THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

GOODNESS OF FIT OF A DISCRIMINANT FUNCTION FROM THE VECTOR SPACE OF DUMMY VARIABLES

by

A. M. Kshirsagar

Technical Report No. 42
Department of Statistics THEMIS Contract

August 14, 1969

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

DEPARTMENT OF STATISTICS
Southern Methodist University

GOODNESS OF FIT OF A DISCRIMINANT FUNCTION

FROM THE VECTOR SPACE OF DUMMY VARIABLES

by

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

1. INTRODUCTION.

If samples of independent observations are available from q + 1 groups, on p variables $\underline{x}' = [x_1 , x_2 , \cdots , x_p]$ having a p-variate normal distribution with variance-covariance matrix Σ , we shall get the following multivariate analysis of variance table:

(1.1)

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
Within groups	n - q	W
TOTAL	n	T = B + W

The matrix B can be looked upon as the matrix of regression s.s. and s.p. of $\underline{x}$ on q "dummy" variables $\underline{y}' = [y_1 , y_2 , \cdots , y_q]$, representing the contrasts among the q + 1 groups [See [3].]. The problem of discrimination among these groups then reduces to the study of the relationship between the vector variates $\underline{x}$ and $\underline{y}$. One particular case of interest occurs

*This research was sponsored by the Office of Naval Research, Contract No. N00014-68-A-0515, Project No. NR 042-260.

when the group means are collinear, or which is the same as, when there is only one true non-zero canonical correlation ρ, between $\underline{x}$ and $\underline{y}$. A single discriminant function is adequate in such a situation and it is either the canonical variate corresponding to ρ in the x-space or the y-space. In some situations, observations on one "dummy" variable out of $y_1, y_2, \cdots, y_q$ (or on one linear combination of the y's) are available, in addition to those on $\underline{x}$. For example, for the data on Egyptian skulls analysed by Barnard [1] and later by Bartlett [2], Rao [10], Williams [12] and Kshirsagar [7], observations on a variable t, denoting time corresponding to the 4 series of skulls were available and the data was analysed to find out whether time could be regarded as a single discriminant function adequate for discriminating among the four series. The hypothesis under consideration is, thus, of goodness of fit of a single discriminant function assigned from the "dummy" variables space, namely the y-space. The hypothesis, however comprises of two aspects: (i) collinearity aspect, i.e., whether one discriminant function is adequate at all, and (ii) direction aspect, i.e., whether the proposed or assigned discriminant function agrees in direction with the true one. Bartlett [3] obtained an over-all criterion for the hypothesis and then factorized it into two independent test statistics for the two aspects of collinearity and direction, of the hypothesis but he considered the proposed discriminant function to be from the x-space. Kshirsagar [7], following Williams' [12] analysis of the Egyptian skulls data, derived general expressions for the direction and collinearity tests, when the assigned discriminant function comes from the y-space. If S_{xt} is the $p \times 1$ vector of corrected s.p. of the proposed discriminator t with $\underline{x}$ and S_{tt} is the corrected s.s. of observations on t, and if $\underline{b} = T^{-1}S_{xt}$, the direction and 'partial' collinearity statistics are respectively,

$$(1.2) \qquad \Lambda'' = \frac{\underline{b}'W\,\underline{b}}{\underline{b}'S_{xt}} \cdot \frac{S_{tt}}{S_{tt} - \underline{b}'S_{xt}} \quad ,$$

$$(1.3) \qquad \Lambda''' = \frac{|W|}{|T|} \cdot \frac{\underline{b}'S_{xt}}{\underline{b}'W\,\underline{b}} \quad .$$

There is an alternative factorization also, yielding collinearity and 'partial' direction statistics. They are respectively

$$(1.4) \qquad \Lambda^{IV} = \frac{|W|}{|T|} \cdot \frac{S_{tx}W^{-1}S_{xt}}{\underline{b}'S_{xt}} \quad ,$$

$$(1.5) \qquad \Lambda^{V} = \frac{S_{tt}}{S_{tt} - b'S_{xt}} \cdot \frac{b'S_{xt}}{S_{tx}W^{-1}S_{xt}} \quad .$$

The distributions of these statistics and their independence have not been explicitly derived so far. The symmetry or "duality" in the relationship of $\underline{x}$ and $\underline{y}$, in the absence of any true association, has been noted by several people [See for example Fisher [4], Bartlett [3], Khatri [5].]. Bartlett's derivation, when the proposed discriminator comes from the x-space, is based on geometrical devices. Kshirsagar [8] proved the same results by analytical method but that method does not work, when the proposed discriminant function is from the y-space, because, even though there is 'duality' of relationship between $\underline{x}$ and $\underline{y}$, there is a fundamental distinction that $\underline{x}$ are random variables while $\underline{y}$ are "dummy" or "pseudo" variables and are fixed. The aim of this paper is to remove this lacuna and obtain the distributions of Λ'' , Λ''' , Λ^{IV} , and Λ^{V} analytically. The method used for this purpose is the elegant method of random orthogonal transformation, described by Wijsman [11].

2. WILKS' Λ DISTRIBUTION.

Before actually proceeding to derive the distributions, we state the following results, which are well known in connection with Wilks' Λ ; (See for example Kshirsagar [6], [9].).

If A_1 and A_2 are two symmetric positive definite matrices of order r , having independent Wishart distributions with the same variance covariance matrix and respective d.f. f_1 and f_2 , then

$$(2.1) \qquad A = (A_1 + A_2)^{-1/2} A_1 (A_1 + A_2)^{-1/2}$$

is independently distributed of $A_1 + A_2$ and has the multivariate Beta distribution

$$(2.2) \quad B_r(A|f_1|f_2)\,dA = \text{const.}\,|A|^{\frac{f_1-p-1}{2}}\,|I - A|^{\frac{f_2-p-1}{2}}\,dA \;.$$

The statistic $|A|$ is known as Wilks' Λ and involves f_1 , f_2 , and r as parameters and according to Bartlett's notation, it is said to have a $\Lambda(f_1 + f_2\ , r\ , f_2)$ distribution. If A is expressed as TT' where $T = [t_{ij}]$ is a lower triangular matrix, then t_{ii}^2 are independently distributed $(i = 1, 2, \cdots , r)$ and have the density $B_1(t_{ii}^2|f_1 + i - 1|f_2)$. Conversely if t_{ii}^2 has this density, $|A| = \prod_1^r t_{ii}^2$ has $\Lambda(f_1 + f_2\ , r\ , f_2)$ distribution.

3. DISTRIBUTIONS OF THE DIRECTION AND COLLINEARITY FACTORS Λ'' AND Λ''' .

For the table (1.1), it is well known that W has the Wishart distribution with $n - q$ d.f. independent of B but, as the group means are not identical, B will have a non-central Wishart distribution. However, under the null hypothesis of goodness of fit of the proposed discriminator t , if we remove from B , the regression s.s. and s.p. matrix of $\underline{x}$ on t viz. $S_{xt}S_{tt}^{-1}S_{tx}$, the residual matrix

$$L = B - \underline{z}\,\underline{z}' \tag{3.1}$$

where

$$\underline{z} = S_{xt}S_{tt}^{-1/2} \tag{3.2}$$

has a central Wishart distribution with q - 1 d.f. The vector $\underline{z}$ has a p-variate normal distribution but $E(z) \neq 0$. In other words, the non-centrality of the distribution of B is entirely removed by regression on t . We have, thus, three independent distributions, of W , L , and $\underline{z}$. From the results of Section 2,

$$M = (L + W)^{-1/2}W(L + W)^{-1/2} \tag{3.3}$$

is independently distributed of L + W and has the density $B_p(M|n-q|q-1)$. Hence it is also independently distributed of

$$\underline{u} = (L + W)^{-1/2}\underline{z} \tag{3.4}$$

Now, make an orthogonal transformation from M to $G = [g_{ij}]$ by

$$G = UMU' \quad , \tag{3.5}$$

where U is an orthogonal matrix whose <u>first</u> row is $(\underline{u}'u)^{-1/2}\underline{u}'$. This is a random orthogonal transformation as $\underline{u}$ is random. Here, therefore, we employ Wijsman's [11] argument. As M and $\underline{u}$ are independent, the conditional distribution of M when $\underline{u}$ is fixed is $B_p(M|n-q|q-1)dM$. By the transformation (3.5), in this, we obtain the conditional distribution of G when $\underline{u}$ is fixed. As $|G| = |M|$ and $|I - G| = |I - M|$, the distribution of G is $B_p(G|n-q|q-1)dG$. But this conditional distribution of G does not involve $\underline{u}$ and so it is also the unconditional distribution of G and further G is independent of $\underline{u}$. A further transformation $G = KK'$ where $K = [k_{ij}]$, a lower triangular matrix, yields $k_{ii}^2 (i = 1, \cdots, p)$ which are independent and have $B_1(k_{ii}^2|n-q + 1-i|q-1)dk_{ii}^2$ as their distribution.

It now only remains to prove that $\Lambda'' = k_{11}^2$ and $\Lambda''' = \prod_{i=2}^{p} k_{ii}^2$, so that Λ'' has $\Lambda(n-1, 1, q-1)$ distribution and Λ''' has $\Lambda(n-2, p-1, q-1)$ distribution independent of Λ''. This can be seen from below:

From (3.1),

$$(3.6) \qquad S_{tt}^{-1/2}\underline{b} = T^{-1}\underline{z} = (L + W + \underline{z}\,\underline{z}')^{-1}\underline{z}$$

$$= \frac{(L + W)^{-1}\underline{z}}{1 + z'(L + W)^{-1}z}$$

and hence, from (1.2)

$$(3.7) \qquad \Lambda'' = \underline{z}'(L + W)^{-1}W(L + W)^{-1}\underline{z}/\underline{z}'(L + W)^{-1}\underline{z}$$

$$= \underline{u}'M\underline{u}/\underline{u}'\underline{u}$$

$$= \text{first element of } UMU' = G$$

$$= g_{11}$$

$$= k_{11}^2 \quad .$$

Similarly,

$$(3.8) \qquad \Lambda''' = \frac{|W|}{|W + L + \underline{z}\,\underline{z}'|} \cdot \frac{\underline{z}'(L + W)^{-1}\underline{z} \cdot \{1 + z'(L + W)^{-1}z\}}{\underline{z}'(L + W)^{-1}W(L + W)^{-1}\underline{z}}$$

$$= \frac{|W|}{|W + L| \cdot (1 + u'u)} \cdot \frac{u'u(1 + u'u)}{u'Mu}$$

$$= |M| \cdot \frac{1}{g_{11}}$$

$$= |G| \cdot \frac{1}{g_{11}}$$

$$= \prod_{2}^{p} k_{ii}^{2}$$

4. DISTRIBUTIONS OF Λ^{IV} AND Λ^{V}.

The procedure is exactly analogous to that in the previous section, except for the change, that for obtaining the distributions of Λ^{IV} and Λ^{V}, we employ an orthogonal matrix U, whose last row is $(\underline{u}'\underline{u})^{-1/2}\underline{u}'$. Hence,

$$(4.1) \qquad \Lambda^{IV} = \frac{|W|}{|W + L + zz'|} \cdot \frac{\underline{z}'W^{-1}\underline{z}}{z'(L + W)^{-1}z/\{1 + z'(L + W)^{-1}z\}}$$

$$= |M|\underline{u}'M^{-1}\underline{u}/\underline{u}'\underline{u}$$

$$= |G| \cdot u'U'G^{-1}Uu/u'u$$

$$= |G| \cdot \text{last element of } G^{-1}\text{, as } Uu = [0, 0, 0, \cdots, 1]' .$$

$$= k_{11}^{2}k_{22}^{2} \cdots k_{p-1,\,p-1}^{2}$$

Similarly

$$\Lambda^{V} = |G|/\Lambda^{IV} \qquad (4.2)$$

$$= k_{pp}^{2} \quad ,$$

and hence Λ^{IV} has $\Lambda(n-1 , p-1 , q-1)$ distribution and Λ^{V} has $\Lambda(n-p, 1 , q-1)$ distribution independent of Λ^{IV} .

5. GEOMETRICAL INTERPRETATION.

Wilks' Λ for the relationship between $\underline{x}$ and $\underline{y}$ is $|W|/|W + B|$. If we eliminate t , the 'residual' criterion is $|W|/|W + L|$ or from Section 3, it is $|M| = |G|$; and Λ'' , Λ''' or Λ^{IV} , Λ^{V} are factors of $|G|$. G is obtained from M by an orthogonal transformation; in other words, we rotate the axes corresponding to $x_1 , x_2 , \cdots , x_p$, so that one of the new axes corresponds to $\underline{u}'\underline{x}$. Observe that $\underline{u}'x = (L + W)^{-1/2}z'x$ corresponds to the sample projection of t , the hypothetical discriminator on the $\underline{x}$-space. Λ'' corresponds to the relationship of this variate with $\underline{y}$, eliminating t and Λ''' measures the relationship of other variables in the $\underline{x}$-space, after $\underline{u}'\underline{x}$ is eliminated. For Λ^{IV} and Λ^{V} , the last row of the orthogonal matrix U was taken to correspond to $\underline{u}'\underline{x}$ and so Λ^{IV} measures the relationship of variates orthogonal to $\underline{u}'\underline{x}$ with $\underline{y}$, eliminating t while Λ^{V} measures the relationship of $\underline{u}'\underline{x}$, when all these variables are eliminated first.

6. MORE THAN ONE HYPOTHETICAL DISCRIMINANT FUNCTION FROM THE y-SPACE.

The above argument and derivation can be easily extended to the case of goodness of fit of $s(s > 1)$, hypothetical discriminant functions from the dummy variables space. It is a straight forward extension, where B is split up as $\sum_{\alpha=1}^{s} \underline{z}_{\alpha}\underline{z}_{\alpha}' + L$ where $\sum_{\alpha=1}^{s} \underline{z}_{\alpha}\underline{z}_{\alpha}'$ is the matrix of regression s.s. and s.p. of $\underline{x}$ on the assigned s discriminant functions from the y-space and carries s d.f. L will have q-s d.f. and under the null hypothesis of adequacy of the proposed discriminant functions, L will have a central Wishart distribution. The residual Wilks' Λ criterion, when the proposed discriminant functions are eliminated will be again $|W|/|W + L| = |M|$ as before (but with q-1 replaced by q-s). We than make a transformation $G = UMU'$ where either the first s rows of U are $(L + W)^{-1/2}\underline{z}_\alpha$ $(\alpha = 1, 2, \cdots, s)$ and the remaining rows are so chosen that U is orthogonal [we can always choose or adjust $\underline{z}_\alpha$ to satisfy $z'_\alpha (L + W)^{-1} z_\beta = 0$, $\alpha \neq \beta$] or that the last s rows of U are $(L + W)^{-1/2} z_\alpha$ $(\alpha = 1, 2, \cdots, s)$ and the remaining rows are suitably chosen. The matrix G is then decomposed as KK' where K is lower triangular and we get two alternative factorizations viz.

$$(6.1) \qquad |G| = \left(\prod_{i=1}^{s} k_{ii}^2\right)\left(\prod_{i=s+1}^{p} k_{ii}^2\right) = \Lambda''\Lambda'''$$

or

$$(6.2) \qquad |G| = \left(\prod_{i=1}^{p-s} k_{ii}^2\right)\left(\prod_{i=p-s+1}^{p} k_{ii}^2\right) = \Lambda^{IV}\Lambda^{V}$$

depending on whether the first or the last s rows of U are $(L + W)^{-1/2} z_\alpha$. Λ'' is the direction factor and Λ''' is the partial collinearity factor. They are independently distributed as $\Lambda(n-s, s, q-s)$ and $\Lambda(n-2s, p-s, q-s)$ respectively. In the alternative factorization (6.2), Λ^{IV} is $\Lambda(n-s, p-s, q-s)$ and Λ^{V} is an independent $\Lambda(n-p, s, q-s)$.

REFERENCES

[1] Barnard, M. M. (1935). "The secular variation of skull characters in four series of Egyptian skulls," Ann. Eugen. Land, 6, 352-71.

[2] Bartlett, M. S. (1947). "Multivariate analysis," J. Roy. Statist. Soc., Suppl. 9, 176-90.

[3] Bartlett, M. S. (1951). "The goodness of fit of a single hypothetical discriminant function in the case of several groups," Ann. Eugen. Land, 16, 199-216.

[4] Fisher, R. A. (1936). "The use of multiple measurements in Taxonamic problems," Ann. Eugen. Land., 7, 179-95.

[5] Khatri, C. G. (1964). "Distribution of the 'generalised' multiple correlation matrix in the dual case," Ann. Math. Stat., 35, 1801-06.

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

[7] Kshirsagar, A. M. (1962). "A note on direction and collinearity factors in canonical analysis," Biometrika, 49, 255-59.

[8] Kshirsagar, A. M. (1964). "Distributions of the direction and collinearity factors in discriminant analysis," Proc. Camb. Phil. Soc., 60, 217-25.

[9] Kshirsagar, A. M. (1964). "Wilks' Λ," J. Ind. Stat. Ass., 2, 1-20.

[10] Rao, C. Radhakrishna (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley & Sons, Inc.

[11] Wijsman, Robert A. (1957). "Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis," Ann. Math. Stat., 28, 415-23.

[12] Williams, E. J. (1959). Regression Analysis. New York: John Wiley & Sons, Inc.