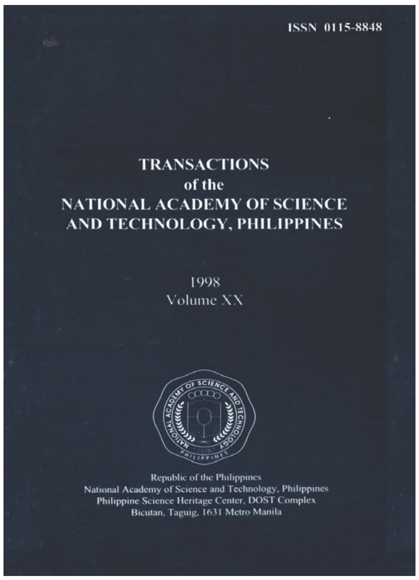
TRANSACTIONSNASTPHL

ISSN 0115-8848 (print)
ISSN 2815-2042 (online)

https://transactions.nast.ph/

Vol. 20 (1998)
doi.org/10.57043/transnastphl.1998.5837

Transactions NAST PHL, is the official journal of the National Academy of Science and Technology Philippines. It has traditionally published papers presented during the Academy's Annual Scientific Meeting since 1979 to promote science-based policy discussions of and recommendations on timely and relevant national issues as part of its functions as a national science academy. Starting in 2021, this journal has been open to contributions from the global scientific community in all fields of science and technology.



On Roots of Likelihood Functional Equations in Classical Univariate and Multivariate Statistical Theory

Tito A. Mijares

Member, National Academy of Science and Technology, Philippines

Citation
Mijares TA. 1998. On roots of likelihood functional equations in classical univariate and multivariate statistical theory. Transactions NAST PHL 20: 395-418. doi.org/10.57043/transnastphl.1998.5837

ON ROOTS OF LIKELIHOOD FUNCTIONAL EQUATIONS IN CLASSICAL UNIVARIATE AND MULTIVARIATE STATISTICAL THEORY

TITO A. MIJARES
Member, National Academy of Science and Technology, Philippines

ABSTRACT

Three possible multivariate analogues of the univariate root of equation $h - \lambda g = 0$ are $|H| - \lambda|G| = 0$, $|H - \lambda G| = 0$, and $a'Ha - a'Ga = 0$. The distribution of the roots in various cases such as in regression equations, likelihood ratio criterion, canonical correlations, and Fisher's linear discriminant function are examined. The general distribution of the roots can be expressed in multivariate beta-prime variates and tests, such as those involving Wilks Λ, are shown to be related to equivalent tests involving other functions (such as the sum) of the roots after transforming the general distribution to multivariate beta variates.

Key words: Sum of the roots test; Bartlett-Nanda-Pillai trace criterion; first elementary symmetric function of the roots, $V_1^{(s)}$, test; Lawley-Hotelling trace $U^{(s)}$ test; multivariate roots tests; largest root; Wilks L; likelihood ratio criterion; Hotelling T^2; distribution of the roots of determinantal equations

1. INTRODUCTION

Many inferential problems in univariate statistical analysis can be reduced often to requiring the sampling distribution of the root of an equation of the form

$$\mathbf{h} - \lambda \mathbf{g} = 0, \quad (1.1)$$

where h and g are mutually independent random variates with specified distributions and parameters. If g is a non-zero constant then the root has a distribution which is proportional to the distribution of h. In normal theory the root λ generally turns out to have a distribution that enables one to perform directly significance testing. If the distribution is unknown a monotonic transformation to a variate function in λ with known distribution may be undertaken to enable one to perform significance testing on the function in λ, and equivalently, on the root itself. For instance, the distribution of the root when h is a standard normal variate distributed independently of variate g that follows the distribution of the square root of a mean chi-square is that of Student-t with a certain degree of freedom; that is, $h \sim n(0,1)$ and $g \sim \sqrt{\frac{\chi^2}{\upsilon}} \Rightarrow \lambda \sim t_\upsilon$. From this $\lambda^2 \sim F(1, \upsilon)$, that is, an F-distribution with 1 and υ degrees of freedom. More generally, if $h^2 \sim \chi^2$ with v_1 d.f.and independently of $g^2 \sim \chi^2$ with v_2 d.f., then $\lambda^2 \sim F(v_1,v_2)$.

This paper examines the multivariate analogues of the root when extended to an equation in vector variates form $\mathbf{h}' = (h_1,h_2, \ldots, h_p)$ and $\mathbf{g}' = (g_1,g_2,\ldots,g_p)$. The root(s) of

$$\text{(1.2)} \qquad \mathbf{h}' - \lambda\mathbf{g}' = (0,0,\ldots,0)$$

are those of p univariate equations $h_i - \lambda\, g_i = 0$, $i = 1,2,\ldots,p$. If each h_i is a standard normal variate and g is a unit vector, then $\mathbf{h}' \sim N_p(\mathbf{0},\mathbf{I}_p)$, a standard multivariate normal distribution with mean at p-component row vector 0 and covariance identity matrix $\mathbf{I}_p$.

The paper also considers the linking of the roots of likelihood functional equations with those of the roots of determinantal equations. It serves to unify perspectives on analogues in univariate and multivariate statistical analyses and theory.

2. DISTRIBUTIONS OF THE MULTIVARIATE ANALOGUES OF THE ROOT IN VECTOR VARIATES EQUATION FORM

Consider the p-component row vector $\mathbf{h}'$. Define its respective mean and measure of variance by the expected value $E(\mathbf{h}') = (0,0,\ldots,0)$ and the $E(hh') = \Sigma = (\sigma_{ij})$, $i,j = 1,2,\ldots,p$. We call the symmetric matrix Σ the covariance matrix of vector h. There are two possible multivariate analogues of the univariate variance σ^2:

$$\text{(2.1)} \qquad \Sigma = (\sigma_{ij}),\ i,j = 1,2,\ldots,p$$

$$\text{(2.2)} \qquad \det \Sigma = |\Sigma| = |\sigma_{ij}|.$$

While the expression in (2.1) has $p(p+1)/2$ distinct variates that of (2.2) is equivalent to a single variate since it represents the determinant of covariance matrix Σ. Thus we can have three possible types of equations representing the multivariate analogues of the univariate equation; these are:

(2.3) $\quad |H| - \lambda|G| = 0,$

(2.4) $\quad |H - \lambda G| = 0,$

(2.5) $\quad a'(H - \lambda G)a = 0$

where $H = (h_{ij})$ and $G = (g_{ij})$ are symmetric positive semi-definite matrices and **a** is a vector variate linearizing matrix (*).

In the next few sections (Sections 4-7) we shall study the distributions of the root of equations of the form (2.3) and some of its associated problems in significance testing. The rest after Section 3 will be devoted to those of the form (2.4) and (2.5). In Section 3 we recall the fundamental distributions derived from the multivariate normal distribution that form the bases for many significance testing in multivariate statistical analysis and theory.

3. WISHART DISTRIBUTION, WILKS GENERALIZED VARIANCE AND HOTELLING T^2

Let x_i', $i = 1,2,\ldots,N$ be vector variate samples from a p-variate normal distribution $N_p(\mu',\Sigma)$ with mean vector variate m' and covariance matrix S. Let S be a sample estimate of Σ and given by

$$(3.1) \qquad S = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})'}{N-1} .$$

We call S the (p x p) sample covariance matrix. The sample mean vector $\bar{x}$ is distributed according to $N_p(\mu,\Sigma/N)$ and $\sqrt{N}(\bar{x}-\mu) \sim N_p(0,\Sigma)$. The distribution of A = (N-1)S is known to have the density

$$(3.2) \qquad \frac{|A|^{\left(\frac{n-p-1}{2}\right)}\exp\left(-\frac{tr\Sigma^{-1}A}{2}\right)}{2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}}|\Sigma|^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right)},$$

where $n = N - 1$ and $\Sigma^{-1} = (\sigma^{ij})$.The distribution of A is often referred to as the Wishart p-variate distribution $W_p(\Sigma, n)$ with covariance matrix S and degrees of freedom n where $n = N - 1$. Wishart (1928) first derived the distribution which he called the "generalized product moment" distribution in samples from a multivariate normal population.

On the other hand, $|S|$ is commonly called the *generalized variance* in current literature. The distribution of $|S|$ is the same as the distribution of $|A|/(N-1)^p$ and $\mathbf{x}_i$ is distributed independently of $\mathbf{x}_j$ ($i \neq j$) according to $N_p(0,\Sigma)$. Through studies of integral equations of certain types Wilks (1932) showed that the distribution of the generalized variance follows the distribution of

$$(3.3) \qquad \frac{|\Sigma|}{(N-1)^p} \times \prod_{i=1}^{p} (\chi_{N-i}{}^2)$$

that is, the constant factor times the product of p independent $\chi_i{}^2$ variates, with the *ith* variate having N-i degrees of freedom.

The role of the Wishart distribution in multivariate statistical analysis is analogous to the role played by the chi-square distribution in univariate statistical analysis. The Wishart distribution has a property which is the multivariate analogue of Cochran's theorem on the sum of independently distributed chi-squares. Let $\mathbf{A}_1$ be distributed according to $W_p(\Sigma, n_1)$ and independently of $\mathbf{A}_2$ which is $W_p(\Sigma, n_2)$. Then the sum $\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{A} \sim W_p(\Sigma, n)$ where $n = n_1 + n_2$.

In the univariate case when $x_1,\ldots,x_N$ are independent samples from a normal population with mean zero and variance σ^2, let $Q_1 = N\times$ square of the sample mean and Q_2 = the sample variance s^2. Then, in terms of the Wishart distribution, $Q_1 \sim W_1(\sigma^2,1)$ since $\sqrt{N}\bar{x} \sim n(0,1)$, a standard normal, and $Q_2 \sim W_1(\sigma^2, N-1)$ since $s^2 \sim$ chi-square with N-1 degrees of freedom. The equation $Q_1 - \lambda^2 Q_2 = 0$ is

$$(3.4) \qquad N\bar{x}^2 - \lambda^2 s^2 = 0$$

and the root λ^2 is known to be distributed as Student t^2, which is distributed according to an F- distribution with 1 and N-1 degrees of freedom; i.e., $\lambda^2 \sim F(1,N-1)$.

If $Y_1,\ldots,Y_N$ are N p-component sample vectors ($N \geq p$) drawn from a multivariate population with distribution $N_p(0,\Sigma)$ the mean vector $\bar{Y} \sim N_p(0, \Sigma/N)$ or $\sqrt{N}\bar{Y} \sim N_p(0, \Sigma)$.

Consider the quadratic forms

$$(3.5) \qquad Q^{(1)} = N\bar{Y}\bar{Y}',$$

$$(3.6) \qquad Q^{(2)} = \sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})' = \sum_{i=1}^{N} Y_i Y_i' - N\bar{Y}\bar{Y}'$$

The form $Q^{(1)} \sim W_p(\Sigma, p)$, since it is of rank p while $Q^{(2)} \sim W_p(\Sigma, N-p)$ since it is of rank N - p. The sum $Q_1+Q_2 = Q \sim W_p(\Sigma, N)$. [In general, if A_i ($i=1,\ldots,q$) are independently distributed according to $W_p(\Sigma, n_i)$ then the sum $A_1 + \ldots + A_q = A \sim W_p(\Sigma, n_1+\ldots+n_q)$.]

We now have an equation of the matric form

$$(3.7) \qquad Q^{(1)} - \lambda^2 Q^{(2)} = O$$

$$Q^{(1)} Q^{(2)-1} - \lambda^2 (I_p) = O$$

The trace of the root in (3.7) then becomes

$$(3.8) \qquad \lambda^2 tr(I_p) = (N-1)tr[(\bar{Y}\bar{Y})S^{-1}]$$

$$p\lambda^2 = (N-1)\bar{Y}S^{-1}\bar{Y},$$

where $(\bar{Y}\bar{Y}) = Q^{(1)}$ and $(N-1)S^{-1} = Q^{(2)-1}$. Let $\bar{Y}S^{-1}\bar{Y} = T^2$. Then $\lambda^2 = [(N-1)/p]$ T^2. The distribution of the root l^2 is proportional to the distribution of Hotelling T^2 which, in turn, is proportional to an F- distribution with p and N-p degrees of freedom. That is,

$$(3.9) \qquad T^2/(N-1) \sim [p/(N-p)]\, F(p,N-p).$$

The distribution of T^2 was derived by Hotelling (1931). Like the Student-t, it has the property of invariance under scale transformation when the mean vector $\mu' = 0$

4. THE ROOT OF A LIKELIHOOD FUNCTIONAL EQUATION

The root could also represent a comparison of the likelihood function under two situations. Given the probability density function $f(x;\Omega)$ of random variable X in some parameter space Ω the likelihood of a sample of size N from this population is defined by $L=\Pi_i f(x_i;\Omega)$. Often we wish to examine the behavior of the likelihood under subspace ω of Ω by comparing maximum $L_{max}(\omega)$ with the maximum $L_{max}(\Omega)$. The root of equation is

$$(4.1) \qquad L_{max}(\omega) - \lambda L_{max}(\Omega) = 0,$$

$$\lambda = L_{max}(\omega) / L_{max}(\Omega).$$

The (2/N)th power of root λ is often known as the likelihood ratio criterion (LRC). In samples of size N from a normal distribution with unknown mean μ and unknown variance σ^2, the root can be shown to be

$$\lambda = \left[\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{\sum_{i=1}^{N} x_i^2} \right]^{\frac{N}{2}} \tag{4.2}$$

Since $\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 + N\bar{x}^2$, the LRC can be written

$$\lambda^{\frac{2}{N}} = \frac{1}{\left\{ 1 + \left[\frac{N\bar{x}^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right] \right\}} \tag{4.3}$$

or $\lambda^{2/N} = \{1+t^2/(N-1)\}^{-1}$, where $t^2=N\text{'}x^2/[(N-1)s^2]$ and $s^2 = \Sigma_i (x_i-\text{'}x)^2/(N-1)$.

The multivariate analogue of the likelihood function in samples of size N from $N_p(\mu, \Sigma)$ is given by

$$L(\mu,\Sigma^{-1}) = \frac{|\Sigma^{-1}|^{\frac{N}{2}}}{(2\pi)^{\frac{pN}{2}}} \exp\left[-\frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right]. \tag{4.4}$$

t can be shown that under subspace $\omega = (\mu_0,\Sigma^{-1})$ of $\Omega = (\mu,\Sigma^{-1})$ the LRC is the (2/N)th) power of root

$$\lambda = \frac{|\Sigma_\Omega|^{\frac{N}{2}}}{|\Sigma_\omega|^{\frac{N}{2}}} = \frac{\left| \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})' \right|^{\frac{N}{2}}}{\left| \sum_{i=1}^{N} (x_i - \mu_0)(x_i - \mu_0)' \right|^{\frac{N}{2}}} \tag{4.5}$$

$$= \frac{|A|^{\frac{N}{2}}}{\left| A + N(\bar{x} - \mu_0)(\bar{x} - \mu_0)' \right|^{\frac{N}{2}}}$$

where Σ_ω and Σ_Ω are the maximum likelihood estimates of Σ under spaces ω and Ω, respectively; $A = \sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})' = (N-1)S$, and S = the sample covariance matrix.The third relation of (4.5) shows $\lambda \to 0$ the greater the departure of the sample mean vector $\bar{\mathbf{x}}$ from the population mean vector μ_0 and $\lambda \to 1$ as $\bar{\mathbf{x}} \to \mu_0$. Relation (4.5) can be further reduced to

$$(4.6) \qquad \lambda^{\frac{2}{N}} = \frac{1}{1 + N(\bar{x} - \mu_0)' A^{-1}(\bar{x} - \mu_0)} = \frac{1}{1 + T^2/(N-1)}$$

where $T^2 \sim N(\bar{\mathbf{x}} - \mu_0)'S^{-1}(\bar{\mathbf{x}} - \mu_0) = (N-1)N(\bar{\mathbf{x}} - \mu_0)'A^{-1}(\bar{\mathbf{x}} - \mu_0)$. By (3.9) we have

$$(4.7) \qquad \lambda^{\frac{2}{N}} = \frac{1}{1 + \frac{p}{N-p}F(p, N-p)}$$

5. THE ROOTS OF REGRESSION EQUATIONS

In simple regression we regress a scalar deviate y from its population mean on a fixed deviate x by an equation of the form

$$(5.1) \qquad y_i - \beta x_i = e_i, \quad i=1,2,\ldots,N$$

As deviates we assume $\Sigma_i y_i = \Sigma_i x_i = \Sigma_i e_i = 0$. We construct the moment equations on the set of **N** points (x_i, y_i) by multiplying both sides of the equation by x_i and adding the results. Thus

$$(5.2) \qquad \sum_{i=1}^{N} x_i y_i - \beta \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i e_i = 0$$

since x_i is fixed so that $\Sigma x_i e_i = \Sigma x_i \times \Sigma e_i = 0$. The root of (5.2) is

$$(5.3) \qquad \hat{\beta} = \left(\sum_{i=1}^{N} x_i^2\right)^{-1}\left(\sum_{i=1}^{N} x_i y_i\right)$$

In the multiple regression model we have a vector **y** of **N** components regressed on known matrix **x** (N×p) weighted by unknown p-component vector β and error vector **e** of **N** components. Thus,

$$(5.4) \qquad \mathbf{y} - \mathbf{x}\beta = \mathbf{e}$$

The moment equations, noting the usual assumptions, become

$$(5.5) \qquad x'y - x'x\beta = 0,$$

so that the root is

$$(5.6) \qquad \beta_{reg} = (x'x)^{-1}x'y$$

The analogues in multivariate regression model are

$$(5.7) \qquad \mathbf{Y} - \mathbf{XB} = \mathbf{U},$$

or

$$(5.8) \qquad (\mathbf{y}_1\ \mathbf{y}_2 \ldots \mathbf{y}_p) - (\mathbf{X}\beta_1\ \mathbf{X}\beta_2 \ldots \mathbf{X}\beta_p) = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_p),$$

where Y is (N×p), X is (N×q), B is (q×p) and U is (N×p). Pre-multiplying both sides by X′ and solving for B, we have the root as the regression estimate

$$(5.9) \qquad \mathbf{B}_{reg} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

which is an analogue of (5.3) and (5.6).

6. THE MAXIMUM LIKELIHOOD ESTIMATES OF PARAMETERS AND THE LIKELIHOOD RATIO CRITERION IN MULTIVARIATE LINEAR REGRESSION

Consider the model defined by (5.7) or (5.8). In most applications **Y** (N×p) is an observed matrix of p response variables on each of N individuals, **X** (N×q) is a known matrix, **B** (q×p) is a matrix of unknown regression parameters (or roots of the moment equations) and **U** is an error matrix coming from a multivariate distribution $N_p(\mathbf{0},\Sigma)$ and unknown covariance matrix Σ independent of **X**. When **X** represents q independent variables observed on each of N individuals (5.7) is called a *multivariate regression model.* When **X** is a design matrix (consisting of 0s and 1s), (5.7) is called a *general linear model.* The columns of **Y** represent the dependent variables which are to be explained by the q independent variables of **X**.

Let y_i' be the ith row vector of matrix Y. Assume that N ≥ p+q and the rank of **X** is q. It can be shown that the maximum likelihood estimate of **B** (which is mathematically equivalent to $\mathbf{B}_{reg}$) is given by

(6.2) $$\mathbf{B}_{\text{mle}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and that for Σ is

(6.2) $$N\Sigma_{mle} = \sum_{i=1}^{N} y_i y_i' + B_{mle}' A B_{mle}$$

where $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}$. It can also be shown that

(a) $\mathbf{B}_{\text{mle}}$ is unbiased for $\mathbf{B}$ and is multivariate normal,

(b)$\mathrm{N}\Sigma_{\text{mle}} \sim \mathrm{W}_{\mathrm{p}}(\Sigma, \mathrm{N\text{-}q})$, and

(c) $\mathbf{B}'_{\text{mle}}\mathbf{A}\mathbf{B}_{\text{mle}} \sim \mathrm{W}_{\mathrm{p}}(\Sigma, \mathrm{q})$ and independent of $\mathrm{N}\Sigma_{\text{mle}}$.

The likelihood function of the $\mathbf{N}$ multivariate normal observations is

(6.3) $$L = \frac{1}{(2\pi)^{\frac{Np}{2}} |\Sigma|^{\frac{N}{2}}} \exp[-\frac{1}{2}(y - B'x)' \Sigma^{-1}(y - B'x)]$$

Suppose we partition $\mathbf{B}' = (\mathbf{B}'_1\ \mathbf{B}'_2)$, where $\mathbf{B}'_1$is $(p \times q_1)$ and $\mathbf{B}'_2$ is $(p \times q_2)$ and $q = q_1 + q_2$, $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})$, and $\Sigma^{-1} = \mathbf{A} = (\mathbf{A}_{ij})$ such that $\mathbf{A}_{11}$ is $(q_1 \times q_1)$, $\mathbf{A}_{12} = \mathbf{A}'_{21}$ is $(q_1 \times q_2)$, and $\mathbf{A}_{22}$ is $(q_2 \times q_2)$. Then the maximum likelihood estimate of Σ over space ω (that is, under the hypothesis $\mathbf{B}_1 = \mathbf{B}_1^*$), can be shown (Anderson,1984; p.293) to be

(6.4) $$N\Sigma_\omega = \sum_{i=1}^{N} \left(y_i' - B_1^{*\prime} x_i^{(1)}\right)\left(y_i' - B_1^{*\prime} x_i^{(1)}\right)' - B_{2\omega}' A_{22} B_{2\omega}$$

Thus the maximum likelihood function over space ω is

(6.5) $$\max \mathrm{L}\omega(\mathbf{B}_2, \Sigma) = (2\pi)^{-(1/2)pN} \mid \Sigma\omega \mid^{-N/2} \exp\{(-1/2)\, pN\}$$

The maximum likelihood of (6.3) over space Ω is

(6.6) $$\max \mathrm{L}\Omega(\mathbf{B}, \Sigma) = (2\pi)^{-(1/2)pN} \mid \Sigma\Omega \mid^{-N/2} \exp\{(-1/2)\, pN\}$$

Thus the root of the equation for testing the hypothesis $\mathbf{B}_1 = \mathbf{B}^*_1$ (that is, the hypothesis about the first q_1 rows of the coefficients) is given by

(6.7) $$\max \mathrm{L}\omega\, (\mathbf{B}_2, \Sigma) - \lambda \max \mathrm{L}\Omega(\mathbf{B}, \Sigma) = 0$$

or

$$\text{(6.8)} \qquad \lambda^{\frac{2}{N}} = \frac{\left|\hat{\Sigma}_{\Omega}\right|}{\left|\hat{\Sigma}_{\omega}\right|}.$$

The denominator in (6.8) has q_1 degrees of freedom while the numerator has ($N - q_1 - q_2$) degrees of freedom. Let $\lambda^{2/N} = U(p,q_1,N\text{-}q_1\text{-}q_2)$. Then the **LRC** for testing the hypothesis $\mathbf{B}_1=\mathbf{B}^*_1$, where $\mathbf{B}_1^*$ is specified, in multivariate regression has distribution $U(p,q_1,N\text{-}q_1\text{-}q_2)$. This is known as the **LRC** for testing $\mathbf{B}_1=\mathbf{B}^*_1$ in regression equations.

7. THE LIKELIHOOD RATIO CRITERION, U(p,m,n)

The likelihood ratio criterion is the root of a determinantal equation of the form

$$\text{(7.1)} \qquad |\mathbf{W}_1| - \lambda\,|\mathbf{W}_2| = 0,$$

where $|\mathbf{W}_i|$ are the determinants of random symmetric matrices following the Wishart distribution of (3.2). Wilks (1932) first showed that the determinant of the sample covariance matrix S, which he termed *generalized* variance, follows the distribution that is proportional to the distribution of products of independent chi-squares given by (3.3). The root $\lambda^{2/N}$ of the likelihood functional equation in (6.8) is a form of **LRC** U(p,m,n). The parameter *n* of the criterion usually represents the "error" degrees of freedom and *m* the "hypothesis" degrees of freedom. Thus *m* + *n* represents the "total" degrees of freedom. Relation (4.6) shows LRC as a function of Hotelling T^2 and relation (4.7) as a function of an F- distribution.

The following are some features of the **LRC**:

(a) The distribution of U(p,m,n) is the distribution of $\prod_{i=1}^{p}\{X_i\}$, where the $\mathbf{X}_i$ are independent, each having the beta density $\beta\left[x;\frac{1}{2}(n+1-i),\frac{1}{2}m\right]$

(b) In the case when q_1 is greater than p and the hypothesis about the first q_1 of the regression coefficients is true, the distribution of $U(p,q_1,N\text{-}q_1\text{-}q_2)$ is the same as that of $U(q_1,p,N\text{-}p\text{-}q_2)$; that is, interchange the role of p and q_1.

Particular cases of (a) are (Anderson, 1958; 195-196).

Case 1. p=1. The density of U(1,m,n) is

$$\text{(7.2)} \qquad f(u) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} u^{\frac{n}{2}-1}(1-u)^{\frac{m}{2}-1}$$

$$= \beta[u;(n/2),(m/2)]$$

Its relation to the F-distribution is

$$U(1,m,n) = \frac{1}{1+\frac{m}{n}F(m,n)} \cdot \tag{7.3}$$

[See, for example, relation (4.6)] Thus, the LRC when p = 1 and the F-distribution are related by

$$\frac{1-U(1,m,n)}{U(1,m,n)} \bullet \frac{n}{m} = F(m,n). \tag{7.4}$$

Case p=2: The distribution of

$$\frac{1-\sqrt{U(2,m,n)}}{\sqrt{U2,m,n}} \bullet \frac{n-1}{m} = F[2m,2(n-1)] \tag{7.5}$$

Case any p and m=2: The distribution of

$$\frac{1-\sqrt{U(p,2,n)}}{\sqrt{U(p,2,n)}} \times \frac{n+1-p}{p} = F[2p, 2(n+1-p)] \tag{7.6}$$

Remarks

1. Understanding what the root of the functional equation does in statistical analysis can help unify perspectives in inference problems. Estimation and hypothesis testing often require knowledge of the probability distribution of the root(s). The distribution of the root of h - λ = 0 (g=1) in sampling from a univariate standard normal population (h) is known to follow a normal distribution and its square (i.e., $h^2 - \lambda^2 = 0$) follows the chi-square distribution with degrees of freedom equal to the sample size. The distribution of the root in sampling from a normal population with mean μ and standard deviation σ, where h is the standardized mean variate, that is, $\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu)$, and independent of g variate $\sqrt{\frac{ns^2}{\sigma^2}}$, where s^2 is the sample variance, follows the Student-t distribution with n-1 degrees of freedom. The square of this root follows the F-distribution with 1 and n-1 degrees of freedom.

2. Calculating the desired probabilities from the distribution of the root is not usually an easy straightforward task. To have practical use, tables of probabilities (particularly, on the significance points), such as those for the standard normal, chi-square, Student-t, and the F distributions, are usually available and are easily accessible in statistical textbooks. However, the distribution of the root of equation

of the form such as that of the **LRC** U(p,m,n) does not usually take simple forms and computing the desired probabilities can be a challenging if not a formidable task (e.g. see Anderson,1958; pp. 195-202). Accordingly, approximate distributions have been utilized to obtain the desired significance points. One such approximation study, for example, showed (Schatzoff, 1966) that the limiting distribution of

$$(7.7) \qquad -[n-(1/2)(p-m+1)]\log U(p,m,n) \sim \chi^2_{pm}$$

The significance point of U(p,m,n) at level α is approximated by a chi-square with *pm* degrees of freedom; that is,

$$(7.8) \qquad -[n-(1/2)(p-m+1] \log U_{p,m,n}(\alpha) \approx \chi^2_{pm}(\alpha),$$

thus, $\Pr[U(p,m,n) \leq \mathbf{u}_{p,m,n}(\alpha)] = \Pr[\chi^2_{pm} \geq \chi^2_{pm}(\alpha)] = \alpha$. Other approximations have been given by Rao(1951) [e.g., Anderson, 1984; p.321] as an **F**-approximation and Box (1949).

8. THE ROOTS OF THE DETERMINANTAL EQUATION $|\mathbf{H} - \lambda\, \mathbf{G}| = 0$

We have seen in the preceding section how the **LRC** comes out as the (2/N)th power of the root of the likelihood functional equation of the forın $|\mathbf{H}| - \lambda|\mathbf{G}| = 0$, the distribution of which follows that of **U**(p,m,n). The symmetric positive definite matrices H and G are distributed according to $W_p(\Sigma, m)$ and $W_p(\Sigma,n)$, respectively. Wilks criterion is a root following the distribution of the product of independent univariate betas [7(a)]. The roots of equation

$$(8.1) \qquad |\mathbf{H} - \lambda\, \mathbf{G}| = 0,$$

however, are the *characteristic roots of* **H** *in the metric of* **G**. These roots possess certain invariant properties under nonsingular transformations which simplify some multivariate distribution problems to those not dependent on parameters. For instance, let C be a nonsingular transformation such that

$$(8.2) \qquad \mathbf{C\Sigma C'} = \mathbf{I}$$

Let

$$(8.3) \qquad \mathbf{H^*} = \mathbf{CHC'}, \quad \mathbf{G^*} = \mathbf{CGC'}$$

Then **H*** and **G*** will be distributed according to Wishart $\mathbf{W}_p(\mathbf{I}_p,m)$ and $\mathbf{W}_p(\mathbf{I}_p,n)$, respectively. The roots of equation of the transformed matrices

$$(8.4) \qquad \begin{aligned} |\mathbf{H^*} - \lambda\, \mathbf{G^*}| &= |\mathbf{CHC'} - \lambda\mathbf{CGC'}| = 0, \\ &= |\mathbf{C(H} - \lambda\mathbf{G)C'}| = 0, \\ &= |\mathbf{H} - \lambda\mathbf{G}| = 0 \end{aligned}$$

are the same as the roots of (8.1).

A more appropriate multivariate analogue of the univariate case for testing the homogeneity of two variances is the null hypothesis that the two p-variate population covariance matrices, Σ_1 and Σ_2, are equal. Using the sample equivalents of the population covariance matrices, Roy (1939) derived the sampling distribution of the nonzero roots of

$$(8.5) \qquad |\mathbf{S}_1 - \lambda\mathbf{S}_2| = 0$$

where $\mathbf{S}_i$ is obtained from independent samples of sizes n_i (> p) from a p-variate normal population $N_p(\mu_i,\Sigma_i)$. The joint density of the roots $0 \le \lambda_p \le \lambda_{p-1} \le \ldots \le \lambda_1$ is of a multivariate Beta-prime type [see (11.2), for instance] and is given (Anderson, 1984; p.530) by

$$(8.6) \qquad K\prod_{i=1}^{p}\frac{\lambda_i^{\frac{m-p-1}{2}}}{(1+\lambda_i)^{\frac{m+n}{2}}}\prod_{i<j}(\lambda_i - \lambda_j)$$

where $m = n_1 - 1$, $n = n_2 - 1$ and

$$(8.7) \qquad K = \frac{\pi^{\frac{p}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{m+n+1-i}{2}\right)}{\prod_{i=1}^{p}\left\{\Gamma\left(\frac{m+1-i}{2}\right)\Gamma\left(\frac{n+1-i}{2}\right)\Gamma\left(\frac{p+1-i}{2}\right)\right\}}$$

A related form of relation (8.1) is the determinantal equation of the form

$$(8.8) \qquad |H - \theta(H+G)| = 0.$$

The roots θ_i of **H** in the metric of (**H** + **G**) are related to λ_i by the relation

$$(8.9) \qquad \theta_i = \frac{\lambda_i}{1+\lambda_i}$$

or

$$(8.10) \qquad \lambda_i = \frac{\theta_i}{1-\theta_i}.$$

Accordingly, the joint density for $0 \le \theta_p \le \theta_{p-1} \le \ldots \le \theta_1 \le 1$ can be deduced from (8.6) using transformation (8.10).

At about the same time and independently of.Roy, Hsu (1939) also obtained the joint density of the roots of equation (8.8) for two cases.

Case I. $n_1 \ge p$ and $n_2 > p$

$$(8.11)\ K_1 \times \left\{\prod_{i=1}^{p} \theta_i\right\}^{(n_1-p-1)/2} \left\{\prod_{i=1}^{p} (1-\theta_i)\right\}^{(n_2-p-1)/2} \prod_{i<j}^{p} (\theta_i - \theta_j),$$

Case 2. $n_1 \le p$ and $n_2 > p$

$$(8.12)\quad K_2 \times \left\{\prod_{i=1}^{n_1} \theta_i\right\}^{(p-n_1-1)/2} \left\{\prod_{i=1}^{n_1} (1-\theta_i)\right\}^{(n_2-p-1)/2} \prod_{i<j}^{p} (\theta_i - \theta_j)$$

He gave explicit values of constants K_1 and K_2. He also proved a theorem leading to the distribution of the canonical correlations below.

9. CANONICAL CORRELATIONS

An investigator, for example, may want to investigate the possible relation between personality variables on the one hand, and academic achievement on the other, among high school students. By using canonical analysis he/she may be able to identify what sort of personality profile, such as determining appropriate linear combination of personality scales, that are most highly correlated with linear combinations of achievement tests.

The theory of canonical correlations is due to Hotelling (1936). Suppose, the random vector **X** of p+q components with mean vector **0** has the covariance matrix Σ. We partition $X' = \left(X^{(1)}\ X^{(2)'}\right)$ into two subvectors of p and q components respectively and its covariance matrix conformably to

$$(9.1)\qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Consider the linear combinations $U=\alpha'\mathbf{X}^{(1)}$ and $V=\gamma'\mathbf{X}^{(2)}$. We want to find vectors α and γ such that the correlation between the two linear functions U and V is a maximum. We require α and γ such that U and V have unit variances, that is

$$(9.2)\qquad E(U^2) = E(\alpha' X^{(1)} X^{(1)'} \alpha) = \alpha' \Sigma_{11} \alpha = 1$$

$$(9.3)\qquad E(V^2) = E(\gamma' X^{(2)} X^{(2)'} \gamma) = \gamma' \Sigma_{22} \gamma = 1$$

The correlation between U and V is

$$(9.4)\qquad E(UV) = E(\alpha' X^{(1)} X^{(2)'} \gamma) = \alpha' \Sigma_{12} \gamma.$$

Using Lagrange multipliers we maximize (9.4) subject to (9.2) and (9.3). Thus differentiating

$$(9.5)\qquad \varphi = \alpha' \Sigma_{12}\gamma - (1/2)\lambda(\alpha' \Sigma_{11}\alpha - 1) - (1/2)\mu(\gamma' \Sigma_{22}\gamma - 1)$$

separately with respect to α and γ and equating to **0** the resulting system of linear functions we arrive at a matrix equation

$$(9.6)\qquad \begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix}\begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = 0.$$

In order that (9.6) have a nontrivial solution the determinant of the matrix of coefficients should vanish, thus

$$(9.7)\qquad \begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \lambda\Sigma_{22} \end{vmatrix} = 0$$

which is a polynomial equation of degree (and of number of nonzero roots) p+q. Relation (9.7) can be further reduced algebraically (Kendall, p.63) to

$$(9.8)\qquad (-\lambda)^{q-p}\left|\lambda^2\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right| = 0$$

which is now only a pxp determinant. It will have p nonzero roots in λ^2. Putting $\lambda^2 = \theta$ and, with some rearrangement, (9.8) can be transformed to

$$(9.9)\qquad \left| \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1} - \theta I_p \right| = 0.$$

If the Σ_{ij} 's are replaced by their sample equivalents $\mathbf{S}_{ij}$'s the joint distribution of the roots $1 \geq \theta_1 \geq \theta_2 \geq \ldots \geq \theta_p \geq 0$ of (9.9) has also been derived by.Hsu (1939) as a generalization of Hotelling's result for p=q=2. If the $(x_1, x_2, \ldots, x_p)$ and $(x_{p+1}, x_{p+2}, \ldots, x_{p+q})$ are two mutually independent sets of variates, $p \leq q$, such that the first set is normally distributed, the joint density of the canonical correlations is

$$(9.10)\qquad K \cdot \left\{\prod_{i=1}^{p}\theta_i\right\}^{(q-p-1)/2}\left\{\prod_{i=1}^{p}(1-\theta_i)\right\}^{(n-p-q-2)/2}\prod_{i<j}^{p}(\theta_i - \theta_j)$$

where

$$(9.11)\qquad K = \frac{\pi^{p/2}\prod_{i=1}^{p}\Gamma(\frac{n-i}{2})}{\prod_{i=1}^{p}\left\{\Gamma(\frac{n-q-i}{2})\Gamma(\frac{p-i+1}{2})\Gamma(\frac{q-i+1}{2})\right\}}$$

which is of the same form as that derived by Roy (1939).

10. FISHER'S LINEAR DISCRIMINANT FUNCTION

The third type of analogue is the root of an equation in quadratic form

$$10.1 \qquad \mathbf{a'Ha} - \lambda \mathbf{a'Ga} = 0$$

given by (2.5). When two or more populations are measured in several (say, p) characters there is special interest on certain linear functions of these characters called by Fisher (1936) as *discriminant functions*. No particular parametric form for the distribution of the populations $\Pi_1, \ldots, \Pi_g$ need be assumed other than merely to look for a rule that best discriminates between them. We look for a linear function $\mathbf{a'x}$ that *maximizes* the ratio of the g between-groups sum of squares to their within-group sum of squares. Let vector

$$(10.1) \qquad y = Xa = \begin{pmatrix} X_1 a \\ \bullet \\ \bullet \\ X_g a \end{pmatrix} = \begin{pmatrix} y_1 \\ \bullet \\ \bullet \\ y_g \end{pmatrix}$$

be a linear combination of the columns of data matrix $\mathbf{X}(n \times p)$ whose n rows are divided into g groups. The total sum of squares of vector y is $\mathbf{y'H_n y} = \mathbf{a'X'H_n Xa}$, where $\mathbf{H_n}$ is the centring matrix; that is, $\mathbf{H_n = I_n - 11'}/n$. If $\bar{y}_i$ is the mean of the *ith* subvector y_i and $\bar{y}$ is the overall mean of the g groups, the sum of squares of the g between-groups is

$$(10.2) \qquad \Sigma_i n_i(\bar{y}_i - \bar{y})^2 = \Sigma_i n_i \{ \mathbf{a'}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\mathbf{a}\}^2 = \mathbf{a'Ba}$$

while the sum of squares of the within-groups is

$$(10.3) \qquad \Sigma_i \mathbf{y'_i H_i y_i} = \Sigma_i \mathbf{a'X_i'H_i X_i a} = \mathbf{a'Wa}.$$

This gives the total sum of squares of the g groups as

$$(10.4) \qquad \mathbf{y'Hy} = \mathbf{a'X'HXa} = \mathbf{a'Ta}.$$

The root of equation

$$(10.5) \qquad \mathbf{a'Ba} - \lambda \mathbf{a'Wa} = 0$$

is the ratio

$$(10.6) \qquad \lambda = \mathbf{a'Ba} \,/\, \mathbf{a'Wa}$$

that is intuitively appealing. Matrices B and W are both (p×p). The ratio is a measure of how far apart the groups are. If a is the vector that maximizes (10.6) [that is, $\lambda = \lambda_{max}$], the linear function a'x is called *Fisher's linear discriminant function* [In the next section, we will see that it is actually the first canonical variate or the eigenvector of matrix BW^{-1} corresponding to the largest root λ_{max}.]

An observation **x** is allocated to population Π_j whose mean score $a'\bar{x}_j$ is the closest to $a'\bar{x}$; that is, we allocate x to Π_j if

$$(10.7) \qquad |\mathbf{a'x} - \mathbf{a'}\bar{\mathbf{x}}_j| < |\mathbf{a'x} - \mathbf{a'}\bar{\mathbf{x}}_i|, \text{ for all } i \neq j.$$

11. THE MULTIVARIATE BETA VARIATES

In the Pearson system of curves there are two types of Beta variates. Their densities are

$$(11.1) \qquad f(x) = x^{m-1}(1-x)^{n-1}/\mathbf{B}(m,n), \quad 0 \leq x \leq 1$$

$$(11.2) \qquad f(y) = y^{m-1}(1+y)^{-m-n}/\mathbf{B}(m,n), \quad 0 \leq y < \infty$$

The density (11.1) is of Type I and is called a beta variate $\beta(m,n)$. It is distinguished from the case of (11.2) which is of Type VI and called a Beta-prime variate b¢(m,n). Type VI is transformed to Type I by using transformation $x = y(1+y)^{-1}$.

In univariate analysis many tests are based on statistics having independent chi-square distributions. If $h \sim \sigma^2\chi_{v1}^2$ and $g \sim \sigma^2\chi_{v2}^2$ are two independent chi-square variates, then the root λ of equation $\sigma^2\chi_{v1}^2 - \lambda\sigma^2\chi_{v2}^2 = 0$ is distributed as a $\beta'(v_1/2, v_2/2)$ variate. For instance, let $F = (N_1s_1^2/n_1)/(N_2s_2^2/n_2)$ where s_i^2 are sample variances and N_i are the sizes of samples from normal populations with unknown means and variances, then $(n_1/n_2)F$ is the ratio of two χ^2-variates with n_1 and n_2 degrees of freedom, respectively. The density of the F-variate is of the Beta-prime form given by

$$(11.3) \qquad f(F) = K_1 F^{\frac{n_1}{2}-1}\left(1 + \frac{n_1 F}{n_2}\right)^{-\frac{n_1+n_2}{2}}, \quad 0 \leq F < \infty$$

where

$$(11.4) \qquad K_1 = \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} / B\left(\frac{n_1}{2}, \frac{n_2}{2}\right).$$

The multivariate analogues of (11.1) and (11.2) are (8.6) and (8.11), or (8.12), respectively. The joint distribution of the roots in (8.6) transforms to the joint distribution of the roots in (8.11) or (8.12) by using transformation $\theta_i = \lambda_i(1+\lambda_i)^{-1}$. Noting the similarities in the forms of the joint distributions of roots in the λ's and in the θ's, Pillai (1954;1955) combined these into the general joint distribution of the non-zero roots in the θ's [in increasing order and given by (11.5) below] for testing three cases of multivariate hypotheses by defining for each case the appropriate set of parameters ; namely, (I). that of equality of the dispersion matrices of two p-variate normal populations;

(II). that of equality of the p-dimensional mean vectors for a k p-variate normal population (which is mathematically identical with the general problem of multivariate analysis of variance of means); and

(III) that of independence between a p-set and a q-set of variates in a $(p+q)$ - variate normal population, with $p \leq q$.

The joint density of the θ's in multivariate Beta variates is given by

$$(11.5) \qquad f(\theta_1, \ldots, \theta_s) = C(s,m,n) \prod_{i=1}^{s} \theta^{m}{}_{i} (1-\theta_i)^{r} \prod_{i>j} (\theta_i - \theta_j),$$
$$0 < \theta_1 \leq \ldots \leq \theta_s < 1.$$

where

$$(11.6) \qquad C(s,m,n) = \frac{\pi^{s/2} \prod_{i=1}^{s} \left[\Gamma\left[\left(\frac{1}{2}\right)(2m+2n+s+i+2) \right] \right]}{\prod_{i=1}^{s} \Gamma\left[\left(\frac{1}{2}\right)(2m+i+1) \right] \Gamma\left[\left(\frac{1}{2}\right)(2n+i+1) \right] \Gamma\left(\frac{i}{2}\right)}$$

The joint density given by (11.5) is now known to have been independently derived by Fisher (1939), Hsu (1939), Girshick (1939) for the case of s=2, Mood (1951), and Roy (1939). See also Mijares (1964a) and Bose (1977).

The sets of parameters of (11.5) for the various cases are defined by:

Case I : $Q = 2m = n_1 - p - 2$
$R = 2n = n_2 - p - 2$
n_1 and n_2 are the sample sizes from p-variate populations.

Case II: $Q = 2m = |\ k - p - 1\ | - 1$
$R = 2n = N - k - p - 1$
N = total size of the k samples from the p-variate population.

Case III. $Q = 2m = q - p - 1$
$R = 2n = k - p - q - 2$
$p + q < k$, the total size of samples from the $(p + q)$-variate population.

Mijares (1964a) includes s = 1 for the univariate and Hotelling T^2 cases. Accordingly, the degrees of freedom associated with the univariate beta parameters are equivalent to $v_2 = Q + 2 = 2m + 2$, $v_1 = R + 2 = 2n + 2$.

Remarks. There are meaningful relations among the roots λ and θ with the chi-square and F variates and their associated degrees of freedom v_1 and v_2. Denote F as the ratio of two mean chi-square variates; that is, $F = (\chi_1^2/v_1) \div (\chi_2^2/v_2)$. Then $(v_1/v_2)F = (\chi_1^2/\chi_2^2) = \lambda$ is the ratio of two chi-square variates and

$$\frac{\chi_1^2}{\chi_1^2+\chi_2^2} + \frac{(\chi_1^2/\chi_2^2)}{(\chi_1^2/\chi_2^2)+1} = \frac{\frac{v_1}{v_2}F}{\left(1+\frac{v_1}{v_2}F\right)}$$

$$(11.7) \qquad = \frac{\lambda}{1+\lambda} = \theta$$

The density of the βeta variate θ is

$$(11.8) \qquad f(\theta) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \theta^{\frac{v1}{2}-1}(1-\theta)^{\frac{v1}{2}-1}$$

and the incomplete beta function is defined by the integral

$$(11.9) \qquad I_{\theta_0}\left(\frac{v_1}{2}, \frac{v_1}{2}\right) = \int_0^{\theta_0} f(\theta)\, d(\theta),\ 0 \leq \theta_0 \leq 1$$

The upper percentage point of the F-distribution, $F_{1-\alpha}$, is related to the lower percentage point θ_α of the incomplete Beta function as follows. From (11.3) $\theta_\alpha = [1 + (v_1/v_2)F_{1-}\alpha]^{-1}$; that is

$$(11.10) \qquad \int_{F_{1-\alpha}}^{\infty} f\left(\frac{v_2}{2}F\right)dF = I_{\theta_0}\left(\frac{v_2}{2}, \frac{v_1}{2}\right)$$

For example, from the F-table $F_{,95}(10,20) = 2.35$. Thus $\theta_{.05} = [1+(10/20)\ 2.35]^{-1} = 0.460$; i.e., $I_{.46}(20,10)=0.05$. In probability terms, $Pr(F \geq F_{.95}) = Pr(F \geq 2.35) = Pr(\theta \leq 0.460) = Pr(\theta \leq \theta_{.05}) = 0.05$. Note the reversal of the degrees of freedom v_1 and v_2 in the F-distribution and the incomplete beta function. For computing the $(1-\alpha)\%$ point, $\theta_{1-\alpha}$, the beta-prime form is more convenient to use. Thus, $\theta_{.95} = 1 - \theta_{.05} = 1 - 0.460 = 0.540 = \lambda(1+\lambda)^{-1}$, where $\lambda = (10/20)(2.35)$. In probability terms since the beta-prime form is a monotonic increasing function of the λ's, $Pr(F \geq F_{.95}) = Pr(\theta \geq \theta_{.95}) = 0.05$. [See first sentence, second paragraph of Section 13 below]

12. WILKS Λ OR PRODUCT OF THE ROOTS

The LRC U(p,m,n) is the root of an equation of the form $|\mathbf{W}_1| - \lambda|\mathbf{W}_2| = 0$, where $\mathbf{W}_1$ is distributed as $W_p(\Sigma, m)$ and $\mathbf{W}_2$ is distributed as $W_p(\Sigma, n)$. In terms of the roots of the equation, $\lambda = |\mathbf{W}_1| \bullet |\mathbf{W}_2|^{-1} = |\mathbf{W}_1\mathbf{W}_2^{-1}|$, which is the product of the roots $\Pi_i\lambda_i$ in the determinantal equation $|\mathbf{W}_1 - \lambda\, \mathbf{W}_2| = |\mathbf{W}_1\mathbf{W}_2^{-1} - \lambda I_p| = 0$. The product $\Pi_i\lambda_i$ is known as Wilks Λ. For testing the null hypothesis **H**: $\mathbf{B}_1 = \mathbf{B}_1{}^*$, where $\mathbf{B}_1{}^*$ is specified (Section 6) we compute the LRC given by (6.8) and compare the number with $u_{p,q1,N-q1-q2}(\alpha)$, the α significance point of the $U(p,q_1,N-q_1-q_2)$ distribution.We reject the null hypothesis H if the computed $U_{p,q1,N-q1-q2} \le u_{p,q1,N-q1-q2}(\alpha)$ and accept H otherwise at α% level of significance.

Wilks test is used in one-way multivariate analysis of variance. We test the hypothesis H_b: $\mu_1 = \ldots = \mu_k$ given that $\Sigma_1 = \ldots = \Sigma_k$. Since under H_b (over subspace ω) the total observations are viewed as one sample, the maximum likelihood estimates are $\bar{\mathbf{x}}$ and S. That is, over subspace ω the "total" SSCP matrix is T = n S derived by considering all the data matrices as a single sample. Under the alternative hypothesis (that is, over space Ω) the maximum likelihood estimates of the μ_i is $'x_i$ and that of the common variance matrix is W/n, where $W = \Sigma_i\, n_i\, S_i$, the "within-groups" sum of squares and cross products (SSCP) matrix, and $n = n_1 + \ldots + n_k$. Thus

$$(12.1) \qquad \lambda_b = \{|\mathbf{W}|/|\mathbf{T}|\}^{n/2},$$

The total SSCP matrix

$$(12.2) \qquad T = \sum_{j=1}^{k} \sum_{j=i}^{k} (x_{ij} - \bar{\mathbf{x}})\,(x_{ij} - \bar{\mathbf{x}})'$$

is distributed as a $W_p(\Sigma, n-1)$. The "between-groups" matrix B is given by

$$(12.3) \qquad B = \sum_{j=1}^{k} n_j\, (\bar{x}_j - \bar{x})\,(\bar{x}_j - \bar{x})'$$

and is distributed as a $W_p(\Sigma, k-1)$. The difference between **T** and **B**,

$$(12.4) \qquad \mathbf{W} = \mathbf{T} - \mathbf{B},$$

is distributed independently of **B** as a $W_p(\Sigma, n-k)$. Thus, LRC for testing H_b is the (2/n)th power of (12.1); that is

$$(12.5) \qquad \lambda_b{}^{2/n} = |\mathbf{W}|/|\mathbf{T}|$$

which is distributed as U(p,n-k,n-1). We reject H_b at α% level of significance if the computed $U(p,n-k,n-1) \le u_{p,n-k,n-1}(\alpha)$; otherwise, accept H_b.

In the context of the usual multivariate analysis-of-variance form, where "hypothesis" matrix **H** = **B** and "error" matrix **W** = **E**, we could have used directly the root of equation $|\mathbf{B}| - \lambda|\mathbf{W}| = 0$, that is

$$(12.6) \qquad \lambda = \frac{|B|}{|W|} = |BW^{-1}| \sim U(p, K-1, n-k).$$

The detenninant $|\mathbf{BW}^{-1}| = \Pi_i\ \lambda_i$ = Wilks Λ.. Thus the LRC for testing $\mathbf{H}_b$ is given by (12.5) and $\Pr[U(p,k\text{-}1,n\text{-}k) \leq \mathbf{u}_{p,k-1,n-k}(\alpha)]=\alpha$ is the significance level when the null hypothesis H_b is true.

We could have used also the roots of the determinantal equation [or the root of equation of the form $|\mathbf{B}| + \lambda|\mathbf{B}+\mathbf{W}| = 0$]

$$(12.7) \qquad |\mathbf{B} - \theta(\mathbf{B}+\mathbf{W})| = 0$$

$$(12.8) \qquad |\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1} - \theta\,\mathbf{I}_p| = 0,$$

Thus $|\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}|$ is the product of the roots $\Pi_i\theta_i = \Pi_i\ \lambda_i(1+\lambda_i)^{-1}$ or Wilks Λ. = $\Pi_i\lambda_i = \Pi_i\theta_i(1-\theta_i)^{-1}$, a monotonic inoreasing function of the θ's.

For testing the hypothesis H_a: $\Sigma_1 = \ldots = \Sigma_k$ (test of homogeneity of covariances), the maximum likelihood estimates of μ_i are $\bar{x}_i$ under both H_a (that is, over subspace ω) and the alternative (that is, over space Ω). The maximum likelihood estimates of Σ_i is S=W/n over subspace ω and S_i over the alternative space **Ω**. Thus the **LRC** for testing $\mathbf{H}_a$ is the root

$$(12.9) \qquad \lambda_a = \Pi_i\,|S_i|^{ni/2} / |S|^{n/2},$$

or

$$(12.10) \qquad \lambda_a = \Pi_i\,|S_iS^{-1}|^{ni/2}.$$

Box (1949) suggested the use of unbiased estimates in place of the maximum likelihood estmates S_i and S. Thus, $S_{ui} = [n_i/(n_i-1)]S_i$ and $S_u = [n/(n-1)]S$. From (12.10) we have

$$(12.11) \qquad -2\log\lambda_a = \Sigma_i\,(n_i-1)\log|S_{ui}^{-1}S_u|$$

Box showed that

$$(12.12) \qquad M = \gamma\,\Sigma_i\,(n_i-1)\log|S_{ui}^{-1}S_u|,$$

where

$$(12.13) \qquad \gamma = 1\ \frac{2p^2+3p-1}{6(p+1)(k-1)}\left(\sum_{i=}^{k}\frac{1}{n_i-1}\ \frac{1}{n-k}\right)$$

follows asymptotically a $\chi^2_{p(p+1)(k-1)/2}$-distribution. The approximation is good if each n_i exceeds 20, and if k and p do not exceed 5. Note that for p =1, the quantity -2log λ_a of (12.10) reduces to Bartlett's test of homogeneity; that is

$$(12.14) \qquad n \log s^2 - \Sigma_i n_i \log s_i^2 \sim \chi^2_{(k-1)}.$$

For the test on complete homogeneity, that is **H**: $\mu_1 = \ldots = \mu_k$ and $\Sigma_1 = \ldots = \Sigma_k$, the root can be derived by noting that $\mathbf{H}_b$ is a conditional type of hypothesis and is true if the hypothesis $\mathbf{H}_a$ is true. Thus, $\mathbf{H} = \mathbf{H}_b\mathbf{H}_a$., implying that the root is equivalent to the product

$$(12.15) \qquad \lambda = \Pi_i \mid S_i S^{-1} \mid^{ni/2},$$

or

$$(12.16) \qquad -2 \log \lambda = n \log |S| - \Sigma_i n_i \log |S_i| .$$

The statistic (12.16), following Box, is asymptotically distributed as a chi-square distribution with (p/2)(k-1)(p+3) degrees of freedom.

Note that Wilks statistic or the LRC, in principle, can be integrated to obtain probabilities. The cases of p=1, k=2 are simply functions of an F-statistic. The evaluation of cases p=2, k=2 are still manageable but in other cases asymptotic expansions are resorted to. However, the evaluations even here could become quite unmanageable (e.g.,Anderson,1958; pp.195-202).

13. THE SUM OF THE ROOTS $V_1^{(s)}$ OR TRACE OF THE MATRIX, TR(*)

It may be observed that computing exact probabilitities of the **LRC** U(p,m,n) could be a tedious one and approximations have usually been resorted to in significance testing of hypotheses [Sections 7(Remark 2) & 12]. Thus transforming the LCR to monotonic functions of its roots and testing the equivalent hypotheses may be resorted to as an alternative.

Intuitively, it appears that good tests should reject the null hypothesis when the roots of the equation of the types (8.1) and (8.8) in some sense are large, indicating large departures of H from its metrics. Pillai (1954,1955) proposed the sum of the roots $V^{(s)} = \Sigma_i \theta_i$ [or tr$\{H(H+G)^{-1}\}$] and the sum of the roots $U^{(s)} = \Sigma_i \lambda_i$ [or tr (HG^{-1})] as test criteria for testing the multivariate hypotheses. He presented tables of $V^{(s)}$ for s = 2, 3, and 4 roots (Pillai, 1957) and their extensions to *s* = 8 along with tables of $U^{(s)}$ for s = 2,..., 8 (Pillai, 1960). The criterion $U^{(s)}$ was also proposed earlier by Lawley (1938) and Hotelling (1931,1951) while that of $V^{(s)}$ has been also earlier proposed by Bartlett (1939) and Nanda (1950). The former criterion is now known (Anderson,1984) as Lawley-Hotelling trace criterion and the latter as Bartlett-Nanda-Pillai trace criterion.

Mijares (1964a), by including the univariate and the Hotelling T^2 cases for s=1, extended the table of $V^{(s)}$ to $V_1^{(s)}$ to s=1,2,...,50, following the generalization of the moment equations of $V_1^{(s)}$ (Mijares, 1958; Pillai and Mijares, 1959),the subscript 1 being to indicate the sum as the *first elementatry symmetric function* (e.s.f.) of the nonzero roots. The generalizations of moment equations of the e.s.f. for any number s of the nonzero roots and their joint moments followed (Mijares,1961; 1964b). He proposed (Mijares, 1962) the set of e.s.f. tests for testing the multivariate hypotheses to overcome problems (Hotelling, 1936) of competing choices as to the best tests under certain situations. Approximations to the distributions of $V_1^{(s)}$ (Mijares, 1962; 1990a) and $U^{(s)}$ (Mijares,1990b) have been given.The normal approximations have reportedly been incorporated in some IBM-generated softwares for testing multivariate hypotheses.

REFERENCES

Anderson, T. W. (1958), *An Introduction to Multivariate Statistical Analysis*, New York, John Wiley & Sons.

(1984), *An Introduction to Multivariate Statistical Analysis, Second Edition*, New York, John Wiley & Sons.

Bartlett, M. S. (1939), "A note on tests of significance in multivariate analysis", *Proc. Camb. Phil. Soc.*, 180-185.

Bose, R. C. (1977), "Early History of Multivariate Statistical Analysis", *Multivariate Analysis IV*. 3-22, Amsterdam, North-Holland Publishing Company.

Box, G. E. P. (1949), "A general distribution theory for a class of likelihood criteria" *Biometrika*, **36**, 317-346

Fisher, R. A. (1930), "The use of multiple measurements in taxonomic problems", *Ann. Eugen.*,7, 179-188.

(1939), "The sampling distribution of some statistics obtained from non-linear equations",*Ann. Eugen.*, **9**, 238-249.

Girshick, M. A. (1939), "On the sampling theory of roots of determinantal equations", *Ann. Math. Stat.* **10**, 203-224.

Hotelling, H.(1931), " The generalization of Student's ratio", *Ann. Math .Stat.*, 2, 360-378.

(1936), "Relation between two sets of variates", *Biometrika*, 28, 321-377.

(1951), "A generalized T test and measure of multivariate dispersion", *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*", Los Angeles and Berkeley, University of California Press, 23-42.

Hsu, P. L. (1939), "On the distribution of the roots of certain determinantal equations", *Ann. Eugen.*, **9**, 250-258.

Kendall, M. (1980), *Multivariate Analysis*, Second Edition, London and High Wycombe, Charles Griffin & Company Ltd.

Mijares, T. A. (1958), "On the distribution of the sum of six roots of a matrix in multivariate analysis", master's thesis, University of the Philippines Library.

(1961), "The moments of elementary symmetric functions of the roots of a matrix in multivariate analysis", *Ann. Math. Stat.*, **32**, 1152-1160.

(1962), "The moments and approximate distributions of the elementary symmetric functions of the roots of a multivariate matrix", doctoral dissertation, Cambridge, Massachussetts, Harvard University.

(1964a), *Percentage Points of the Sum $V_1^{(s)}$ of s Roots (s = 1-50)*, Manila Statistical Center, University of the Philippines.

(1964b), "On elementary symmetric functions of the roots of a multivariate matrix: distributions", *Ann. Math. Stat.*, **35**, 1186-98.

(1990a), "The normal approximation to the Bartlett-Nanda-Pillai trace test in multivariate analysis", *Biometrika*, 77, 230-3.

(1990b), "On the normal approximation to the Lawley-Hotelling trace criterion", *Biometrika*, 77,443.

Mood, A.M. (1951), "On the distribution of the characteristic roots of normal second-moment matrices", *Ann. Math. Stat.* 22, 266-273

Nanda, D. N. (1950), "Distribution of the sum of the roots of a determinantal equation under a certain condition", *Ann. Math. Stat.*, 21, 432-439.

Pillai, K. C. S. (1954), On some distribution problems in multivariate analysis, Mimeograph Series No. 88, Institute of Statistics, University of North Carolina.

(1955), "Some new test criteria in multivariate analysis", *Ann Math Stat.*, 26, 117- 121

(1957), *Concise Tables for Statisticians.* Manila, The Statistical Center, University of the Philippines.

and Mijares, T. A. (1959), "On the moments of the trace of a matrix and approximations to its distribution", *Ann. Math. Stat.*, 30, 1135-40

(1960), *Statistical Tables for Tests of Multivariate Hypotheses.* Manila, The Statistical Center, University of the Philippines.

Rao, C. R. (1951), "An asymptotic expansion of the distribution of Wilks's criterion", *Bulletin of the International Statistical Institute*, 33, Part II, 177-180.

Schatzoff, M. (1966), "Exact distributions of Wilks's likelihood criterion", *Biometrika*, 53, 347-358.

Wilks, S. S., (1932), "Certain generalizations in the analysis of variance", *Biometrika*, 24, 471-494.

Wishart, J., (1928), "The generalized product moment distribution in samples from a normal multivariate population", *Biometrika*, 20A, 32-52.