivariate linear hypothesis, including the case of a subhypothesis;

(d) The hypothesis of equality of the covariance matrices of r k-variate normal populations.

The reader is referred to Section 2 of [20] in which the information measures are defined and their properties summarized, with particular reference to prop-

Received March 26, 1954; revised September 8, 1955.

erty (iii) on p. 90 of [20]. (Proofs may be found in [22].) Based on the non-decreasing property of $I(1:2)$ and $J(1,2)$ for sufficient statistics, we follow a principle that may be termed maximizing information in order to attain sufficiency or near sufficiency. It seems intuitively reasonable that such an approach should have certain optimum properties. In certain cases the results are closely related to likelihood ratio tests and this relation is under investigation for the general case. Asymptotic distributions occurring herein verify conclusions derivable from a general asymptotic theory, the details of which are in preparation (cf. Wilks [32]), and in two cases a better approximation than the general theory provides is derived.

It might be remarked that appropriate multivariate extensions of the results in Cramér [6], particularly pp. 11 and 12, and Daniels [7] also provide an alternative basis for a general asymptotic theory.

There are few general (automatic) procedures for finding test criteria. The approach using information theory as a means of determining test procedures may be of interest and use because it is, so to speak, an automatic procedure.

Matrix notation, methods, and results are used and assumed known to the reader. The notation, and such results of [20] as are needed, will be used without further summary herein.

2. Components of information. Since $I(1:2)$ and $J(1,2)$ are additive for independent random variables, for a random sample of n observations $I_n(1:2) = nI(1:2)$ and $J_n(1,2) = nJ(1,2)$, where $I(1:2)$ and $J(1,2)$ are given, respectively, by (2.8) and (2.7) of [20].

As is well known, the averages and the variances and covariances in samples from a multivariate normal population are independently distributed, respectively, in a normal distribution and in the Wishart distribution (see, for example, Wilks [33]). Computing the appropriate values from the respective distributions, it is readily found (see, for example, Hoyt [14]), that for the averages

$$
\begin{aligned}
I'(1:2;\bar{x}) &= \tfrac{1}{2}\log\frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} - \frac{k}{2} + \tfrac{1}{2}\operatorname{tr}\sigma_{(1)}\sigma_{(2)}^{-1} + \frac{n}{2}\delta'\sigma_{(2)}^{-1}\delta \\
&= \frac{n}{2}\delta'\sigma^{-1}\delta \qquad \text{for } \sigma_{(1)} = \sigma_{(2)} = \sigma,
\end{aligned} \tag{2.1}
$$

$$
\begin{aligned}
J'(1,2;\bar{x}) &= \tfrac{1}{2}\operatorname{tr}[(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})] + \frac{n}{2}\delta'(\sigma_{(1)}^{-1} + \sigma_{(2)}^{-1})\delta \\
&= n\delta'\sigma^{-1}\delta \qquad \text{for } \sigma_{(1)} = \sigma_{(2)} = \sigma,
\end{aligned} \tag{2.2}
$$

where $\delta = \mu_{(1)} - \mu_{(2)}$, and for the sample unbiased variances and covariances,

$$
I'(1:2;S) = \frac{n-1}{2}\left(\log\frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} - k + \operatorname{tr}\sigma_{(1)}\sigma_{(2)}^{-1}\right), \tag{2.3}
$$

$$
J'(1,2;S) = \frac{n-1}{2}\operatorname{tr}[(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})]. \tag{2.4}
$$

We thus have from the preceding,

$$nI(1:2) = I'(1:2; \bar{x}) + I'(1:2; S), \tag{2.5}$$

$$nJ(1,2) = J'(1,2; \bar{x}) + J'(1,2; S). \tag{2.6}$$

3. Estimates of information. The procedure we shall use (replacing population parameters in $I(1:2)$, $J(1,2)$ by unbiased estimates appropriate to the hypotheses) is based on a principle of maximizing information, as may be seen by the following heuristic discussion.

Suppose that $g_2(y)$ and $g^*(y)$ are densities, satisfying the conditions of Section 4 of [21], such that for given $g_2(y)$ we require

$$I^* = \int g^*(y) \log \frac{g^*(y)}{g_2(y)} \, d\gamma(y) \tag{3.1}$$

to be a maximum, subject to

$$\int g^*(y) \, d\gamma(y) = 1, \qquad \int y g^*(y) \, d\gamma(y) = a. \tag{3.2}$$

This is equivalent to maximizing

$$U = \int \left(g^*(y) \log \frac{g^*(y)}{g_2(y)} + k g^*(y) + l y g^*(y) \right) d\gamma(y), \tag{3.3}$$

where k and l are arbitrary constants to be determined. The usual variational procedures lead to

$$\delta U = 0 = \int \delta g^*(y) \left[\log \frac{g^*(y)}{g_2(y)} + 1 + k + ly \right] d\gamma(y) \tag{3.4}$$

or

$$\log \frac{g^*(y)}{g_2(y)} + 1 + k + ly = 0. \tag{3.5}$$

This means that

$$g^*(y) = e^{-1-k-ly} g_2(y) \tag{3.6}$$

or, by integration, that

$$1 = e^{-1-k} \int e^{-ly} g_2(y) \, d\gamma(y) = e^{-1-k} M_2(t), \tag{3.7}$$

where we have replaced $-l$ by t.

Thus,

$$g^*(y) = \frac{e^{ty} g_2(y)}{M_2(t)}, \qquad M_2(t) = \int e^{ty} g_2(y) \, d\gamma(y); \tag{3.8}$$

since this means that

$$I^* = \int g^*(y) \log \frac{g^*(y)}{g_2(y)} \, d\gamma(y) = at - \log M_2(t), \tag{3.9}$$

the desired maximum occurs for that value of t which maximizes $at - \log M_2(t)$, or when $I^* = -\log m_2(a)$, in the notation of Section 4 of [21]. (It might be noted that $k \log m_2(a) = -kI^*$, where k is Boltzmann's constant, is the entropy of the distribution whose density is $g_2(y)$ (cf. Khinchin [18]).) Also,

$$J^* = \int (g^*(y) - g_2(y)) \log \frac{g^*(y)}{g_2(y)} \, d\gamma(y) = t(a)(a - E_2(y)). \tag{3.10}$$

For a simple hypothesis, the parameters of $g_2(y)$ are completely specified. For a composite hypothesis, say $\theta \in \Theta$, where θ is a vector of the parameters and Θ is some subset of the parameter space, we use as the appropriate test of the null hypothesis (that relative to $g_2(y)$) the value given by

$$\hat{I} = \min_{\theta \in \Theta} I^* = I^*(\hat{\theta}) \tag{3.11}$$

and, correspondingly,

$$\hat{J} = J^*(\hat{\theta}). \tag{3.12}$$

We shall carry through the foregoing in detail for some cases; in others, we shall apply the procedure of replacement of the parameters by unbiased estimates.

4. Single sample. Consider problem (a) of Section 1. For a sample of n observations from a multivariate normal population with mean matrix $\mu' = (\mu_1, \mu_2, \cdots, \mu_k)$ and covariance matrix σ, the moment-generating function of the sample averages $\bar{x}' = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ and V_{ii} and $2V_{ij}$, $i \neq j$, the elements of the matrix $V = NS$, where N is the number of degrees of freedom and S is the sample unbiased covariance matrix, is known to be given by ([33], p. 121)

$$M_2(t, T) = |I - 2\sigma T|^{-N/2} \exp\left(t'\mu + \tfrac{1}{2}t' \frac{\sigma}{n} t\right), \tag{4.1}$$

where $t' = (t_1, t_2, \cdots, t_k)$, $T = (t_{ij})$, $i, j = 1, 2, \cdots, k$.

We want to determine g^* of (3.9) (the conjugate distribution of Khinchin [18]) so as to have the observed unbiased estimates as its parameters, that is to say, $a = (\bar{x}, V)$, which means that we seek the values of t and T which will maximize (cf. [18], Section 33)

$$I^* = t'\bar{x} + \operatorname{tr} TV - t'\mu - \tfrac{1}{2}t' \frac{\sigma}{n} t + \frac{N}{2} \log |I - 2\sigma T|. \tag{4.2}$$

Differentiating with respect to t and T (see [8], p. 364), we have

$$\bar{x} - \mu - \frac{\sigma t}{n} = 0, \qquad \operatorname{tr} (dT)V - N \operatorname{tr} (I - 2\sigma T)^{-1}\sigma(dT) = 0, \tag{4.3}$$

from which are derived

$$(4.4)\qquad t = n\sigma^{-1}(\bar{x} - \mu)\,, \qquad T = \tfrac{1}{2}\sigma^{-1} - \tfrac{1}{2}S^{-1}.$$

Using the values given by (4.4), (4.2) becomes

$$(4.5)\qquad I^* = \frac{n}{2}(\bar{x} - \mu)'\sigma^{-1}(\bar{x} - \mu) + \frac{N}{2}(\log |\sigma| / |S| - k + \operatorname{tr} S\sigma^{-1}).$$

The null hypothesis specifies σ but not μ. It is clear that for variations of μ, I^* in (4.5) is a minimum for $\hat{\mu} = \bar{x}$, or

$$(4.6)\qquad \hat{I} = \min_{\mu} I^* = I^*(\hat{\mu}) = \frac{N}{2}(\log |\sigma| / |S| - k + \operatorname{tr} S\sigma^{-1}).$$

Note that (4.6) is (2.3) with $\sigma_{(1)} = S, \quad \sigma_{(2)} = \sigma$.

The case of a single sample of n observations from a k-variate normal population was considered in some detail by Hoyt [14]. Hoyt showed that asymptotically $2\hat{I}$ has a chi-square distribution with $k(k + 1)/2$ d.f., and to a closer approximation, R. A. Fisher's B distribution.

In considering tests of significance in factor analysis, Bartlett [4], using a "homogeneous" likelihood function, and Rippe [26], using the likelihood-ratio test procedure for the problem of tests of significance of components in matrix factorization, arrived at the statistic $2\hat{I}$ and the same conclusion as to its asymptotic chi-square distribution.

For the hypothesis of independence of variates, i.e.,

$$(4.7)\qquad \sigma = (\sigma_{ij}), \qquad \sigma_{ij} = 0, \qquad i \neq j,$$

we may write (4.6) as

$$(4.8)\qquad 2\hat{I} = -N \log |R| + N\left[\sum_{i=1}^{k}\left(\frac{S_{ii}}{\sigma_{ii}} + \log \frac{\sigma_{ii}}{S_{ii}} - 1\right)\right],$$

where R is the matrix of sample correlation coefficients, or

$$(4.9)\qquad 2\hat{I} = 2\hat{I}'_R + 2\hat{I}'_S\,,$$

where $2\hat{I}'_R = -N \log |R|$ (cf. [20], p. 94). Wilks [31] has shown that when (4.7) holds, the s_{ii} and r_{ij} are independent, so that $\hat{I}'_R$ and $\hat{I}'_S$ are independent. It follows from the discussion of Section 9 that, asymptotically, $2\hat{I}'_R$ has a chi-square distribution with $k(k - 1)/2$ d.f., and $2\hat{I}'_S$ has a chi-square distribution with k d.f. It is shown in Section 9 that a better approximation to the distribution of $2\hat{I}'_R$ is given by Fisher's B distribution ([10], p. 14.665) with

$$\beta^2 = k(k - 1)(2k + 5)/12N, \quad B^2 = 2\hat{I}'_R\,, \quad n_1 = k(k - 1)/2.$$

5. Homogeneity of means. Consider problem (b) of Section 1. We will first discuss the case for two samples. Suppose we have two independent samples, having, respectively, n_1, n_2 independent observations from k-variate normal populations with respective covariance matrices Σ_1, Σ_2. We want to test the null hypothesis that the population mean vectors are equal, i.e.,

$$H_2 : \mu_{(1)} = \mu_{(2)} = \mu, \qquad \Sigma_1, \Sigma_2, \tag{5.1}$$

with no specification about Σ_1 and Σ_2.

Using the notation already introduced in Section 4, we want to determine g^*, with $a = (\bar{x}_{(1)}, \bar{x}_{(2)}, V_1, V_2)$, which means that we seek the values of $t_{(i)}$, T_i, $i = 1, 2$, which will maximize

$$\begin{aligned} I^* = {} & t'_{(1)}\bar{x}_{(1)} - t'_{(1)}\mu - \tfrac{1}{2}t'_{(1)}\frac{\Sigma_1}{n_1}t_{(1)} + \operatorname{tr} T_1V_1 + \frac{N_1}{2}\log|I - 2\Sigma_1T_1| \\ & + t'_{(2)}\bar{x}_{(2)} - t'_{(2)}\mu - \tfrac{1}{2}t'_{(2)}\frac{\Sigma_2}{n_2}t_{(2)} + \operatorname{tr} T_2V_2 + \frac{N_2}{2}\log|I - 2\Sigma_2T_2|. \end{aligned} \tag{5.2}$$

Following the procedure as used for (4.3), we find that the sought for values are given by

$$\begin{aligned} t_{(1)} &= n_1\Sigma_1^{-1}(\bar{x}_{(1)} - \mu), \qquad & t_2 &= n_2\Sigma_2^{-1}(\bar{x}_{(2)} - \mu), \\ T_1 &= \tfrac{1}{2}\Sigma_1^{-1} - \tfrac{1}{2}S_1^{-1}, \qquad & T_2 &= \tfrac{1}{2}\Sigma_2^{-1} - \tfrac{1}{2}S_2^{-1}, \end{aligned} \tag{5.3}$$

for which values I^* of (5.2) becomes

$$\begin{aligned} I^* = {} & \frac{n_1}{2}(\bar{x}_{(1)} - \mu)'\Sigma_1^{-1}(\bar{x}_{(1)} - \mu) + \frac{n_2}{2}(\bar{x}_{(2)} - \mu)'\Sigma_2^{-1}(\bar{x}_{(2)} - \mu) \\ & + \frac{N_1}{2}\left(\log\frac{|\Sigma_1|}{|S_1|} - k + \operatorname{tr} S_1\Sigma_1^{-1}\right) + \frac{N_2}{2}\left(\log\frac{|\Sigma_2|}{|S_2|} - k + \operatorname{tr} S_2\Sigma_2^{-1}\right). \end{aligned} \tag{5.4}$$

The null hypothesis requires equality of the means with no specification on the covariance matrices. It is clear that for variations of Σ_1 and Σ_2, I^* will be a minimum for $\hat{\Sigma}_1 = S_1$, $\hat{\Sigma}_2 = S_2$, and for $\hat{\mu}$ satisfying

$$0 = n_1S_1^{-1}(\bar{x}_{(1)} - \hat{\mu}) + n_2S_2^{-1}(\bar{x}_{(2)} - \hat{\mu}) \tag{5.5}$$

or

$$\hat{\mu} = (n_1S_1^{-1} + n_2S_2^{-1})^{-1}(n_1S_1^{-1}\bar{x}_{(1)} + n_2S_2^{-1}\bar{x}_{(2)}). \tag{5.6}$$

For convenience let $d = \bar{x}_{(1)} - \bar{x}_{(2)}$, $A = n_1S_1^{-1}$, $B = n_2S_2^{-1}$; substituting in (5.4) we get

$$\begin{aligned} & 2\hat{I}(\hat{\mu}, \hat{\Sigma}_1, \hat{\Sigma}_2) \\ & \quad = \operatorname{tr}\{[B(A + B)^{-1}A(A + B)^{-1}B + A(A + B)^{-1}B(A + B)^{-1}A]\,dd'\}. \end{aligned} \tag{5.7}$$

But $B(A + B)^{-1}A = [A^{-1}(A + B)B^{-1}]^{-1} = (B^{-1} + A^{-1})^{-1}$ and $A(A + B)^{-1}B = [B^{-1}(A + B)A^{-1}]^{-1} = (B^{-1} + A^{-1})^{-1}$, so that finally

$$
\begin{aligned}
2\hat{I} &= \operatorname{tr}[(B^{-1} + A^{-1})^{-1}\, dd'] \\
&= d'(B^{-1} + A^{-1})^{-1}\, d \qquad (5.8)\\
&= (\bar{x}_{(1)} - \bar{x}_{(2)})' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}).
\end{aligned}
$$

It is readily found that in this case $\hat{J} = 2\hat{I}$.

For the single variate case, see Fisher [10], pp. 35.174–35.180.

Linear discriminant function. Consider $y = \alpha' x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$, the same linear compound for each sample. Since y is normally distributed, we seek α so as to maximize

$$
(5.9) \qquad 2\hat{I}'(\hat{\mu}, \hat{\Sigma}_1, \hat{\Sigma}_2; y) = \frac{\alpha'\, dd'\, \alpha}{\alpha' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right) \alpha}.
$$

As is easily determined (cf. [20], p. 91), the maximum occurs for

$$
\alpha = \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} d \qquad \text{and} \qquad 2\hat{I}'(y) = 2\hat{I}.
$$

r-samples. Suppose we have r independent samples, having, respectively, n_i, $i = 1, 2, \cdots, r$, independent observations each, from k-variate normal populations with respective covariance matrices Σ_i, $i = 1, 2, \cdots, r$, and we want to test the null hypothesis that the population mean vectors are equal, i.e.,

$$
(5.10) \qquad H_2 : \mu_{(1)} = \mu_{(2)} = \cdots = \mu_{(r)} = \mu, \qquad \Sigma_1, \Sigma_2, \cdots, \Sigma_r,
$$

with no specification about the Σ_i.

Without repeating the details, we find, in this case, that

$$
(5.11) \quad I^* = \sum_{i=1}^{r} \frac{n_i}{2} (\bar{x}_{(i)} - \mu)' \Sigma_i^{-1} (\bar{x}_{(i)} - \mu) + \sum_{i=1}^{r} \frac{N_i}{2} \left(\log \frac{|\Sigma_i|}{|S_i|} - k + \operatorname{tr} S_i \Sigma_i^{-1}\right),
$$

$$
(5.12) \qquad \hat{\Sigma}_i = S_i, \qquad \hat{\mu} = \left(\sum_{i=1}^{r} n_i S_i^{-1}\right)^{-1} \left(\sum_{i=1}^{r} n_i S_i^{-1} \bar{x}_{(i)}\right) = \bar{x}.
$$

$$
(5.13) \qquad 2\hat{I} = \sum_{i=1}^{r} n_i (\bar{x}_{(i)} - \bar{x})' S_i^{-1} (\bar{x}_{(i)} - \bar{x}).
$$

On the null hypothesis, $2\hat{I}$ has an asymptotic chi-square distribution with $(r - 1)k$ d.f.

Covariance matrices equal. If we assume that the population covariance matrices are equal, i.e., that $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_r = \Sigma$, and want to test the null hypothesis that the population mean vectors are equal, then, without repeating the details, we find, in this case, that

$$
(5.14) \quad I^* = \sum_{i=1}^{r} \frac{n_i}{2} (\bar{x}_{(i)} - \mu)' \Sigma^{-1} (\bar{x}_{(i)} - \mu) + \frac{N}{2} \left(\log \frac{|\Sigma|}{|S|} - k + \operatorname{tr} S \Sigma^{-1}\right),
$$

where $NS = N_1S_1 + \cdots + N_rS_r$, $N = N_1 + N_2 + \cdots + N_r$, and that

$$\hat{\Sigma} = S, \quad n\hat{\mu} = n\bar{x} = n_1\bar{x}_1 + \cdots + n_r\bar{x}_r, \quad n = n_1 + n_2 + \cdots + n_r, \tag{5.15}$$

$$\begin{aligned} 2\hat{I} &= \sum_{i=1}^{r} n_i(\bar{x}_{(i)} - \bar{x})'S^{-1}(\bar{x}_{(i)} - \bar{x}) \\ &= \text{tr } S^{-1}(n_1\, d_{(1)}\, d'_{(1)} + \cdots + n_r\, d_{(r)}\, d'_{(r)}) \\ &= \text{tr } S^{-1}\, S^*, \end{aligned} \tag{5.16}$$

where $d_{(i)} = \bar{x}_{(i)} - \bar{x}$, $S^* = \sum_{i=1}^{r} n_i\, d_{(i)}\, d'_{(i)}$ (cf. Hotelling [13]). It is readily found that in this case $\hat{J} = 2\hat{I}$.

Asymptotically, $2\hat{I} = \hat{J}$ has a chi-square distribution with $k(r - 1)$ d.f. on the null hypothesis (cf. [25], p. 372). This will be shown in Section 10.

Linear discriminant function. Consider $y = \alpha'x = \alpha_1x_1 + \alpha_2x_2 + \cdots + \alpha_kx_k$, the same linear compound for each sample. Since y is normally distributed, we have for the y's, using (5.16),

$$\begin{aligned} 2\hat{I}'(y) &= \frac{n_1(\alpha'\, d_{(1)})^2 + \cdots + n_r(\alpha'\, d_{(r)})^2}{\alpha'S\alpha} \\ &= \frac{\alpha'(n_1\, d_{(1)}\, d'_{(1)} + \cdots + n_r\, d_{(r)}\, d'_{(r)})\alpha}{\alpha'S\alpha} \\ &= \frac{\alpha'S^*\alpha}{\alpha'S\alpha}, \end{aligned} \tag{5.17}$$

with the symbols as defined in (5.16). For the linear compound for which $2\hat{I}'(y)$ is a maximum, the usual calculus procedures yield the result that the α's must satisfy $S^*\alpha = lS\alpha$, where l is the largest root of the equation $|S^* - lS| = 0$, which has (almost everywhere) p positive and $(k - p)$ zero roots, where $p \leqq \min(k, r - 1)$ (cf. [28]). If we denote the positive roots in descending order as $l_1, l_2, \cdots, l_p$,

$$\begin{aligned} 2\hat{I} = \hat{J} = \text{tr } S^{-1}S^* &= l_1 + l_2 + \cdots + l_p \\ &= \hat{J}'(l_1) + \hat{J}'(l_2) + \cdots + \hat{J}'(l_p). \end{aligned} \tag{5.18}$$

The discrimination efficiency of the linear compound associated with l_i is given by

$$\text{Eff} = \frac{\hat{J}'(l_i)}{\hat{J}} = \frac{l_i}{l_1 + l_2 + \cdots + l_p}. \tag{5.19}$$

In this case, asymptotically we have, on the null hypothesis, the chi-square decomposition (cf. [25], p. 373)

$$\begin{aligned} \hat{J}'(l_p) &= l_p, \ |k - (r - 1)| + 1 \quad \text{d.f.} \\ \hat{J}'(l_{p-1}) &= l_{p-1}, \ |k - (r - 1)| + 3 \text{ d.f.} \\ \cdots \quad & \quad \cdots \qquad \cdots \\ \hline \hat{J} &= l_1 + l_2 + \cdots + l_p, \ k(r - 1) \quad \text{d.f.} \end{aligned}$$

This is to be taken in the sense that $l_{m+1} + \cdots + l_p$ is asymptotically a chi-square, not that $l_{m+1}, \cdots, l_p$ have asymptotic independent chi-square distributions (see (10.3)).

EXAMPLE. Consider the following data from a problem discussed by Bartlett and which was also considered in [20], p. 93. (See further references therein.) Here, $r = 8$, $k = 2$, $n = n_1 + \cdots + n_8 = 57$,

$$49S = \begin{pmatrix} 136{,}972.6 & 58{,}549.0 \\ 58{,}549.0 & 71{,}496.1 \end{pmatrix}, \qquad S^* = \begin{pmatrix} 12{,}496.8 & -6{,}786.6 \\ -6{,}786.6 & 32{,}985.0 \end{pmatrix},$$

and the roots of $|S^* - lS| = 0$ are given by $l_1 = 44.68667$, $l_2 = 3.09106$. Also,

$$\begin{array}{rl} \hat{J}'(l_2) = 3.09106 & 6 \text{ d.f.} \\ \hat{J}'(l_1) = 44.68667 & 8 \text{ d.f.} \\ \hline \hat{J} = 47.77773 & 14 \text{ d.f.} \end{array}$$

Since only $\hat{J}'(l_1)$ is significant, the linear discriminant function $y = x_2 - 0.535x_1$, associated with l_1, is affected by the treatments and is practically sufficient.

6. Multivariate linear hypothesis. Consider problem (c) of Section 1. Let $Z_{(i)} = Y_{(i)} - BX_{(i)}$, $i = 1, 2, \cdots, n$, where $Z'_{(i)} = (Z_{i1}, \cdots, Z_{ik_2})$, $Y'_{(i)} = (y_{i1}, \cdots, y_{ik_2})$, $X'_{(i)} = (x_{i1}, \cdots, x_{ik_1})$, $B = (\beta_{rs})$, $r = 1, 2, \cdots, k_2$, $s = 1, 2, \cdots, k_1$, $k_1 \geqq k_2$, and the $Z_{(i)}$ are independent k_2-variate normal random vectors with zero means and common covariance matrix Σ. The $Y_{(i)}$ are stochastic and the $X_{(i)}$ are considered known. The usual unbiased estimate of B is given by (see [2], pp. 103–104) $\hat{B} = (Y'X)(X'X)^{-1}$, where

$$Y' = (Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}), \quad X' = X_{(1)}, X_{(2)}, \cdots, X_{(n)}),$$

and that of Σ is given by $(n - k_1)\hat{\Sigma} = \hat{Z}'\hat{Z} = (Y' - \hat{B}X')(Y - X\hat{B}') = Y'Y - \hat{B}X'X\hat{B}' = Y'Y - (Y'X)(X'X)^{-1}(X'Y)$.

Let us now consider the hypotheses

$$(6.1) \qquad \begin{array}{ll} H_1: E_1(Y_{(i)}) = BX_{(i)}, & i = 1, 2, \cdots, n, \\ H_2: E_2(Y_{(i)}) = 0, & \text{i.e., } B = 0. \end{array}$$

As in pp. 90–91 of [20], we have

$$2I(1:2) = J(1, 2) = \delta'\sigma^{-1}\delta$$

$$(6.2) \qquad = (X'_{(1)}B', X'_{(2)}B', \cdots, X'_{(n)}B') \begin{bmatrix} \Sigma^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Sigma^{-1} \end{bmatrix} \begin{bmatrix} BX_{(1)} \\ BX_{(2)} \\ \vdots \\ BX_{(n)} \end{bmatrix}$$

$$= X'_{(1)}B'\Sigma^{-1}BX_{(1)} + \cdots + X'_{(n)}B'\Sigma^{-1}BX_{(n)}$$
$$= \mathrm{tr}\ \Sigma^{-1}B(X_{(1)}X'_{(1)} + \cdots + X_{(n)}X'_{(n)})B'$$
$$= \mathrm{tr}\ \Sigma^{-1}BX'XB'.$$

Using the estimates given above, we get as the estimate of $J(1, 2)$ (cf. [20], p. 96)

$$\begin{aligned} 2\hat{I}(1:2) &= \hat{J}(1, 2) = \mathrm{tr}\ \hat{\Sigma}^{-1}\hat{B}X'X\hat{B}' \\ &= (n - k_1)\ \mathrm{tr}\ (Y'Y - (Y'X)(X'X)^{-1}(X'Y))^{-1}(Y'X)(X'X)^{-1}(X'Y) \qquad (6.3) \\ &= (n - k_1)\ \mathrm{tr}\ S_{22.1}^{-1}S_{21}S_{11}^{-1}S_{12}, \end{aligned}$$

where $X'X = nS_{11}$, $X'Y = nS_{12}$, $Y'X = nS_{21}$, $Y'Y = nS_{22}$, and

$$S_{22.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}.$$

We may also express $\hat{J}(1, 2)$ as $(n - k_1)$ times the sum of the k_2 roots (almost everywhere positive) of the determinantal equation $|S_{21}S_{11}^{-1}S_{12} - lS_{22.1}| = 0$.

As in the preceding section,

$$(6.4) \qquad 2\hat{I}(1:2) = \hat{J}(1, 2) = (n - k_1)(l_1 + l_2 + \cdots + l_{k_2}),$$

asymptotically on the null hypothesis H_2, has a chi-square distribution with k_1k_2 d.f. (see Section 10). By replacing $S_{22.1}$ by its value as given above,

$$|S_{21}S_{11}^{-1}S_{12} - lS_{22.1}| = 0 = |S_{21}S_{11}^{-1}S_{12} - r^2S_{22}|, \text{ where } l = r^2/(1 - r^2).$$

The r's thus defined are Hotelling's canonical correlation coefficients. (See [20], pp. 95–99, and further references therein.) We may also write (6.4) as (cf. [20], p. 97)

$$(6.5) \qquad 2\hat{I}(1:2) = \hat{J}(1, 2) = (n - k_1)\left(\frac{r_1^2}{1 - r_1^2} + \frac{r_2^2}{1 - r_2^2} + \cdots + \frac{r_{k_2}^2}{1 - r_{k_2}^2}\right).$$

On the null hypothesis $B = 0$, the results are equivalent to those for the null hypothesis that in a k-variate normal population, the set of the first k_1 variates is uncorrelated with the set of the last k_2 variates, $k = k_1 + k_2$. The latter hypothesis is the one considered in [20], pp. 95–99.

Linear discriminant function. For the problem of this section, consider $w_i = \alpha'Y_{(i)} = \alpha_1Y_{i1} + \alpha_2Y_{i2} + \cdots + \alpha_{k_2}Y_{ik_2}$, $i = 1, 2, \cdots, n$, the same linear compound of the y's for each observation. Since the w's are normally distributed with $\sigma_{w_i}^2 = \alpha'\Sigma\alpha$, we have for the w's

$$\begin{aligned} 2I'(1:2; w) = J'(1,2; w) &= \frac{(\alpha'BX_{(1)})^2 + \cdots + (\alpha'BX_{(n)})^2}{\alpha'\Sigma\alpha} \\ &= \frac{\alpha'B(X_{(1)}X'_{(1)} + \cdots + X_{(n)}X'_{(n)})B'\alpha}{\alpha'\Sigma\alpha} \qquad (6.6) \\ &= \frac{\alpha'BX'XB'\alpha}{\alpha'\Sigma\alpha}. \end{aligned}$$

To find the linear compound for which $J'(1,2; w)$ is a maximum, the usual calculus procedures yield the result that the α's must satisfy $BX'XB'\alpha = \lambda\Sigma\alpha$, where λ is the largest root of the equation $|BX'XB' - \lambda\Sigma| = 0$. Denoting the k_2 positive roots in descending order as $\lambda_1, \lambda_2, \cdots, \lambda_{k_2}$,

$$(6.7) \quad \begin{aligned} 2I(1:2) = J(1,2) &= \operatorname{tr} \Sigma^{-1}BX'XB' = \lambda_1 + \lambda_2 + \cdots + \lambda_{k_2} \\ &= J'(1,2; \lambda_1) + \cdots + J'(1,2: \lambda_{k_2}). \end{aligned}$$

Using the estimates as in (6.3) and (6.4), we have

$$(6.8) \quad \begin{aligned} 2\hat{I}(1:2) = \hat{J}(1,2) &= (n - k_1)(l_1 + l_2 + \cdots + l_{k_2}) \\ &= \hat{J}'(1,2; l_1) + \hat{J}'(1,2; l_2) + \cdots + \hat{J}'(1,2; l_{k_2}). \end{aligned}$$

The canonical correlations enter as before.

In this case, too, asymptotically on the null hypothesis H_2, we have the chi-square decomposition (see Section 10)

$$\begin{array}{ll} \hat{J}'(1,2; l_{k_2}) = (n - k_1)l_{k_2} = (n - k_1)r_{k_2}^2/(1 - r_{k_2}^2) & k_1 - k_2 + 1 \text{ d.f.} \\ \hat{J}'(1,2; l_{k_2-1}) = (n - k_1)l_{k_2-1} = (n - k_1)r_{k_2-1}^2/(1 - r_{k_2-1}^2) & k_1 - k_2 + 3 \text{ d.f.} \\ \cdots\cdots\cdots\cdots\cdots\cdots & \cdots \\ \hat{J}'(1,2; l_1) = (n - k_1)l_1 = (n - k_1)r_1^2/(1 - r_1^2) & k_1 + k_2 - 1 \text{ d.f.} \\ \hline \hat{J}(1,2) = (n - k_1)\sum_{i=1}^{k_2} l_i = (n - k_1)\sum_{i=1}^{k_2} r_i^2 \Big/ (1 - r_i^2) & k_1k_2 \text{ d.f.} \end{array}$$

This is to be taken in the sense that $(n - k_1)(l_{m+1} + \cdots + l_{k_2})$ is asymptotically a chi-square, not that $(n - k_1)l_{m+1}, \cdots, (n - k_1)l_{k_2}$ have asymptotic independent chi-square distributions. (See (10.4).)

EXAMPLE. By way of illustration, we use the data already discussed by Hotelling and the values derived in [20], p. 98, where it was found that $r_1^2 = .1556$, $r_2^2 = .0047$, $r_1^2/(1 - r_1^2) = .1843$, $r_2^2/(1 - r_2^2) = .0047$; and since $n - k_1 = 139 - 2 = 137$ (there were 140 observations but the values were computed about the sample averages),

$$\begin{array}{rl} \hat{J}'(1,2; r_2) = & .6439 \text{ 1 d.f.} \\ \hat{J}'(1,2; r_1) = & 25.2491 \text{ 3 d.f.} \\ \hline \hat{J}(1,2) = & 25.8930 \text{ 4 d.f.} \end{array}$$

Since only $\hat{J}'(1,2; r_1)$ is significant, the linear discriminant function associated with r_1, $w = -2.4404\, y_1 + y_2$, is the only such linear function and is practically sufficient, confirming the inference made in [20], p. 99.

Subhypothesis. We return to the problem at the beginning of this section and separate the k_1 x's into two sets of q_1 and q_2, $k_1 = q_1 + q_2$. With a corresponding partition of the matrix B, we now have $Z_{(i)} = Y_{(i)} - CX_{(1i)} - DX_{(2i)}$, where $X_{(i)} = \begin{pmatrix} X_{(1i)} \\ X_{(2i)} \end{pmatrix}$; $B = (C, D)$, where C and D are, respectively, $k_2 \times q_1$, $k_2 \times q_2$

matrices; or $Z' = Y' - CX_1' - DX_2'$, with Z and Y as previously defined and

$$X' = (X_{(1)}, X_{(2)}, \cdots, X_{(n)}) = \left(\begin{pmatrix} X_{(11)} \\ X_{(21)} \end{pmatrix}, \cdots, \begin{pmatrix} X_{(1n)} \\ X_{(2n)} \end{pmatrix}\right) = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix}.$$

With the same assumptions as to the $Z_{(i)}$, we now consider the hypotheses

$$\begin{aligned} H_1: E_1(Y_{(i)}) &= C_1 X_{(1i)} + D_1 X_{(2i)}, \\ H_2: E_2(Y_{(i)}) &= C_2 X_{(1i)} + D_2 X_{(2i)}, \end{aligned} \qquad i = 1, 2, \cdots, n. \tag{6.9}$$

Applying the same procedures as previously, it is found that now

$$\begin{aligned} 2I(1:2) &= J(1,2) \\ &= \operatorname{tr} \Sigma^{-1} \left\{ ((C_1 - C_2), (D_1 - D_2)) \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} (C_1 - C_2)' \\ (D_1 - D_2)' \end{pmatrix} \right\}, \end{aligned} \tag{6.10}$$

where

$$X'X = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} (X_1 X_2) = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

In particular, we wish to test the null hypothesis that $D_2 = 0$. For C_1 and D_1, the estimation procedure previously used for B, [2], yields here

$$(\hat{C}_1, \hat{D}_1) \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = (Y'X_1, Y'X_2), \tag{6.11}$$

or

$$\begin{aligned} \hat{C}_1 S_{11} + \hat{D}_1 S_{21} &= Y'X_1, \\ \hat{C}_1 S_{12} + \hat{D}_1 S_{22} &= Y'X_2; \end{aligned} \tag{6.12}$$

and for C_2,

$$\hat{C}_2 S_{11} = Y'X_1. \tag{6.13}$$

From (6.12) it is readily found that

$$\hat{D}_1 = Y'X_{2.1} S_{22.1}^{-1}, \qquad \hat{C}_1 = Y'X_1 S_{11}^{-1} - \hat{D}_1 S_{21} S_{11}^{-1}, \tag{6.14}$$

where $X_{2.1} = X_2 - X_1 S_{11}^{-1} S_{12}$, $S_{22.1} = S_{22} - S_{21} S_{11}^{-1} S_{12}$.

For the estimate of Σ, we have as before

$$(n - k_1)\hat{\Sigma} = Y'Y - (\hat{C}_1, \hat{D}_1) \begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22} \end{pmatrix} \begin{pmatrix} \hat{C}_1' \\ \hat{D}_1' \end{pmatrix}; \tag{6.15}$$

and for the estimate of $J(1,2)$,

$$\hat{J}(1,2) = \operatorname{tr} \hat{\Sigma}^{-1}\left\{((\hat{C}_1 - \hat{C}_2), \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix}(\hat{C}_1 - \hat{C}_2)' \\ \hat{D}_1'\end{pmatrix}\right\}. \tag{6.16}$$

Using the values given in (6.13) and (6.14), it is found that

$$\begin{aligned} (\hat{C}_1, \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix}\hat{C}_1' \\ \hat{D}_1'\end{pmatrix} &= \hat{C}_2 S_{11} \hat{C}_2' + \hat{D}_1 S_{22.1} \hat{D}_1' \\ &= Y'X_1 S_{11}^{-1} X_1' Y + Y'X_{2.1} S_{22.1}^{-1} X_{2.1}' Y, \\ ((\hat{C}_1 - \hat{C}_2), \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix}(\hat{C}_1 - \hat{C}_2)' \\ \hat{D}_1'\end{pmatrix} &= \hat{D}_1 S_{22.1} \hat{D}_1'. \end{aligned} \tag{6.17}$$

It is readily verified that

$$X_1 S_{11}^{-1} X_1' X_{2.1} S_{22.1}^{-1} X_{2.1}' = 0, \tag{6.18}$$

and since $X_{2.1}'X_{2.1} = S_{22.1}$,

$$(I - X_1 S_{11}^{-1} X_1' - X_{2.1} S_{22.1}^{-1} X_{2.1}') X_{2.1} S_{22.1}^{-1} X_{2.1}' = 0; \tag{6.19}$$

that is to say, the two factors in $\hat{J}(1,2)$ are independent.

These results are summarized in Tables 1 and 2.

We omit a discussion of linear discriminant functions for this case.

7. Homogeneity of covariance matrices. Consider problem (d) of Section 1. For its special interest, we consider first the case of two samples and then the general case.

(7.1) *Two samples.* Suppose that we have two independent samples with n_1

TABLE 1

Due to	d.f.	Generalized sum of squares
$\hat{C}_2$	q_1	$\hat{C}_2 S_{11} \hat{C}_2' = Y'X_1 S_{11}^{-1} X_1' Y$
Difference	q_2	$\hat{D}_1 S_{22.1} \hat{D}_1' = Y'X_{2.1} S_{22.1}^{-1} X_{2.1}' Y$
$\hat{C}_1, \hat{D}_1$	k_1	$\hat{B}X'X\hat{B}' = Y'X_1 S_{11}^{-1} X_1' Y + Y'X_{2.1} S_{22.1}^{-1} X_{2.1}' Y$
Difference	$n - k_1$	$Y'Y - \hat{B}X'X\hat{B}' = (n - k_1)\hat{\Sigma}$
Total	n	$Y'Y$

TABLE 2

Test	Asymptotic distribution on the	Null hypothesis
$\operatorname{tr} \hat{\Sigma}^{-1}\hat{B}X'X\hat{B}'$	chi-square $k_1 k_2$ d.f.	$B = 0$; i.e., $C_2 = 0$, $D_2 = 0$
$\operatorname{tr} \hat{\Sigma}^{-1}\hat{D}_1 S_{22.1} \hat{D}_1'$	chi-square $q_2 k_2$ d.f.	$D_2 = 0$

and n_2 independent observations, respectively, from k-variate normal populations for which we make no specification about the means, and suppose that for the population covariance matrices we have the two hypotheses, $H_1: \Sigma_1 \neq \Sigma_2$ and $H_2: \Sigma_1 = \Sigma_2 = \Sigma$.

Using the notation already introduced in Section 4, we want to determine g^* with $a = (\bar{x}_{(1)}, \bar{x}_{(2)}, V_1, V_2)$, which means that we seek the values of $t_{(i)}$, T_i, $i = 1,2$, which will maximize (cf. (5.2))

$$
\begin{aligned}
I^* &= t'_{(1)}\bar{x}_{(1)} - t'_{(1)}\mu_{(1)} - \tfrac{1}{2}t'_{(1)} \frac{\Sigma}{n_1} t_{(1)} + \operatorname{tr} T_1 V_1 + \frac{N_1}{2} \log |I - 2\Sigma T_1| \\
&\quad + t'_{(2)}\bar{x}_{(2)} - t'_{(2)}\mu_{(2)} - \tfrac{1}{2}t'_{(2)} \frac{\Sigma}{n_2} t_{(2)} + \operatorname{tr} T_2 V_2 + \frac{N_2}{2} \log |I - 2\Sigma T_1|.
\end{aligned} \tag{7.1.1}
$$

Following the procedure as used for (4.3), we find that the sought-for values are given by (cf. (5.3))

$$
\begin{aligned}
t_{(1)} &= n_1\Sigma^{-1}(\bar{x}_{(1)} - \mu_{(1)}), \qquad & t_{(2)} &= n_2\Sigma^{-1}(\bar{x}_{(2)} - \mu_{(2)}), \\
T_1 &= \tfrac{1}{2}\Sigma^{-1} - \tfrac{1}{2}S_1^{-1}, & T_2 &= \tfrac{1}{2}\Sigma^{-1} - \tfrac{1}{2}S_2^{-1},
\end{aligned} \tag{7.1.2}
$$

for which values I^* of (7.1.1) becomes

$$
\begin{aligned}
I^* &= \frac{n_1}{2} (\bar{x}_{(1)} - \mu_{(1)})'\Sigma^{-1}(\bar{x}_{(1)} - \mu_{(1)}) + \frac{n_2}{2} (\bar{x}_{(2)} - \mu_{(2)})'\Sigma^{-1}(\bar{x}_{(2)} - \mu_{(2)}) \\
&\quad + \frac{N_1}{2}\left(\log \frac{|\Sigma|}{|S_1|} - k + \operatorname{tr} S_1 \Sigma^{-1}\right) + \frac{N_2}{2}\left(\log \frac{|\Sigma|}{|S_2|} - k + \operatorname{tr} S_2 \Sigma^{-1}\right).
\end{aligned} \tag{7.1.3}
$$

For variations of $\mu_{(1)}$, $\mu_{(2)}$, and Σ, I^* will be a minimum for $\hat{\mu}_{(1)}$, $\hat{\mu}_{(2)}$, and $\hat{\Sigma}$ satisfying (see [8])

$$
\begin{aligned}
& n_1 \hat{\Sigma}^{-1}(\bar{x}_{(1)} - \hat{\mu}_{(1)}) = 0, \qquad n_2 \hat{\Sigma}^{-1}(\bar{x}_{(2)} - \hat{\mu}_{(2)}) = 0, \\
& 0 = -\frac{n_1}{2} (\bar{x}_{(1)} - \hat{\mu}_{(1)})'\hat{\Sigma}^{-1}\, d\Sigma\hat{\Sigma}^{-1}(\bar{x}_{(1)} - \hat{\mu}_{(1)}) \\
& \quad - \frac{n_2}{2} (\bar{x}_{(2)} - \hat{\mu}_{(2)})'\hat{\Sigma}^{-1}\, d\Sigma\hat{\Sigma}^{-1}(x_{(2)} - \hat{\mu}_{(2)}) + \frac{N_1}{2} \operatorname{tr} \hat{\Sigma}^{-1}\, d\Sigma \\
& \quad - \frac{N_1}{2} \operatorname{tr} S_1 \hat{\Sigma}^{-1}\, d\Sigma\hat{\Sigma}^{-1} + \frac{N_2}{2} \operatorname{tr} \hat{\Sigma}^{-1}\, d\Sigma - \frac{N_2}{2} \operatorname{tr} S_2\, \hat{\Sigma}^{-1}\, d\Sigma\hat{\Sigma}^{-1},
\end{aligned} \tag{7.1.4}
$$

from which we find that

$$
\hat{\mu}_{(1)} = \bar{x}_{(1)}, \qquad \hat{\mu}_{(2)} = \bar{x}_{(2)}, \qquad (N_1 + N_2)\hat{\Sigma} = N_1S_1 + N_2S_2 = NS, \tag{7.1.5}
$$

where $N = N_1 + N_2$; consequently (cf. Wilks [31], p. 489),

$$
2\hat{I} = N_1 \log \frac{|S|}{|S_1|} + N_2 \log \frac{|S|}{|S_2|}. \tag{7.1.6}
$$

It is readily found that the corresponding $\hat{J}$ is given by (cf. [20], p. 91)

$$\hat{J} = \frac{N_1 N_2}{2(N_1 + N_2)} (\text{tr } S_1 S_2^{-1} + \text{tr } S_2 S_1^{-1} - 2k). \tag{7.1.7}$$

It will be shown in Section 8 that $2\hat{I}$, for large N_1 and N_2, on the null hypothesis H_2, has a chi-square distribution with $k(k + 1)/2$ d.f., and to a better approximation, a non-central chi-square distribution, R. A. Fisher's B distribution.

Linear discriminant function. We seek a linear compound, the same for both samples, $y = \alpha' x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_x$, which will maximize (see (7.1.7))

$$\hat{J}'(y) = \frac{N_1 N_2}{2(N_1 + N_2)} \left(\frac{\alpha' S_1 \alpha}{\alpha' S_2 \alpha} + \frac{\alpha' S_2 \alpha}{\alpha' S_1 \alpha} - 2 \right). \tag{7.1.8}$$

The usual calculus procedures yield the result that α is obtained as a solution of $S_1 \alpha = F S_2 \alpha$, where F is a root of the determinantal equation $|S_1 - F S_2| = |N_1 S_1 - l N_2 S_2| = 0$, and $F = N_2 l / N_1$. It is found that the same linear function results from maximizing (see 7.1.6))

$$\hat{I}'(y) = \frac{N_1}{2} \log \frac{\alpha' S \alpha}{\alpha' S_1 \alpha} + \frac{N_2}{2} \log \frac{\alpha' S \alpha}{\alpha' S_2 \alpha}. \tag{7.1.9}$$

If the roots of the determinantal equation, which are almost everywhere positive, are $F_1, F_2, \cdots, F_k$ arranged in ascending order, then, as was shown in [20], Section 5, the maximum of $\hat{J}'(y)$ occurs for the linear compound associated with F_1 or F_k according as $F_1 F_k < 1$ or $F_1 F_k > 1$.

It may also be shown, readily, that

$$\begin{aligned} \hat{I} &= \hat{I}'(l_1) + \hat{I}'(l_2) + \cdots + \hat{I}'(l_k), \\ \hat{J} &= \hat{J}'(F_1) + \hat{J}'(F_2) + \cdots + \hat{J}'(F_k), \end{aligned} \tag{7.1.10}$$

where

$$\begin{aligned} \hat{I}'(l_i) &= \frac{N_1}{2} \log \frac{N_1}{N_1 + N_2} \frac{1 + l_i}{l_i} + \frac{N_2}{2} \log \frac{N_2}{N_1 + N_2} (1 + l_i) \\ &= \frac{N_1}{2} \log \frac{N_1}{N_1 + N_2} + \frac{N_2}{2} \log \frac{N_2}{N_1 + N_2} \\ &\qquad + \frac{N_1 + N_2}{2} \log (1 + l_i) - \frac{N_1}{2} \log l_i, \\ \hat{J}'(F_i) &= \frac{N_1 N_2}{2(N_1 + N_2)} \frac{(F_i - 1)^2}{F_i}. \end{aligned} \tag{7.1.11}$$

It is conjectured that for large N_1 and N_2, (assuming that the corresponding population parameters have null hypothesis values) the quantity $2\hat{I}'(l_{m+1}) + \cdots + 2\hat{I}'(l_k)$, the terms arranged in descending order of efficiency, has a chi-square distribution with $(k - m)(k - m + 1)/2$ d.f.

(7.2) *r-samples.* Suppose that we have r independent samples, respectively, of $N_1, N_2, \cdots, N_r$, independent observations each, from k-variate normal populations for which we assume the means equal, and that for the population covariance matrices we have the two hypotheses, $H_1: \Sigma_1, \Sigma_2, \cdots, \Sigma_r$ and $H_2: \Sigma_1 = \Sigma_2 = \cdots = \Sigma_r = \Sigma$. Thus, for the r samples, corresponding to H_1 and H_2, we have, respectively,

$$\sigma_{(1)} = \begin{bmatrix} \Sigma_1 & & & & & \cdots & 0 \\ & \ddots & & & & & \\ & & \Sigma_1 & & & & \\ & & & \Sigma_2 & & & \\ \vdots & & & & \ddots & & \vdots \\ & & & & & \Sigma_2 & \\ & & & & & \ddots & \\ & & & & & \Sigma_r & \\ & & & & & & \ddots \\ 0 & & & & & \cdots & \Sigma_r \end{bmatrix} \begin{matrix} N_1 \\ \\ \\ \\ N_2 \\ \vdots \\ \\ \\ \\ N_r \end{matrix}, \qquad N = N_1 + N_2 + \cdots + N_r, \tag{7.2.1}$$

$$\sigma_{(2)} = \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} \begin{matrix} N_1 \\ N_2 \\ \vdots \\ N_r \end{matrix}, \tag{7.2.2}$$

$$\begin{aligned} I(1:2) &= \tfrac{1}{2} \log \frac{|\Sigma|^{N_1+N_2+\cdots+N_r}}{|\Sigma_1|^{N_1} \cdots |\Sigma_r|^{N_r}} + \tfrac{1}{2} \operatorname{tr} \left(\begin{pmatrix} \Sigma_1-\Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r-\Sigma \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma^{-1} \end{pmatrix} \right) \\ &= \sum_{i=1}^{r} \frac{N_i}{2} \left(\log \frac{|\Sigma|}{|\Sigma_i|} + \operatorname{tr} \Sigma_i \Sigma^{-1} \right) - \frac{kN}{2}, \end{aligned} \tag{7.2.3}$$

$$\begin{aligned} J(1,2) &= \tfrac{1}{2} \operatorname{tr} \left[\begin{pmatrix} \Sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r \end{pmatrix} - \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} \right] \\ &\qquad \cdot \left[\begin{pmatrix} \Sigma^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma^{-1} \end{pmatrix} - \begin{pmatrix} \Sigma_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r^{-1} \end{pmatrix} \right] \\ &= \sum_{i=1}^{r} \frac{N_i}{2} \left(\operatorname{tr} \Sigma_i \Sigma^{-1} + \operatorname{tr} \Sigma \Sigma_i^{-1} \right) - kN. \end{aligned} \tag{7.2.4}$$

If we compute the sample values about the sample averages, then we estimate $I(1:2)$ and $J(1,2)$, by taking $N_1, N_2, \cdots, N_r$, as degrees of freedom, and replace $\Sigma_1, \Sigma_2, \cdots, \Sigma_r$, respectively, by the sample unbiased covariance matrices $S_1, S_2, \cdots, S_r$, and Σ by S, where $NS = N_1S_1 + \cdots + N_rS_r$. Thus we have (cf. Box [5] and Wilks [31], p. 489),

$$\hat{I}(1:2) = \sum_{i=1}^{r} \frac{N_i}{2}\left(\operatorname{tr} S_iS^{-1} + \log \frac{|S|}{|S_i|}\right) - \frac{kN}{2} = \sum_{i=1}^{r} \frac{N_i}{2} \log \frac{|S|}{|S_i|}, \tag{7.2.5}$$

$$\begin{aligned}\hat{J}(1,2) &= \sum_{i=1}^{r} \frac{N_i}{2}(\operatorname{tr} S_iS^{-1} + \operatorname{tr} SS_i^{-1}) - kN = \sum_{i=1}^{r} \frac{N_i}{2} \operatorname{tr} SS_i^{-1} - \frac{kN}{2} \\ &= \sum_{i<j} \frac{N_iN_j}{2N}(\operatorname{tr} S_iS_j^{-1} + \operatorname{tr} S_jS_i^{-1} - 2k).\end{aligned} \tag{7.2.6}$$

We omit at this time a discussion of linear discriminant functions for this case.

8. Asymptotic distribution of $\hat{I}(1:2)$ for the homogeneity of covariance matrices.

On the hypothesis H_2 of Section 7.2, we let

$$N_iS_i = \Sigma^{1/2}V_i\Sigma^{1/2}, \qquad NS = \Sigma^{1/2}V\Sigma^{1/2}, \qquad i = 1, 2, \cdots, r. \tag{8.1}$$

These equations define transformations linear in the elements of the matrices S_i, S or V_i, V. The Jacobians of these transformations are given by [8],

$$\left|\frac{1}{N_i}\Sigma\right|^{(k+1)/2} \quad \text{and} \quad \left|\frac{1}{N}\Sigma\right|^{(k+1)/2}.$$

The Wishart distributions of the elements of S_i, S are thereby transformed into the respective probability densities of the elements of V_i, V, given by

$$\frac{(\frac{1}{2})^{kN_i/2}e^{-(1/2)\operatorname{tr} V_i}|V_i|^{(N_i-k-1)/2}}{\pi^{k(k-1)/4}\prod_{\alpha=1}^{k}\Gamma\left(\frac{N_i+1-\alpha}{2}\right)}, \qquad \frac{(\frac{1}{2})^{kN/2}e^{-(1/2)\operatorname{tr} V}|V|^{(N-k-1)/2}}{\pi^{k(k-1)/4}\prod_{\alpha=1}^{k}\Gamma\left(\frac{N+1-\alpha}{2}\right)}. \tag{8.2}$$

Applying the transformations in (8.1) to $\hat{I}(1:2)$ in (7.2.5), we get

$$\hat{I}(1:2) = \sum_{\beta=1}^{r} \frac{N_\beta}{2}\left(\log \frac{|V|}{|V_\beta|} + k \log \frac{N_\beta}{N}\right). \tag{8.3}$$

Since the r samples are independent, the characteristic function of the distribution of

$$\sum_{\beta=1}^{r} N_\beta \log \frac{|V|}{|V_\beta|} = N \log |V| - \sum_{\beta=1}^{r} N_\beta \log |V_\beta|$$

is given by (cf. Box [5], p. 321)

$$\phi(t) = \int \left(\prod_{\beta=1}^{r} \frac{(\frac{1}{2})^{kN_\beta/2} e^{-(1/2)\operatorname{tr} V_\beta} |V_\beta|^{(N_\beta(1-2it)-k-1)/2}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N_\beta + 1 - \alpha}{2}\right)} \right) |V|^{Nit} \prod_{\beta=1}^{r} \prod_{\gamma,\delta=1}^{k} dV_{\beta\gamma\delta}$$

$$(8.4) \quad = \left(\prod_{\beta=1}^{r} \prod_{\alpha=1}^{k} \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} \right) \cdot \int \frac{(\frac{1}{2})^{kN/2} e^{-(1/2)\operatorname{tr} V} |V|^{[N(1-2it)-k-1]/2+Nit} \prod_{\gamma,\delta=1}^{k} dV_{\gamma,\delta}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N(1-2it)+1-\alpha}{2}\right)}$$

$$= \prod_{\alpha=1}^{k} \left(\frac{\Gamma\left(\frac{N+1-\alpha}{2}\right)}{\Gamma\left(\frac{N(1-2it)+1-\alpha}{2}\right)} \prod_{\beta=1}^{r} \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} \right),$$

where the middle result follows from the reproductive property of the Wishart distribution [33]. We will use Stirling's approximation

$$\log \Gamma(p) = \tfrac{1}{2} \log 2\pi + (p - \tfrac{1}{2}) \log p - p + \tfrac{1}{12}p - \tfrac{1}{360}p^3 + O(1/p^5)$$

to get an approximate value for large N_β in (8.4). We have that

$$\log \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} =$$

$$(8.5) \quad \begin{aligned} &\frac{N_\beta(1-2it)-\alpha}{2} \cdot \log \frac{N_\beta(1-2it)+1-\alpha}{2} - \frac{N_\beta(1-2it)+1-\alpha}{2} \\ &+ \frac{1}{6(N_\beta(1-2it)+1-\alpha)} - \frac{1}{45(N_\beta(1-2it)+1-\alpha)^3} \\ &- \frac{N_\beta-\alpha}{2} \log \frac{N_\beta+1-\alpha}{2} \\ &+ \frac{N_\beta+1-\alpha}{2} - \frac{1}{6(N_\beta+1-\alpha)} + \frac{1}{45(N_\beta+1-\alpha)^3} + O(1/N_\beta^5), \end{aligned}$$

and after some algebraic manipulation, the right member of (8.5) may be written as

$$-itN_\beta \log \frac{N_\beta}{2} + \frac{N_\beta(1-2it)-\alpha}{2} \log(1-2it) + N_\beta it + \frac{(3\alpha^2-1)2it}{12N_\beta(1-2it)} + O(1/N_\beta^2).$$

We therefore have

$$\log \phi(t) = \sum_{\alpha=1}^{k}\Big(itN \log \frac{N}{2} - \frac{N(1-2it)-\alpha}{2}\cdot \log(1-2it) - Nit - \frac{(3\alpha^2-1)it}{6N(1-2it)} - 0(1/N^2)\Big)$$

$$(8.6)\quad + \sum_{\alpha=1}^{k}\sum_{\beta=1}^{r} - itN_\beta \log \frac{N_\beta}{2} + \frac{N_\beta(1-2it)-\alpha}{2}\cdot\log(1-2it) + N_\beta it + \frac{(3\alpha^2-1)it}{6N_\beta(1-2it)} + 0(1/N_\beta^2)$$

$$= -it\sum_{\beta=1}^{r} kN_\beta \log\frac{N_\beta}{N} - \frac{(r-1)k(k+1)}{4}\log(1-2it) + \frac{it(2k^3+3k^2-k)}{12(1-2it)}\Big(\sum_{\beta=1}^{r}\frac{1}{N_\beta} - \frac{1}{N}\Big) + \sum_{\beta=1}^{r} 0(1/N_\beta^2) - 0(1/N^2).$$

Neglecting the last term in (8.6), we have that

$$(8.7)\quad \phi(t) = (1-2it)^{-(r-1)k(k+1)/4}\exp\Big(-it\sum_{\beta=1}^{r} kN_\beta\log\frac{N_\beta}{N} + \frac{Cit}{1-2it}\Big),$$

where $C = (2k^3+3k^2-k)(\sum_{\beta=1}^{r} 1/N_\beta - 1/N)/12$.

Because of (8.3) and (8.4), writing $\zeta = 2\hat{I}(1:2)$, the probability density of ζ is given by

$$(8.8)\qquad D(\zeta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-it\zeta + Cit/(1-2it)}\,dt}{(1-2it)^{(r-1)k(k+1)/4}}.$$

If we neglect the term with C, it follows that $D(\zeta)$ is a chi-square distribution with $(r-1)k(k+1)/2$ d.f.; otherwise, by integrating (8.8) (see [23], p. 86), we get, since ζ is real and positive and $(r-1)k(k+1)/4 > 0$,

$$(8.9)\qquad D(\zeta) = \tfrac{1}{2}e^{-C/2-\zeta/2}\Big(\frac{\zeta}{C}\Big)^{(n-1)/2} I_{n-1}(\sqrt{C\zeta}),$$

where $n = (r-1)k(k+1)/4$ and $I_{n-1}(\sqrt{C\zeta})$ is the Bessel function of purely imaginary argument [30]

$$I_{n-1}(\sqrt{C\zeta}) = \sum_{j=0}^{\infty}\frac{(\frac{1}{2})^{(n-1)/2+j}\Big(\frac{C\zeta}{2}\Big)^{(n-1)/2+j}}{j!\,\Gamma(n+j)}.$$

The distribution given by (8.9) is the non-central chi-square distribution and is Fisher's B distribution ([10, p. 14. 665) if we write $C = \beta^2$, $\zeta = B^2$, $2n = n_1$. The case for $k = 1$ is the Bartlett test for homogeneity of variance [3], [5].

The approximation to the logarithm of the characteristic function of ζ, i.e., $-n\log(1-2it) + cit/(1-2it)$, corresponds to that of Box [5], formula 29,

p. 323, retaining only the first term in his sum; i.e., $(\alpha_1/\mu)[1/(1 - 2it) - 1]$ (there is a misprint in the formula) is $Cit/(1 - 2it)$ as used here, as may be verified by using the appropriate formulas with $\beta = 0$ on pp. 324–325 of [5].

For large n we may approximate $I_{n-1}(\sqrt{C\zeta})$ in (8.9) by writing

$$\begin{aligned} I_{n-1}(\sqrt{C\zeta}) &= \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{(C\zeta/4)^j \Gamma(n)}{j!\Gamma(n+j)} \\ &\approx \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{C\zeta}{4n}\right)^j \\ &= \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} e^{C\zeta/4n} \end{aligned}$$

and thereby get

$$D(\zeta) \approx \tfrac{1}{2} \frac{e^{-C/2-\zeta[1-(C/2n)]/2}}{\Gamma(n)} \left(\frac{\zeta}{2}\right)^{n-1}. \tag{8.10}$$

If we set $\zeta[1 - (C/2n)] = \chi^2$, (8.10) yields

$$\begin{aligned} D(\chi^2)\, d\chi^2 &= \frac{e^{-C/2}}{\left(1 - \dfrac{C}{2n}\right)^n} \cdot \frac{e^{-\chi^2/2}}{\Gamma(n)} \left(\frac{\chi^2}{2}\right)^{n-1} d\frac{\chi^2}{2} \\ &\approx \frac{e^{-\chi^2/2}(\chi^2/2)^{n-1}\, d\chi^2/2}{\Gamma(n)}, \end{aligned} \tag{8.11}$$

or $\zeta[1 - (C/2n)]$ asymptotically has a chi-square distribution with

$$2n = (r-1)k(k+1)/2 \text{ d.f.}$$

It is readily verified that $1 - (C/2n) = \rho$, Box's scale factor in the chi-square approximation ([5], p. 329).

For other approximations to (8.9), see Abdel-Aty [1].

Examples: (a). For the first example we use the data given by Smith [29], Table 2, which he used to calculate a linear discriminant function for a group of 25 normal persons and 25 psychotics. Here $k = 2$, $r = 2$,

$$S_1 = \begin{pmatrix} 6.92 & -5.27 \\ -5.27 & 40.89 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 36.75 & 13.92 \\ 13.92 & 287.92 \end{pmatrix}, \quad S = \begin{pmatrix} 21.83 & 4.33 \\ 4.33 & 164.40 \end{pmatrix},$$

$$N_1 = N_2 = 24, \quad N = 48, \quad |S_1| = 255.1859, \quad |S_2| = 10387.2936, \quad |S| = 3570.1031,$$

$$2\hat{I} = 24 \log (3570.1031/255.1859) + 24 \log (3570.1031/10387.2936) = 37.7268,$$

$$C = (16 + 12 - 2)(2/24 - 1/48)/12 = .135416 = \beta^2, \quad \beta = .368,$$

$$n = (2-1)(2)(3)/4, \quad 2n = 3 = n_1,$$

$$\zeta = 2\hat{I} = 37.7268 = B^2, \quad B = 6.14.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 3$ and $\beta = .2$ and $.4$ to be, respectively, 2.8140 and 2.8680. We therefore reject the null hypothesis of equality of the population covariance matrices. Smith [29] does remark that the correlations are not significant, but the variances of the psychotics are significantly greater than those of the normals.

(b) For the second example, we use the data given by Kossack [19] for a problem of classifying an A.S.T.P. pre-engineering trainee as to whether he would do unsatisfactory or satisfactory work in his first-term mathematics course. The three variables used are x_1, a mathematics placement test score; x_2, a high school mathematics score; x_3, the Army General Classification Test score. There were 96 trainees who did unsatisfactory work and 209 who performed satisfactory work. Here $k = 3$, $r = 2$, $N_1 = 95$, $N_2 = 208$, $N = 303$,

$$S_1 = \begin{pmatrix} 133.8592 & 7.0572 & 2.0717 \\ 7.0572 & 4.1288 & -2.0109 \\ 2.0717 & -2.0109 & 27.7016 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 217.1505 & 14.0692 & 35.7085 \\ 14.0692 & 3.9820 & .4031 \\ 35.7085 & .4031 & 72.7206 \end{pmatrix},$$

$$S = \begin{pmatrix} 191.04 & 11.871 & 25.162 \\ 11.871 & 4.0280 & -.35378 \\ 25.162 & -.35378 & 58.606 \end{pmatrix}, \quad |S_1| = 13313, \quad |S_2| = 43779,$$

$$|S| = 34053, 2\hat{I} = 95 \log \frac{34053}{13313} + 208 \log \frac{34053}{43779} = 227.0867,$$

$$C = (54 + 27 - 3)(1/95 + 1/208 - 1/303)/12 = .078221 = \beta^2, \quad \beta = .28,$$

$$n = (2 - 1)(3)(4)/4, \quad 2n = 6 = n_1,$$

$$\zeta = 2\hat{I} = 227.0867 = B^2, \quad B = 15.06.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 6$ and $\beta = .2$ and $.4$ to be, respectively, 3.5602 and 3.5951. We therefore reject the null hypothesis of equality of the population covariance matrices. An assumption of equality is, however, implicit in the procedure used by Kossack.

(c) For the third example, we use the data given by Pearson and Wilks [24], for five samples of twelve observations each on the strength and hardness in aluninum die-castings. Based on their data (note that they did not use the unbiased estimates), the details of which are not repeated here,

$$k = 2, \quad r = 5, \quad N_1 = \cdots = N_5 = 11, \quad N = 55,$$

$$\log |S_1| = 5.82588, \quad \log |S_2| = 6.63942, \quad \log |S_3| = 5.31904,$$

$$\log |S_4| = 6.66973, \quad \log |S_5| = 5.35937, \quad \log |S| = 6.13953,$$

$$2\hat{I} = 55(6.13953) - 11(29.81344) = 9.726,$$

$$C = (16 + 12 - 2)(5/11 - 1/55)/12 = .945454 = \beta^2, \quad \beta = .972,$$

$$n = (5 - 1)(2)(3)/4, \quad 2n = 12 = n_1,$$

$$\zeta = 2\hat{I} = 9.726 = B^2, \quad B = 3.12.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 7$ (the largest there tabulated), and $\beta = 0.8$ and 1.0 to be, respectively, 3.9144 and 4.0005. Since the tabulated values increase with increasing n_1 for a fixed β, we do not, in this case, reject the null hypothesis of equality of population covariance matrices. This is consistent with the conclusion reached by Pearson and Wilks [24].

9. Asymptotic distribution of $\hat{I}'_R$. In (4.9) we defined $\hat{I}'_R$ and made certain statements about its asymptotic distribution which we will now confirm.

It is known that the logarithm of the characteristic function of the distribution of $2\hat{I}'_R$ is given by (see [31], p. 492; [4])

$$(9.1)\quad \log \phi(t) = (k-1)\log \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N}{2} - Nit\right)} + \sum_{\alpha=1}^{k-1} \log \frac{\Gamma\left(\frac{N(1-2it)-\alpha}{2}\right)}{\Gamma\left(\frac{N-\alpha}{2}\right)}.$$

Employing Stirling's approximation as in (8.5), and retaining comparable terms as in (8.7), we have

$$(9.2)\quad \log \phi(t) = -\frac{k(k-1)}{4}\log(1-2it) + \frac{Cit}{1-2it},$$

where $C = k(k-1)(2k+5)/12N$.

The statement at the end of Section 4 then follows from (9.2), (8.8), and (8.9). From (8.11) we may also deduce that

$$2\hat{I}'_R\left(1 - \frac{k(k-1)(2k+5)}{6Nk(k-1)}\right) = -(N - \tfrac{1}{6}(2k+5))\log|R|$$

asymptotically has a chi-square distribution with $k(k-1)/2$ d.f. This latter result is given by Bartlett [4].

10. Asymptotic distribution of $\hat{J}(1,2)$ for the linear hypothesis. From results derived by Fisher [9], Girshick [11], Hsu [15], [16], [17], and Roy [27], it is known that the probability density of the distribution of the roots of $|S^* - lS| = 0$ (see (5.18)), for $(n-r)$ large, is given by

$$(10.1)\quad \frac{(\frac{1}{2})^{(r-1)p/2}\pi^{p/2}(l_1\cdots l_p)^{(r-p-2)/2}}{\prod_{\alpha=1}^{p}\Gamma\left(\frac{r-\alpha}{2}\right)\Gamma\left(\frac{p+1-\alpha}{2}\right)} e^{-\frac{1}{2}(l_1+\cdots+l_p)}\prod_{i>j}(l_j - l_i)$$

and that of the roots of $|S_{21}S_{11}^{-1}S_{12} - lS_{22\cdot1}| = 0$ (see (6.3), (6.4)), for $(n-k_1)$ large, is given by

$$(10.2)\quad \frac{(\frac{1}{2})^{k_1k_2/2}\pi^{k_2/2}(V_1\cdots V_{k_2})^{(k_1-k_2-1)/2}}{\prod_{\alpha=1}^{k_2}\Gamma\left(\frac{k_1+1-\alpha}{2}\right)\Gamma\left(\frac{k_2+1-\alpha}{2}\right)} e^{-\frac{1}{2}(V_1+\cdots+V_{k_2})}\prod_{i>j}(V_j - V_i),$$

where $V_i = (n-k_1)l_i$.

From (10.1) and (10.2) it is readily derived that the characteristic functions of the asymptotic distributions of $2\hat{I}(1:2) = \hat{J}(1,2)$ in (5.18) and (6.4) are, respectively, $(1 - 2it)^{-(r-1)p/2}$ and $(1 - 2it)^{-k_1k_2/2}$, whence the conclusion as to their chi-square distributions. The chi-square decompositions in Section 5 and Section 6 follow from the fact that asymptotically the distributions of

$$l_{m+1}, \cdots, l_p$$

of (10.1) and $V_{m+1}, \cdots, V_{k_2}$ of (10.2), assuming that the corresponding population parameters have the null hypothesis values, are independent of the distribution of the remaining roots and are given, respectively, by

$$(10.3)\qquad \frac{(\frac{1}{2})^{(r-1-m)(p-m)/2}\pi^{(p-m)/2}}{\prod_{\alpha=m+1}^{p}\Gamma\left(\frac{r-m-\alpha}{2}\right)\Gamma\left(\frac{p-m+1-\alpha}{2}\right)} (l_{m+1}\cdots l_p)^{(r-p-2)/2}e^{-\frac{1}{2}(l_{m+1}+\cdots+l_p)}\prod_{i>j}(l_j - l_i),$$

$$(10.4)\qquad \frac{(\frac{1}{2})^{(k_1-m)(k_2-m)/2}\pi^{(k_2-m)/2}}{\prod_{\alpha=m+1}^{k_2}\Gamma\left(\frac{k_1-m+1-\alpha}{2}\right)\Gamma\left(\frac{k_2-m+1-\alpha}{2}\right)} (V_{m+1}\cdots V_{k_2})^{(k_1-k_2-1)/2}e^{-\frac{1}{2}(V_{m+1}+\cdots+V_{k_2})}\prod_{i>j}(V_j - V_i).$$

The characteristic function of the distribution of $2\hat{I}$ of (7.1.10) could also have been derived from the distribution of the roots of $|N_1S_1 - lN_2S_2| = 0$, given by,

$$(10.5)\qquad \pi^{k/2}\left(\prod_{\alpha=1}^{k}\frac{\Gamma\left(\frac{N_1+N_2+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_1+1-\alpha}{2}\right)\Gamma\left(\frac{N_2+1-\alpha}{2}\right)\Gamma\left(\frac{k+1-\alpha}{2}\right)}\right) \frac{(l_1\cdots l_k)^{(N_1-k-1)/2}\prod_{i>j}(l_j - l_i)}{((1+l_1)\cdots(1+l_k))^{(N_1+N_2)/2}}.$$

11. Concluding remarks. The validity of the conjecture at the end of Section 7.1 is under investigation, as well as the distributions of $\hat{J}$ and $\hat{J}'(F_i)$ of Section 7, and related power functions.

It might also be mentioned that we have a basis for assessing the cost of trading observations for dimensions. If there is more than one significant linear discriminant function, then N_1 observations with the linear function associated with λ_1 (one dimension) would be as effective as N observations with the original multidimensional variables, where $NJ(1,2) = N_1J'(1,2;\lambda_1)$. Similar conclusions hold for more than one linear function.

Procedures similar to those used herein to estimate $I(1:2)$ and $J(1,2)$ are also applicable to problems of testing appropriate hypotheses for other than normal populations.

I am indebted to the referees for comments which have contributed to improvements in this paper.

REFERENCES

[1] S. H. ABDEL-ATY, "Approximate formulae for the percentage points and the probability integral of the non-central X^2 distribution," *Biometrika*, Vol. 41 (1954), pp. 538–540.

[2] T. W. ANDERSON, "The asymptotic distribution of certain characteristic roots and vectors," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 103–130.

[3] M. S. BARTLETT, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc. London*, Ser. A, Vol. 160 (1937), pp. 268–282.

[4] M. S. BARTLETT, "Tests of significance in factor analysis," *Brit. J. Psychology, Stat. Sec.*, Vol. 3 (1950), pp. 77–85.

[5] G. E. P. BOX, "A general distribution theory for a class of likelihood criteria," *Biometrika*, Vol. 36 (1949), pp. 317–346.

[6] H. CRAMÉR, "Sur un nouveau théorème-limite de la théorie des probabilitiés," *Actualités Scientifiques et Industrielles*, No. 736, Hermann & Cie, Paris, 1938.

[7] H. E. DANIELS, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 631–650.

[8] W. L. DEEMER AND I. OLKIN, "The Jacobians of certain matrix transformations useful in multivariate analysis," *Biometrika*, Vol. 38 (1951), pp. 345–367.

[9] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[10] R. A. FISHER, *Contributions to Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1950.

[11] M. A. GIRSHICK, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203–224.

[12] S. W. GREENHOUSE, "On the problem of discrimination between statistical populations," M. A. Thesis, The George Washington University, 1954.

[13] H. HOTELLING, "A generalized T test and measure of multivariate dispersion," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 23–41.

[14] J. P. HOYT, "Estimates and asymptotic distributions of certain statistics in information theory," Dissertation, The Graduate Council of The George Washington University, 1953.

[15] P. L. HSU, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[16] P. L. HSU, "On the limiting distribution of the canonical correlations," *Biometrika*, Vol. 32 (1941–42), pp. 38–45.

[17] P. L. HSU, "On the limiting distribution of roots of a determinantal equation," *J. London Math. Soc.*, Vol. 16 (1941), pp. 183–194.

[18] A. I. KHINCHIN, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, New York, 1949.

[19] C. F. KOSSACK, "On the mechanics of classification," *Ann. Math. Stat.*, Vol. 16 (1945) pp. 95–98.

[20] S. KULLBACK, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.

[21] S. KULLBACK, "Certain inequalities in information theory and the Cramér-Rao inequality," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 745–751.

[22] S. KULLBACK AND R. A. LEIBLER, "On information and sufficiency," *Ann. Math. Stat.* Vol. 22 (1951), pp. 79–86.

[23] N. W. McLachlan, *Complex Variable and Operational Calculus with Technical Applications*, Cambridge University Press, 1939.

[24[E. S. Pearson and S. S. Wilks, "Methods of statistical analysis appropriate for k samples of two variables," *Biometrika*, Vol. 25 (1933), pp. 353–378.

[25] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., New York, 1952.

[26] D. D. Rippe, "Statistical rank and sampling variation of the results of factorization of covariance matrices," Doctoral Thesis on file at the University of Michigan, 1951.

[27] S. N. Roy, "p-statistics, or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.

[28] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220–238.

[29] C. A. B. Smith, "Some examples of discrimination," *Ann. Eugenics*, Vol. 13 (1947), pp. 272–282.

[30] G. N. Watson, *Bessel Functions*, (2d ed.), The Macmillan Co., New York, 1944.

[31] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–491.

[32] S. S. Wilks, "The large sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 60–62.

[33] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1943.