THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

DIRECTION AND COLLINEARITY STATISTICS

IN DISCRIMINANT ANALYSIS

by

A. M. Kshirsagar

Technical Report No. 43
Department of Statistics THEMIS Contract

August 20, 1969

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

DIRECTION AND COLLINEARITY STATISTICS

IN DISCRIMINANT ANALYSIS

by

A. M. Kshirsagar*
Southern Methodist University
Dallas, Texas 75222

1. INTRODUCTION.

If samples of sizes $n_1, n_2, \cdots, n_k$ are available from $k(=q+1)$ multivariate normal populations with different mean vectors $\underline{\mu}_\alpha (\alpha = 1, \cdots, k)$ and the same variance-covariance matrix (Σ) and if $\underline{x}' = [x_1, x_2, \cdots, x_p]$ denotes the vector of p variates on which measurements are made, one obtains the following multivariate analysis of variance table:

source	degrees of freedom (d.f.)	p × p matrix of sums of squares and sums of products (s.s. and s.p.)	
Between Groups	q	B_x	
Within Groups	$n - q$	W_x	
TOTAL	$n_1 + \cdots + n_k - 1 = n$	$B_x + W_x$	(1.1)

Sometimes one is interested in testing whether a single hypothetical function $\underline{\ell}'\underline{x} = \sum_1^p \ell_i x_i$ is adequate for discrimination among the k groups or populations. This hypothesis, H, of goodness of fit of $\underline{\ell}'\underline{x}$ consists of two aspects, (i) collinearity aspect, i.e., whether a single discriminant function could be adequate at all, and (ii) direction aspect, i.e.,

---

*This research was sponsored by the Office of Naval Research Contract No. N00014-68-A-0515. Reproduction in whole or in part is permitted for any purpose of the United States Government.

whether the single function which could discriminate adequately, agrees with the assigned function $\underline{\ell}'\underline{x}$. The name collinearity aspect follows from the fact that a single discriminant function is adequate only when the group means are collinear. Bartlett (1951) constructed an over-all criterion for testing H and then factorized it into two factors corresponding to the direction and collinearity aspects of H. Wilks' Λ criterion is $\Lambda = |W|/|B + W|$ and the 'residual' Wilks' Λ, when the hypothetical function is eliminated, is

$$\Lambda_R = \Lambda/\{\underline{\ell}'W_x\underline{\ell}/\underline{\ell}'(B_x + W_x)\underline{\ell}\} = \Lambda/\Lambda_1 \tag{1.2}$$

This is the over-all criterion and Bartlett (1951) has factorized it as

$$\Lambda_R = \Lambda_2 \cdot \Lambda_3 \quad ,$$

where

$$\Lambda_2 = \frac{1 - \underline{\ell}'B_x(B_x + W_x)^{-1}B_x\underline{\ell}/\underline{\ell}'B_x\underline{\ell}}{\Lambda_1} \tag{1.3}$$

and

$$\Lambda_3 = \frac{\Lambda}{\Lambda_1\Lambda_2} \quad . \tag{1.4}$$

Λ_2 is the direction factor and Λ_3 is the 'partial' collinearity factor. There is an alternative factorization also viz.

$$\Lambda_R = \Lambda_4 \cdot \Lambda_5 \tag{1.5}$$

where,

$$\Lambda_4 = \left|W_x + \frac{B_x\underline{\ell}\underline{\ell}'B_x}{\underline{\ell}'B_x\underline{\ell}}\right| / |W_x + B_x| \tag{1.6}$$

and

$$\Lambda_5 = \Lambda/\Lambda_1\Lambda_4 \quad . \tag{1.7}$$

Λ_4 is the collinearity factor and Λ_5 is the 'partial' direction factor. The distributions of these factors under the null hypothesis H have been derived by Bartlett by an ingenuous geometrical device. In this paper, a simpler representation of these factors is given and this immediately leads to an analytical method of derivation of their distributions. The method is then extended to the test of goodness of fit of more than one hypothetical discriminant functions.

2. DISTRIBUTIONS OF Λ_2 , Λ_3 , Λ_4 AND Λ_5 .

If H is true, $\underline{\ell}'\underline{x}$ is the only variable, that has different mean values in the k groups and any other linear combination of $\underline{x}$, uncorrelated with $\underline{\ell}'\underline{x}$ has the same mean in all the groups and is of no use for discrimination. There is no loss of generality in assuming $\underline{\ell}'\Sigma\underline{\ell} = 1$, as otherwise we can replace $\underline{\ell}'\underline{x}$ by $\underline{\ell}'\underline{x}/(\ell'\Sigma\ell)^{1/2}$ and this does not alter Λ_2 , Λ_3 , Λ_4 and Λ_5 given by (1.3), (1.4), (1.6) and (1.7). Let, therefore, L be a p × p matrix with $\underline{\ell}'$ as its first row and such that $L\Sigma L' = I$. Transform from $\underline{x}$ to $\underline{z} = [z_1 , z_2 , \cdots , z_p]'$ by

$$\underline{z} = L\underline{x} \quad . \tag{2.1}$$

Then z_1 is our hypothetical discriminant function; $z_2 , \cdots , z_p$ are all uncorrelated with z_1, as $L\Sigma L' = I$, and under the null hypothesis H , they have the same mean in all the k groups. For $\underline{z}$, the "Between Groups" matrix, as obtained from (1.1), is

$$B = LB_xL' = [b_{ij}] \tag{2.2}$$

and the "Within Groups" matrix is

$$W = LW_xL' = [w_{ij}] \quad . \tag{2.3}$$

It is obvious that W has the Wishart density

$$W_p(W|n-q) = \text{const.}\,|W|^{\frac{(n-q)-p-1}{2}} e^{-\frac{1}{2}\,\mathrm{tr}\,W} \quad , \tag{2.4}$$

while the matrix B has an independent non-central Wishart distribution. However, the non-centrality is due only to the differences in the means of z_1 alone in the k groups and thus effects b_{11} only. b_{11} is a non-central χ^2 with q d.f., so B has the density

$$W_p(B|q)\,\phi(b_{11}) \quad , \tag{2.5}$$

where $\phi(b_{11})$ is the contribution due to non-centrality. For our purpose, an explicit expression for $\phi(b_{11})$ is not necessary but can be easily written down if required.

It is a well-known property of the Wishart distribution (2.4) that

$$w_{11}\ ,\ \frac{w_{i1}}{\sqrt{w_{11}}}\ (i = 2, \cdots, p) \quad \text{and} \quad w^*_{ij} = w_{ij} - \frac{w_{i1}w_{j1}}{w_{11}}\ (i,j = 2,\cdots, p) \tag{2.6}$$

are independently distributed; $\frac{w_{i1}}{\sqrt{w_{11}}}$ are standard normal and independent variables while the matrix $W^* = [w^*_{ij}]$ of order p-1 has the Wishart distribution $W_{p-1}(W^*|n-q-1)dW^*$, based on n-q-1 d.f. This result follows very easily from the Bartlett decomposition (See Kshirsagar (1959).) of a Wishart matrix. In a similar manner, b_{11} , $\frac{b_{i1}}{\sqrt{b_{11}}}$ (i = 2, $\cdots$, p) ,

$b^*_{ij} = b_{ij} - \dfrac{b_{i1}b_{j1}}{b_{11}}$ $(i,j = 2, \cdots, p)$ are independently distributed. $\dfrac{b_{i1}}{\sqrt{b_{11}}}$ are standard normal independent variables; $B^* = [b^*_{ij}]$ has the Wishart distribution $W_{p-1}(B^*|q-1)dB^*$. It is easy to see, therefore, that

$$\xi_i = \sqrt{\frac{b_{11}w_{11}}{b_{11} + w_{11}}}\left(\frac{b_{i1}}{b_{11}} - \frac{w_{i1}}{w_{11}}\right) \quad , \quad (i = 2, \cdots, p) \tag{2.7}$$

are standard normal variables and are independently distributed of B^* and W^* . Let $\underset{\sim}{\xi}$ denote the column vector of $\xi_i (i = 2, \cdots, p)$.

We now use the following well-known results connected with Wishart matrices:

If an $r \times r$ symmetric, positive definite matrix A has the Wishart distribution $W_r(A|f)dA$, and if $\underline{u}_\alpha (\alpha = 1, 2, \cdots, s)$ are s independent r-vectors (column) of standard normal independent variables, independently distributed of A, then

$$y = |A|/|A + \sum_{\alpha=1}^{s} \underline{u}_\alpha \underline{u}'_\alpha|$$

is independently distributed of

$$A + \sum_{\alpha=1}^{s} \underline{u}_\alpha \underline{u}'_\alpha \tag{2.8}$$

(See Kshirsagar (1961, 1964).). The distribution of y , in Bartlett's (1951) notation, is called the $\Lambda(f + s , r , s)$ distribution. To be more precise, y is distributed as $\prod_1^s \eta_i$, where $\eta_i (i = 1, \cdots, s)$ are independent and have the distribution

$$\text{const. } \eta_i^{\frac{f+1-i}{2} - 1} (1 - \eta_i)^{\frac{s}{2} - 1} d\eta_i \tag{2.9}$$

Applying this result to the three independent quantities W^* , B^* and ξ , we find immediately that

$$C_1 = \frac{|W^*|}{|W^* + B^*|} \tag{2.10}$$

has the $\Lambda(n-2\ ,\ p-1\ ,\ q-1)$ distribution and is independent of $W^* + B^*$ and hence also of

$$D_1 = \frac{|W^* + B^*|}{|W^* + B^* + \underline{\xi}\underline{\xi}'|}\ , \tag{2.11}$$

which has the $\Lambda(n-1\ ,\ p-1\ ,\ 1)$ distribution. Alternatively,

$$D_2 = \frac{|W^*|}{|W^* + \underline{\xi}\underline{\xi}'|} \tag{2.12}$$

has the $\Lambda(n-q\ ,\ p-1\ ,\ 1)$ distribution and is independent of $W^* + \underline{\xi}\underline{\xi}'$ and hence of

$$C_2 = \frac{|W^* + \underline{\xi}\underline{\xi}'|}{|W^* + \underline{\xi}\underline{\xi}' + B^*|}\ , \tag{2.13}$$

which has the $\Lambda(n-1\ ,\ p-1\ ,\ q-1)$ distribution.

In the next section we show that $C_1 = \Lambda_3$, $D_1 = \Lambda_2$, $C_2 = \Lambda_4$ and $D_2 = \Lambda_5$. This then completes the derivation of the distributions of of the direction and collinearity factors of Wilks' Λ , under H .

3. DIRECTION AND COLLINEARITY FACTORS.

Let $\underline{k}'$ denote the row vector

$$\left[\sqrt{b_{11}}\ ,\ b_{21}/\sqrt{b_{11}}\ ,\ \cdots\ ,\ b_{p1}/\sqrt{b_{11}}\right] \tag{3.1}$$

and let L denote the matrix

$$\left[\begin{array}{c|c} w_{11} & w_{12}\ w_{13}\ \cdots\cdots\ w_{1p} \\ \hline w_{21} & \\ \cdot & w_{ij} + b^*_{ij} \\ \cdot & \\ \cdot & (i,j = 2, \cdots , p) \\ w_{p1} & \end{array}\right] , \tag{3.2}$$

Then it is easy to see that,

$$B + W = L + \underline{k}\underline{k}' \quad . \tag{3.3}$$

By using the well-known result,

$$\begin{array}{c} 1 \\ p-1 \end{array} \left|\begin{array}{c|c} a & \underline{x}' \\ \hline \underline{x} & P \end{array}\right| = a \left| P - \frac{\underline{x}\underline{x}'}{a} \right| \tag{3.4}$$

$$\qquad\quad 1 \quad p-1$$

we find that

$$|B + W| = (b_{11} + w_{11}) \, |W^* + B^* + \underline{\xi}\underline{\xi}'| \tag{3.5}$$

and

$$|L| = w_{11} |W^* + B^*| \quad . \tag{3.6}$$

From (1.3), (2.2), (2.3), (3.3) and from the fact that $\underline{\ell}'$ is the first row of L , it is easy to see that

$$\Lambda_2 = \frac{1 - \dfrac{\underline{\ell}'\{L^{-1}BL'^{-1}\}\{L'(B + W)L\}\{L^{-1}BL'^{-1}\}\underline{\ell}}{\underline{\ell}'\{L^{-1}BL'^{-1}\}\underline{\ell}}}{\underline{\ell}'\{L^{-1}WL'^{-1}\}\underline{\ell}/\underline{\ell}'\{L^{-1}(B + W)L'^{-1}\}\underline{\ell}}$$

$$\Lambda_2 = \frac{(b_{11} + w_{11})}{w_{11}} \cdot \{1 - \underline{k}'(L + \underline{k}\underline{k}')^{-1}\underline{k}\}$$

$$= \frac{(b_{11} + w_{11})}{w_{11}} \cdot \frac{1}{1 + \underline{k}'L^{-1}\underline{k}}$$

$$= \frac{(b_{11} + w_{11})}{w_{11}} \cdot \frac{|L|}{|L + \underline{k}\underline{k}'|} \tag{3.7}$$

$$= \frac{|W^* + B^*|}{|W^* + B^* + \underline{\xi}\underline{\xi}'|} \quad . \tag{3.8}$$

The last step follows from (3.3), (3.5) and (3.6). This shows that Λ_2 is the same as D_1 of (2.11). It is easy to see that

$$\Lambda_1 = w_{11}/(b_{11} + w_{11}) \tag{3.9}$$

and so from (1.4), (3.3) and (3.7),

$$\Lambda_3 = \frac{\Lambda}{\Lambda_1\Lambda_2} = \frac{|W|}{|L|} \tag{3.10}$$

$$= \frac{w_{11}|W^*|}{w_{11}|W^* + B^*|} \quad . \tag{3.11}$$

The last step follows from (3.6). Thus Λ_3 is the same as C_1 given by (2.10).

Similarly, for the alternative factorization, from (1.6), (2.2) and (2.3), it follows that

$$\Lambda_4 = |W + \underline{k}\underline{k}'|/|W + B| \tag{3.12}$$

$$= |W^* + \underline{\xi}\underline{\xi}'|/|W^* + B^* + \underline{\xi}\underline{\xi}'| \quad . \tag{3.13}$$

The last step is a result of the application of (3.4) to $W + \underline{k}\underline{k}'$ and using (3.5), together with

$$\xi_i\xi_j = \frac{b_{i1}b_{j1}}{b_{11}} + \frac{w_{i1}w_{j1}}{w_{11}} - \frac{(b_{i1} + w_{i1})(b_{j1} + w_{j1})}{b_{11} + w_{11}} \quad . \tag{3.14}$$

Thus Λ_4 is C_2 and so, from (1.7), (3.9) and (3.12),

$$\begin{aligned} \Lambda_5 &= \Lambda/\Lambda_1\Lambda_4 \\ &= \frac{(b_{11} + w_{11})}{w_{11}} \cdot \frac{|W|}{|W + \underline{\underline{k}}\underline{\underline{k}}'|} \\ &= |W^*|/|W^* + \underline{\underline{\xi}}\underline{\underline{\xi}}'| \end{aligned} \tag{3.15}$$

on account of the application of (3.4) to W and $W + \underline{\underline{k}}\underline{\underline{k}}'$. Thus Λ_5 and D_2 are the same.

REFERENCES

Bartlett, M. S. (1951). "Goodness of fit of a single hypothetical discriminant function in the case of several groups," Ann. Eugen., 16, 199-214.

Kshirsagar, A. M. (1959). "Bartlett decomposition and Wishart distribution," Ann. Math. Stat., 30, 239-41.

Kshirsagar, A. M. (1961). "The non-central multivariate Beta distribution," Ann. Math. Stat., 32, 104-11.

Kshirsagar, A. M. (1964). "Wilks' Λ ," J. Ind. Stat. Asso., 2, 1-20.