Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 3240–3248
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Multiple discoveries: Distribution of roots of determinantal equations

T.W. Anderson[a, b, *]

[a] *Department of Economics, Stanford University, Stanford, CA 94305, USA*
[b] *Department of Statistics, Stanford University, Stanford, CA 94305, USA*

Available online 30 March 2007

Abstract

Five statisticians including S.N. Roy found the distribution of the characteristic roots of one sample matrix in the matrix of another almost simultaneously. This paper summarizes and contrasts their different methods of deriving the basic theorem.

MSC: 62H10; 03B55

Keywords: Roots of determinantal equations; Wishart distribution; *p*-statistics; Canonical correlations

1. Introduction

The history of science records many cases of a discovery made by two scientists independently and simultaneously, such as Newton and Leibnitz with calculus, Wallace and Darwin with evolution, and Church and Turing with decidability. However, there are not so many instances of *five* scientists developing an idea more or less independently and simultaneously, but that was the situation with S.N. Roy together with R.A. Fisher, P.L. Hsu, M.A. Girshick, and A.M. Mood in 1939. The development was the distribution of the characteristic or latent roots or eigenvalues of one sample covariance matrix in the metric of another sample covariance matrix.

There were good reasons for the wide-spread interest in these characteristic roots and functions of them. Hotelling (1936) had proposed canonical correlations to characterize the relations between two sets of variables. The distribution of the squares of the canonical correlations when the population canonical correlations are zero is the same as the distribution of the characteristic roots of one sample covariance matrix in the metric of another.

Fisher (1938) had considered multivariate analysis of variance and regression models. He asked for the linear combination of observed variables that maximized the effect variance relative to the error variance; the maximized ratio is the largest characteristic root of one covariance matrix in the matrix of another. Bose and Roy (1938) treated a similar situation.

* Corresponding author at. Department of Statistics, Stanford University, Stanford, CA 94305, USA. Tel.: +1 650 723 4732; fax: +1 650 725 8977.
E-mail address: twa@stanford.edu.

0378-3758/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.jspi.2007.03.008

2. The problem

The mathematical problem is to find the distribution of the roots of the equation

$$|\mathbf{A} - f(\mathbf{A} + \mathbf{B})| = 0, \tag{2.1}$$

where $\mathbf{A}$ and $\mathbf{B}$ $(p \times p)$ have Wishart distributions with common population covariance matrix $\mathbf{\Sigma}$ and degrees of freedom m and n, respectively. The density of the Wishart distribution $W(\mathbf{\Sigma}, m)$ of $\mathbf{A}$, for example, is

$$w(\mathbf{A}|\mathbf{\Sigma}, m) = \frac{1}{2^{(1/2)mp}|\mathbf{\Sigma}|^{(1/2)m}\Gamma_p((1/2)m)}|\mathbf{A}|^{(1/2)(m-p-1)}\mathrm{e}^{-(1/2)\,\mathrm{tr}\,\mathbf{\Sigma}^{-1}\mathbf{A}}, \tag{2.2}$$

where p is the dimensionality of $\mathbf{A}$ and

$$\Gamma_p\left(\frac{1}{2}m\right) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(m+1-i)\right]. \tag{2.3}$$

Here $\mathbf{A} = \sum_{\alpha=1}^{m}\mathbf{x}_\alpha\mathbf{x}_\alpha'$, and $\mathbf{x}_1, \ldots, \mathbf{x}_m$ have the normal distribution $\mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$. Equivalently, when $\mathbf{x}_1, \ldots, \mathbf{x}_{m+1}$ have the normal distribution $\mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, the matrix $\mathbf{A} = \sum_{\alpha=1}^{m+1}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})'$ has the Wishart density (2.2). Here $\bar{\mathbf{x}} = [1/(m+1)]\sum_{\alpha=1}^{m+1}\mathbf{x}_\alpha$. Let the roots of (2.1) be $f_1 > f_2 > \cdots > f_p$ $(1 > f_1,\ \ f_p > 0)$.

Theorem 1. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are independently distributed as* $W(\mathbf{\Sigma}, m)$ *and* $W(\mathbf{\Sigma}, n)$, *respectively, the density of* $f_1, \ldots, f_p$ *is*

$$\frac{\pi^{(1/2)p^2}\Gamma_p[(1/2)(m+n)]}{\Gamma_p((1/2)m)\Gamma_p((1/2)n)\Gamma_p((1/2)p)}\prod_{i=1}^{p} f_i^{(1/2)(m-p-1)}\prod_{i=1}^{p}(1-f_i)^{(1/2)(n-p-1)}\prod_{i<j}(f_i - f_j) \tag{2.4}$$

over $1 > f_1 > \cdots > f_p > 0$ *and* 0 *otherwise*.

Note that the roots of (2.1) are the same as the roots of

$$|\mathbf{CAC}' - f(\mathbf{CAC}' + \mathbf{CBC}')| = 0. \tag{2.5}$$

Thus for any nonsingular $\mathbf{C}$ the roots of (2.1) are invariant with respect to transformations

$$\mathbf{A}^* = \mathbf{CAC}', \quad \mathbf{B}^* = \mathbf{CBC}'. \tag{2.6}$$

If $\mathbf{C}$ is chosen so that $\mathbf{C\Sigma C}' = \mathbf{I}$, then $\mathbf{A}^*$ and $\mathbf{B}^*$ have densities $w(\mathbf{A}^*|m, \mathbf{I}_p)$ and $w(\mathbf{B}^*|n, \mathbf{I}_p)$, respectively. Hence, in deriving Theorem 1, we can assume $\mathbf{\Sigma} = \mathbf{I}_p$.

3. R.A. Fisher (1890–1962)

The work of Fisher and Hsu was published in adjacent papers in the 1939 volume of the *Annals of Eugenics* (of which Fisher was the editor). Fisher (1939) actually proved Theorem 1 for $p = 2$; he invited Hsu "to give a complete demonstration of an analytic solution" for general p. We shall review Fisher's derivation including a little more detail than he gave.

The density of $\mathbf{A} + \mathbf{B} = \mathbf{G}$, say, is (2.2) with m replaced by $m + n$. The conditional density of $\mathbf{A}$ given $\mathbf{G}$ is of the form

$$\text{constant}\ \frac{|\mathbf{A}|^a|\mathbf{G} - \mathbf{A}|^b}{|\mathbf{G}|^{a+b}}. \tag{3.1}$$

If $\mathbf{G} = \mathbf{I}$, the conditional density for $p = 2$ is

$$\frac{\Gamma_2[(1/2)(m+n)]}{\Gamma_2((1/2)m)\Gamma_2((1/2)n)}|\mathbf{A}|^{(1/2)(m-3)}|\mathbf{I} - \mathbf{A}|^{(1/2)(n-3)}. \tag{3.2}$$

Fisher transforms a_{11}, a_{12}, a_{22} to ϕ_1, ϕ_2, and a_{12}; the Jacobian is $\phi_1 - \phi_2$. Integration with respect to a_{12} yields the marginal density of f_1 and f_2 as

$$\frac{\pi^2\Gamma_2[(1/2)(m+n)]}{\Gamma_2((1/2)m)\Gamma_2((1/2)n)}(f_1 f_2)^{(1/2)(m-3)}[(1-f_1)(1-f_2)]^{(1/2)(n-3)}(f_1-f_2). \tag{3.3}$$

Fisher announced the general Theorem 1.

4. P.L. Hsu (1909–1970)

Hsu (1939) gave a complete derivation of Theorem 1. Let $d_1 > \cdots > d_p$ be the roots of

$$|\mathbf{A} - d\mathbf{B}| = 0. \tag{4.1}$$

Then $d_i = f_i/(1-f_i)$, $i = 1, \dots, p$. Define the $p \times p$ matrix $\mathbf{W}$ and the diagonal matrix $\mathbf{D}$ by

$$\mathbf{A} = \mathbf{WDW}', \quad \mathbf{B} = \mathbf{WW}' \tag{4.2}$$

and $w_{1j} \geqslant 0$, $j=1, \dots, p$. The density of $\mathbf{D}$ and $\mathbf{W}$ is obtained by substitution in $w(\mathbf{A}|m, \mathbf{I}_p)w(\mathbf{B}|n, \mathbf{I}_p)$ and multiplying by the Jacobian of transformation (4.2). The marginal density of $\mathbf{D}$ is the integral of the joint density with respect to $\mathbf{W}$.

The most difficult part of this procedure is the calculation of the Jacobian. Hsu gives the details for $p=3$; he asserts that “the method of proof will hold good for any p”. For the purpose of the present exposition we give the details for $p=2$.

Let $\mathbf{J}$ denote the matrix of partial derivatives of a_{11}, a_{12}, a_{22}, b_{11}, b_{12}, b_{22} (row indicators) with respect to d_1, d_2, w_{11}, w_{12}, w_{21}, w_{22} (column indicators). Then

$$\mathbf{J} = \begin{bmatrix} w_{11}^2 & w_{12}^2 & 2w_{11}d_1 & 2w_{12}d_2 & 0 & 0 \\ w_{11}w_{21} & w_{12}w_{22} & w_{21}d_1 & w_{22}d_2 & w_{11}d_1 & w_{12}d_2 \\ w_{21}^2 & w_{22}^2 & 0 & 0 & 2w_{21}d_1 & 2w_{22}d_2 \\ 0 & 0 & 2w_{11} & 2w_{12} & 0 & 0 \\ 0 & 0 & w_{21} & w_{22} & w_{11} & w_{12} \\ 0 & 0 & 0 & 0 & 2w_{21} & 2w_{22} \end{bmatrix}. \tag{4.3}$$

Let the cofactor of w_{ij} in the matrix $\mathbf{W}$ be W_{ij}. (In the case of $p=2$, $W_{ij} = (-1)^{i+j}w_{2-i,2-j}$.) Now multiply $\mathbf{J}$ on the left by the matrix

$$\mathbf{K} = \begin{bmatrix} W_{11}^2 & 2W_{11}W_{21} & W_{21}^2 & -W_{11}^2 d_1 & -2W_{11}W_{21}d_1 & -W_{21}^2 d_1 \\ W_{12}^2 & 2W_{12}W_{22} & W_{22}^2 & -W_{12}^2 d_2 & -2W_{12}W_{22}d_2 & -W_{22}^2 d_2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.4}$$

The product $\mathbf{KJ}$ is

$$\mathbf{KJ} = \begin{bmatrix} |\mathbf{W}|^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & |\mathbf{W}|^2 & 0 & 0 & 0 & 0 \\ w_{21}^2 & w_{22}^2 & 0 & 0 & 2w_{21}d_1 & 2w_{22}d_2 \\ 0 & 0 & 2w_{11} & 2w_{12} & 0 & 0 \\ 0 & 0 & w_{21} & w_{22} & w_{11} & w_{12} \\ 0 & 0 & 0 & 0 & 2w_{21} & 2w_{22} \end{bmatrix}. \tag{4.5}$$

The determinant of (4.5) is

$$|\mathbf{KJ}| = 8|\mathbf{W}|^4 \begin{vmatrix} 0 & 0 & w_{21}d_1 & w_{22}d_2 \\ w_{11} & w_{12} & 0 & 0 \\ w_{21} & w_{22} & w_{11} & w_{12} \\ 0 & 0 & w_{21} & w_{22} \end{vmatrix} = 8|\mathbf{W}|^5 \begin{vmatrix} w_{21}d_1 & w_{22}d_2 \\ w_{21} & w_{22} \end{vmatrix}$$
$$= 8|\mathbf{W}|^5 w_{21}w_{22}(d_1 - d_2). \tag{4.6}$$

The determinant of **K** is

$$|\mathbf{K}| = 2\begin{vmatrix} W_{11}^2 & W_{11}W_{21} \\ W_{12}^2 & W_{12}W_{22} \end{vmatrix} = 2W_{11}W_{12}\begin{vmatrix} W_{11} & W_{21} \\ W_{12} & W_{22} \end{vmatrix} = 2W_{11}W_{12}|\mathbf{W}|. \tag{4.7}$$

The Jacobian is the determinant of **J**:

$$|\mathbf{J}| = \frac{|\mathbf{KJ}|}{|\mathbf{K}|} = 4|\mathbf{W}|^4(d_1 - d_2). \tag{4.8}$$

Hsu carried out the details for $p = 3$; they are a little more complicated than the details for $p = 2$. Hsu asserts that for general p his method leads to the Jacobian

$$|\mathbf{J}| = 2^p|\mathbf{W}|^{p+2}\prod_{i<j}(d_i - d_j). \tag{4.9}$$

The density of **D** and **W** is

$$\frac{2^p}{2^{(1/2)(m+n)p}\Gamma_p((1/2)m)\Gamma_p((1/2)n)}|\mathbf{W}'\mathbf{W}|^{(1/2)(m+n-p)}\prod d_i^{(1/2)(m-p-1)}$$
$$\cdot \exp\left[-\frac{1}{2}\sum_{i,j=1}^{p}(1+d_i)w_{ij}^2\right]\prod_{i<j}(d_i - d_j). \tag{4.10}$$

To obtain the marginal density of **D** we need to integrate (4.10) with respect to the elements of **W**.

Lemma.

$$\int \cdots \int |\mathbf{W}'\mathbf{W}|^{(1/2)(m+n-p)}\frac{1}{(2\pi)^{(1/2)p^2}}\exp\left[-\frac{1}{2}\sum_{i,j=1}^{p}(1+d_i)w_{ij}^2\right]\prod_{i,j=1}^{p} dw_{ij}$$
$$= 2^{(1/2)p(m+n-p)}\frac{\Gamma_p(m+n)}{\Gamma_p(p)}\prod_{i=1}^{p}(1+d_i)^{-(1/2)(m+n)}. \tag{4.11}$$

Proof. The left-hand side of (4.11) is $\mathscr{E}|\mathbf{WW}'|^{(1/2)(m+n-p)}$ when $\mathbf{W}=(w_1, \ldots, w_p)$ and $w_1, \ldots, w_p$ are independently distributed according to $\mathrm{N}[\mathbf{0}, (\mathbf{I}+\mathbf{D})^{-1}]$. Then $\mathbf{H} = \mathbf{WW}'$ has the density (conditional on **D**) of

$$w[\mathbf{H}|(\mathbf{I}+\mathbf{D})^{-1}, p] = \frac{(\mathbf{I}+\mathbf{D})^{(1/2)p}}{2^{(1/2)p^2}\Gamma_p((1/2)p)}|\mathbf{H}|^{-1/2}\mathrm{e}^{-(1/2)\mathrm{tr}(\mathbf{I}+\mathbf{D})\mathbf{H}}. \tag{4.12}$$

The integral on the left-hand side of (4.11) is

$$\frac{|\mathbf{I}+\mathbf{D}|^{(1/2)p}}{2^{(1/2)p^2}\Gamma_p((1/2)p)}\int \cdots \int |\mathbf{H}|^{(1/2)(m+n-p-1)}\exp\left[-\frac{1}{2}\mathrm{tr}(\mathbf{I}+\mathbf{D})\mathbf{H}\right]\mathrm{d}\mathbf{H}$$
$$= \frac{|\mathbf{I}+\mathbf{D}|^{(1/2)p}}{2^{(1/2)p^2}\Gamma_p((1/2)p)} \cdot \frac{\Gamma_p[(1/2)(m+n)]2^{(1/2)(m+n)p}}{|\mathbf{I}+\mathbf{D}|^{(1/2)(m+n)p}}\int \cdots \int w[\mathbf{H}|(\mathbf{I}+\mathbf{D})^{-1}, m+n]\,\mathrm{d}\mathbf{H} \tag{4.13}$$

which is the right-hand side of (4.11). □

Hsu refers to Wilks (1932) for this result.

The integral of (4.10) with respect to **W** gives the density of **D** as

$$\frac{2^p}{\Gamma_p((1/2)m)\Gamma_p((1/2)n)}\frac{\pi^{(1/2)p^2}\Gamma_p[(1/2)(m+n)]}{\Gamma_p((1/2)p)}\prod_{i=1}^{p}(1+d_i)^{-(1/2)(m+n)}. \tag{4.14}$$

Substitution of d_i by $f_i/(1-f_i)$ yields the theorem.

Deemer and Olkin (1951) gave more details of Hsu's development.

5. S.N. Roy (1906–1964)

Roy also derived the distribution of the characteristic roots (referred to as "p-statistics") for arbitrary p. He gave a geometric interpretation of the characteristic roots in the $(m+n)$-dimensional space of observations. Then $d_i=\tan^2\phi_i$. In place of transformation (4.2) of **A**, **B** to **D**, **W**, Roy makes transformation

$$a_{ij}=\sum_{k=1}^{p}v_{ik}v_{jk}\sin^2\phi_k, \tag{5.1}$$

$$b_{ij}=\sum_{k=1}^{p}v_{ik}v_{jk}\cos^2\phi_k. \tag{5.2}$$

Then

$$\mathbf{A}+\mathbf{B}=\mathbf{V}\mathbf{V}'. \tag{5.3}$$

The joint density of **A** and **B** is transformed to the density of **V** and $\mathbf{\Phi}=\operatorname{diag}(\phi_1,\ldots,\phi_p)$. Roy treated explicitly the Jacobian for $p=3$. For $p=2$ the Jacobian is

$$\begin{vmatrix} -2v_{11}^2c_1s_1 & -2v_{12}^2c_2s_2 & 2v_{11}c_1^2 & 2v_{12}c_2^2 & 0 & 0\\ -2v_{11}v_{21}c_1s_1 & -2v_{12}v_{22}c_2s_2 & v_{21}c_1^2 & v_{22}c_2^2 & v_{11}c_1^2 & v_{12}c_2^2\\ -2v_{21}^2c_1s_1 & -2v_{22}^2c_2s_2 & 0 & 0 & 2v_{21}c_1^2 & 2v_{22}c_2^2\\ 2v_{11}^2c_1s_1 & 2v_{12}^2c_2s_2 & 2v_{11}s_1^2 & 2v_{12}s_2^2 & 0 & 0\\ 2v_{11}v_{21}c_1s_1 & 2v_{12}v_{22}c_2s_2 & v_{21}s_1^2 & v_{22}s_2^2 & v_{11}s_1^2 & v_{12}s_2^2\\ 2v_{21}^2c_1s_1 & 2v_{22}^2c_2s_2 & 0 & 0 & 2v_{21}s_1^2 & 2v_{22}s_2^2 \end{vmatrix}, \tag{5.4}$$

where $c_i=\cos\phi_i$ and $s_i=\sin\phi_i$. Addition of the fourth row to the first, the fifth to the second, and the sixth to the third in (5.4) gives $4c_1s_1c_2s_2$ times

$$\begin{vmatrix} 0 & 0 & v_{11} & v_{12} & 0 & 0\\ 0 & 0 & v_{21} & v_{22} & v_{11} & v_{12}\\ 0 & 0 & 0 & 0 & v_{21} & v_{22}\\ v_{11}^2 & v_{12}^2 & 2v_{11}s_1^2 & 2v_{12}s_2^2 & 0 & 0\\ v_{11}v_{21} & v_{12}v_{22} & v_{21}s_1^2 & v_{22}s_2^2 & v_{11}s_1^2 & v_{12}s_2^2\\ v_{21}^2 & v_{22}^2 & 0 & 0 & 2v_{21}s_1^2 & 2v_{22}s_2^2 \end{vmatrix}. \tag{5.5}$$

The matrix in (5.5) is essentially a re-arrangement of the rows of (4.3) and replacement of d_1 and d_2 by s_1^2 and s_2^2, respectively.

Roy actually treated the case of $p=3$ in detail by manipulation of rows and columns. He asserts that for general p the Jacobian is

$$\text{constant}\prod_{i=1}^{p}c_is_i\prod_{i<j}(s_i^2-s_j^2)|\mathbf{V}|^{p+2}. \tag{5.6}$$

Roy obtained the joint density of $|\mathbf{V}|$ and the ϕ's. The integration of the joint density with respect to **V** gives the density of the ϕ's. However, Roy did not evaluate the normalizing constant in this paper. In his 1957 monograph he gave the full distribution.

Roy referred to a paper by Bose and Roy (1939) and a paper by Mahalanobis et al. (1937). The paper reviewed here has the footnote "Paper received 20th June, 1939". That issue of *Sankhyā* was received in the United States and in Great Britain much later because of the disruptions of World War II.

6. M.A. Girshick (1906–1955)

Girshick (1939), a student of Hotelling, treated the distribution of canonical correlations (Hotelling, 1936). Let the observation vector $\mathbf{X}$ be partitioned into p_1 and p_2 $(\geqslant p_1)$ components, respectively, as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}. \tag{6.1}$$

Let the sum of squares be similarly partitioned:

$$\mathbf{C} = \sum_{\alpha=1}^{n} x_\alpha x_\alpha' = \begin{bmatrix} \sum_{\alpha=1}^{n} \mathbf{x}_\alpha^{(1)} \mathbf{x}_\alpha^{(1)'} & \sum_{\alpha=1}^{n} \mathbf{x}_\alpha^{(1)} \mathbf{x}_\alpha^{(2)'} \\ \sum_{\alpha=1}^{n} \mathbf{x}_\alpha^{(2)} \mathbf{x}_\alpha^{(1)'} & \sum_{\alpha=1}^{n} \mathbf{x}_\alpha^{(2)} \mathbf{x}_\alpha^{(2)'} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}. \tag{6.2}$$

The sample canonical correlations are the roots of

$$0 = \begin{bmatrix} -r\mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & -r\mathbf{C}_{22} \end{bmatrix} = r^{p_1-p_2} |\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} - r^2\mathbf{C}_{11}| \cdot |-\mathbf{C}_{22}|. \tag{6.3}$$

This equation is of the form of (2.1) with $\mathbf{A}$ replaced by $\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$, $\mathbf{A}+\mathbf{B}$ by $\mathbf{C}_{11}$, and f by r^2. If the vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are normally distributed and uncorrelated (that is, $\mathscr{E}\mathbf{X}^{(1)}\mathbf{X}^{(2)'} = \mathbf{0}$), then $\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} = \mathbf{A}$ and $\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} = \mathbf{B}$ are independently distributed according to Wishart distributions $W(\mathbf{\Sigma}_{11}, p_2)$ and $W(\mathbf{\Sigma}_{11}, n-p_2)$, respectively. However, Girshick did not make the use of this fact. Instead he approached the problem for $p_1 = 2$ $(\leqslant p_2)$ by finding the joint distribution of

$$q^2 = \frac{\begin{vmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{vmatrix}}{|\mathbf{C}_{11}||\mathbf{C}_{22}|} = \frac{|\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}|}{|\mathbf{C}_{11}|} \tag{6.4}$$

and

$$z = \frac{\begin{vmatrix} \mathbf{0} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{vmatrix}}{|\mathbf{C}_{11}| \cdot |\mathbf{C}_{22}|} = \frac{|-\mathbf{C}_{12}\mathbf{C}_{22}\mathbf{C}_{21}|}{|\mathbf{C}_{11}|}. \tag{6.5}$$

For $p_1 = 2$ the roots r_1^2 and r_2^2 are functions of

$$q^2 = r_1^2 r_2^2, \quad z = (1-r_1^2)(1-r_2^2) = 1 - (r_1^2 + r_2^2) + r_1^2 r_2^2. \tag{6.6}$$

Under the assumption of normality, $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent and $\mathscr{E}\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{0}$. The distribution of the roots of (6.3) conditional on $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_n^{(2)}$ does not depend on what this array is. In his proof Girshick took

$$(\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_n^{(2)}) = (\mathbf{I}_{p_2}, \mathbf{0}). \tag{6.7}$$

Then $\mathbf{C}_{22} = \mathbf{I}_{p_2}$, and

$$\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} = \sum_{\alpha=1}^{p_2} \mathbf{x}_\alpha^{(1)} \mathbf{x}_\alpha^{(1)'}, \tag{6.8}$$

$$\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} = \sum_{\alpha=p_2+1}^{n} \mathbf{x}_\alpha^{(1)} \mathbf{x}_\alpha^{(1)'}. \tag{6.9}$$

Moreover, the covariance matrix of $\mathbf{X}^{(1)}$ can be taken as $\mathbf{I}_2$. Then $\mathbf{y}=(x_{11},\ldots,x_{1n})'$ and $\mathbf{z}=(x_{21},\ldots,x_{2n})'$ are independently normally distributed with covariance $\mathbf{I}_n$. Since they are normally distributed, the distributions are spherically symmetric. Girshick transformed the two vectors into polar coordinates. The quantities q^2 and z are functions of four angles, which are uniformly distributed. Two angles are integrated out to get the distribution of q^2 and z. Use of (6.6) leads to the distribution of the roots.

Since Girshick had not completed the proof of Theorem 1 before Fisher and Hsu, his results did not fulfill the requirements for a Ph.D. He had yet another hurdle in getting his doctorate. In 1944 he presented some preliminary results on the distributions of $\mathbf{A}=\sum_{\alpha=1}^{m}\mathbf{x}_\alpha\mathbf{x}_\alpha'$, when $\mathbf{x}_\alpha$ is an observation on $\mathbf{X}_\alpha$ distributed according to $\mathrm{N}(\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma})$, $\alpha=1,\ldots,n$. Independently I had derived the distribution of $\mathbf{A}$ when $\sum_{\alpha=1}^{n}\boldsymbol{\mu}_\alpha\boldsymbol{\mu}_\alpha'$ had rank 1 or 2. We published a joint paper (Anderson and Girshick, 1944). Girshick subsequently received his Ph.D. on the basis of research in sequential analysis.

7. A.M. Mood (1915–)

Mood also derived the distribution of the roots at about the same time as the other four authors referred to above. Before he could publish his results, the issue of the *Annals of Eugenics* with the papers of Fisher and Hsu arrived in Princeton where Mood was a graduate student. I remember S.S. Wilks, Mood's and my dissertation advisor, telling me about Mood seeing their papers; Wilks said "You should have seen Alex's face!". Those papers meant that he could not use these results for his dissertation.

When I became editor of the *Annals of Mathematical Statistics* in 1950, I encouraged Mood to prepare a paper with his results for publication because I thought he had some interesting ideas—different from the other authors—for calculating the Jacobian and calculating the constant in the density of the roots.

Mood (1951) gave an explicit procedure for calculating the Jacobian of the transformation $\mathbf{A}=\mathbf{QFQ}'$, $\mathbf{B}=\mathbf{Q}(\mathbf{I}-\mathbf{F})\mathbf{Q}'$. To facilitate comparison with the approach of Hsu, I modify Mood's description to the transformation $\mathbf{A}=\mathbf{WDW}'$, $\mathbf{B}=\mathbf{WW}'$. The Jacobian is a polynomial in the elements of $\mathbf{D}$ (diagonal) and $\mathbf{W}$. He showed that d_i-d_j, $i<j$, was a factor of this polynomial. To illustrate the argument consider (4.3) for $p=2$. To show d_1-d_2 is a factor write (4.3) with d_2 replaced by d_1. Then subtract d_1 times the fourth row from the first row, d_1 times the fifth row from the second row, and d_1 times the sixth row from the third row obtaining

$$\begin{vmatrix} w_{11}^2 & w_{12}^2 & 0 & 0 & 0 & 0 \\ w_{11}w_{21} & w_{12}w_{22} & 0 & 0 & 0 & 0 \\ w_{21}^2 & w_{22}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2w_{11} & 2w_{12} & 0 & 0 \\ 0 & 0 & w_{21} & w_{22} & w_{11} & w_{12} \\ 0 & 0 & 0 & 0 & 2w_{21} & 2w_{22} \end{vmatrix} = 4|\mathbf{W}| \begin{vmatrix} w_{11}^2 & w_{12}^2 & 0 & 0 \\ w_{11}w_{21} & w_{12}w_{22} & 0 & 0 \\ w_{21}^2 & w_{22}^2 & 0 & 0 \\ 0 & 0 & w_{11} & w_{12} \end{vmatrix} = 0 \tag{7.1}$$

because every submatrix from the last pair of columns has determinant 0. Thus d_1-d_2 is a factor of (4.3). In general d_i-d_j for $i<j$ is a factor, and $\prod_{i<j}(d_i-d_j)$ is a factor. Since $|\mathbf{J}|$ is of total degree $p(p-1)/2$ in the d's, the other factor of $|\mathbf{J}|$ is a function of $\mathbf{W}$ alone. The density of $\mathbf{W}$ and $\mathbf{D}$ factors into a function of $\mathbf{W}$ and a function of $\mathbf{D}$. Hence $\mathbf{D}$ has density

$$\text{constant}\prod_{i=1}^{p} d_i^{(1/2)(m-p-1)}\prod_{i<j}(d_i-d_j). \tag{7.2}$$

Correspondingly the density of $f_1,\ldots,f_p$ is

$$K\prod_{i=1}^{p} f_i^{(1/2)(m-p-1)}\prod_{i=1}^{p}(1-f_i)^{(1/2)(n-p-1)}\prod_{i<j}(f_i-f_j). \tag{7.3}$$

It remains to find the constant K so that the integral of (7.3) over the range $1>f_1>\cdots>f_p\geqslant 0$.

Define

$$L(\alpha, \beta) = \int_0^1 \cdots \int_0^{f_{p-1}} \prod_{i=1}^{p} f_i^{\alpha} \prod_{i=1}^{p} (1-f_i)^{\beta} \prod_{i<j} (f_i - f_j). \tag{7.4}$$

Then

$$K = \frac{1}{L[(1/2)(m-p-1), (1/2)(n-p-1)]}. \tag{7.5}$$

Since

$$\prod_{i=1}^{p} f_i = \frac{|\mathbf{A}|}{|\mathbf{A}+\mathbf{B}|}, \quad \prod_{i=1}^{p} (1-f_i) = \frac{|\mathbf{B}|}{|\mathbf{A}+\mathbf{B}|}, \tag{7.6}$$

$$\mathscr{E} \frac{|\mathbf{A}|^r |\mathbf{B}|^s}{|\mathbf{A}+\mathbf{B}|^{r+s}} = \frac{L[(1/2)(m-p-1+r), (1/2)(n-p-1)+s]}{L[(1/2)(m-p-1), (1/2)(n-p-1)]}. \tag{7.7}$$

On the other hand, since **A** and **B** have Wishart distributions, so

$$\begin{aligned} \mathscr{E} \frac{|\mathbf{A}|^r |\mathbf{B}|^s}{|\mathbf{A}+\mathbf{B}|^{r+s}} &= \int \cdots \int \frac{1}{2^{(1/2)(m+n)p} \Gamma_p((1/2)m) \Gamma_p((1/2)n)} \frac{|\mathbf{A}|^{(1/2)(m-p-1)+r} |\mathbf{B}|^{(1/2)(n-p-1)+s}}{|\mathbf{A}+\mathbf{B}|^{r+s}} \\ &\quad \times \exp[-(1/2)\mathrm{tr}(\mathbf{A}+\mathbf{B})]\, \mathrm{d}\mathbf{A}\, \mathrm{d}\mathbf{B} \\ &= \frac{\Gamma_p((1/2)m+r)\Gamma_p((1/2)n+s)}{\Gamma_p((1/2)m)\Gamma_p((1/2)n)} 2^{(r+s)p} \\ &\quad \times \int \cdots \int |\mathbf{A}+\mathbf{B}|^{-(r+s)} w(\mathbf{A}|m+2r, \mathbf{I}_p) w(\mathbf{B}|n+2s, \mathbf{I})\, \mathrm{d}\mathbf{A}\, \mathrm{d}\mathbf{B}. \end{aligned} \tag{7.8}$$

Since $\mathbf{G} = \mathbf{A} + \mathbf{B}$ has the density $w(\mathbf{G}|m+n+2r+2s, \mathbf{I}_p)$ when **A** and **B** have densities $w(\mathbf{A}|m+2r, \mathbf{I}_p)$ and $w(\mathbf{B}|n+2s, \mathbf{I}_p)$, (7.8) is

$$\begin{aligned} &= \frac{\Gamma_p((1/2)m+r)\Gamma_p((1/2)n+s)}{\Gamma((1/2)m)\Gamma((1/2)n)} 2^{(r+s)p} \int \cdots \int |\mathbf{G}|^{-(r+s)} w(\mathbf{G}|m+n+2r+2s, \mathbf{I})\, \mathrm{d}\mathbf{G} \\ &= \frac{\Gamma_p((1/2)m+n)\Gamma_p((1/2)n+s)}{\Gamma_p((1/2)m)\Gamma_p((1/2)n)} 2^{(r+s)p} \frac{\Gamma_p[(1/2)(m+n)]}{\Gamma_p[(1/2)(m+n)+r+s]} \int \cdots \int w(\mathbf{G}|m+n, \mathbf{I}_p)\, \mathrm{d}\mathbf{G} \\ &= \frac{\Gamma_p((1/2)m+r)\Gamma_p((1/2)n+s)\Gamma_p\,[(1/2)(m+n)]}{\Gamma_p((1/2)m)\Gamma_p((1/2)n)\Gamma_p\,[(1/2)(m+n)+r+s]}. \end{aligned} \tag{7.9}$$

Now set (7.7) equal to (7.9), and let $m = n = p + 1$. Then

$$L(r, s) = L(0, 0) \frac{\Gamma_p(r+1/2)\Gamma_p(s+1/2)\Gamma_p(p+1)}{\Gamma_p^2[(1/2)(p+1)]\Gamma_p(r+s+p+1)}. \tag{7.10}$$

Finally, Mood showed that

$$L(0, 0) = \int_0^1 \cdots \int_0^{f_{p-1}} \prod_{i \neq j} (f_i - f_j) \prod_{i=1}^{p} \mathrm{d} f_i = \prod_{i=1}^{p} \frac{\Gamma(p-1+i)\Gamma(2p+1-2i)}{\Gamma(2p+1-i)}. \tag{7.11}$$

When (7.11) is used in (7.10), the theorem results.

8. My relations with S.N. Roy

Because I was interested in multivariate statistical analysis from my research for the doctoral dissertation, I became familiar with the work of S.N. Roy early. During World War II international mails were slow and irregular; issues of *Sankhyā* were usually delayed. Nevertheless, the contributions of the Indian statisticians, particularly Mahalanobis, Bose, Roy, and C.R. Rao, were read carefully.

In the spring term of the academic year 1948–1949 Roy was a visitor at Columbia University, my home institution; his teaching included a course in multivariate analysis. After Columbia he accepted a position at the University of North Carolina, where he stayed for the rest of his career. Over the years Roy and I had many interesting and useful discussions. While we studied common topics, our approaches to problems were often different. His book, *Some Aspects of Multivariate Analysis* (Roy, 1957), was published about the same time as my book *An Introduction to Multivariate Statistical Analysis* (Anderson, 1958). However, his book concentrated on summarizing and extending his research on distributions and confidence regions, while my book was more general expository.

Roy made many important advances in multivariate analysis, and he trained many students in the field. It was tragic that he passed away so young.

References

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
Anderson, T.W., Girshick, M.A., 1944. Some extensions of the Wishart distribution. Ann. Math. Statist. 15, 345–357.
Bose, R.C., Roy, S.N., 1938. The distribution of the studentised D^2-statistic. Sankhyā 4, 19–38.
Deemer, W.L., Olkin, I., 1951. The Jacobian of certain matrix transformations useful in multivariate analysis. Based on lectures of P.L. Hsu at the University of North Carolina 1947. Biometrika 38, 345–367.
Fisher, R.A., 1938. The statistical utilization of multiple measurements. Ann. Eugenics 8, 376–386.
Fisher, R.A., 1939. The sampling distribution of some statistics obtained from non-linear equations. Ann. Eugenics 9, 238–249.
Girshick, M.A., 1939. On the sampling theory of roots of determinantal equations. Ann. Math. Statist. 10, 203–224.
Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28, 321–377.
Hsu, P.L., 1939. On the distribution of the roots of certain determinantal equations. Ann. Eugenics 9, 250–258.
Mahalanobis, P.C., Bose, R.C., Roy, S.N., 1937. Normalisation of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. Sankhyā 3, 1–40.
Mood, A.M., 1951. On the distribution of the characteristic roots of normal second-moment matrices. Ann. Math. Statist. 22, 266–273.
Roy, S.N., 1939. p-Statistics or some generalisations in analysis of variance appropriate to multivariate problems. Sankhyā 4, 381–396.
Roy, S.N., 1957. Some Aspects of Multivariate Analysis. Wiley, New York.
Wilks, S.S., 1932. Certain generalizations in the analysis of variance. Biometrika 24, 471–494.