Distribution of p-Statistics on the Non-null Hypothesis

In a couple of notes published in this journal [1, 2] sometime ago, I announced that I had obtained the joint distribution of a class of p-statistics intended to test for the hypothesis that two samples have come from the same p-variate normal population and intended also to serve a variety of other purposes. If k^2_i ($i=1, 2, \ldots p$) be the p roots of the determinantal equation

$$| n_{ij} - k^2 n'_{ij} | = 0, \qquad (1)$$

where $\| n_{ij} \|$ and $\| n'_{ij} \|$ are the dispersion matrices of two samples Σ and Σ' of sizes n and n′ supposed to have been drawn from the same p-variate normal population with dispersion matrix $\| \alpha_{ij} \|$, then the distribution I had obtained was of the form

$$\text{Const.}\ (k_1^2 - k_2^2)\ (k_1^2 - k_p^2)(k_2^2 - k_3^2)\ldots(k_2^2 - k_p^2)\ldots(k^2_{p-1} - k_p^2) \times \prod_{i=1}^{p} \frac{k_i^{n-p-1}\, dk_i}{(1+\lambda^2 k_i^2)^{\frac{n+n'-2}{2}}} \qquad (2)$$

λ^2 being equal to n/n'

This has also come out in a paper since published in *Sankhya.* [3] Following up my previous investigation I proceeded to tackle the case where the hypothesis tested is not true *i.e.*, where the two samples are supposed to have been drawn from two populations with different dispersion matrices $\| \alpha_{ij} \|$ and $\| \alpha'_{ij} \|$ In that attempt I have since obtained the joint distribution of the k'_is for this general case of which the previous one is a special case.

If ξ_i ($i=1, 2, \ldots p$) denote the p roots of the determinantal equation $| \alpha_{ij} - \xi^2 \alpha'_{ij} | = 0$ (3) and if $\lambda^2 = n/n'$ and if we define quantities ζ_{ij} such that

$$\left.\begin{array}{l} \zeta^m_{ij} = \Gamma\left(\frac{m}{2}\right).\ \zeta^m_{ij}\ \zeta^n_{i'j'} = \Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) \text{ when} \\ i \neq i' \text{ and } j \neq j' \text{ but } \zeta^m_{ij}\zeta^n_{i'j'} = \Gamma\left(\frac{m+n}{2}\right) \text{ when} \\ i = i' \text{ and } j = j' \end{array}\right\} \qquad (4)$$

then the joint distribution of *p*-statistics k_i ($i=1, 2, \ldots p$) comes out in the form

$$\text{Const.}\ (k_1^2 - k_2^2)\ldots(k_1^2 - k_p^2)(k_2^2 - k_3^2) \ldots(k_2^2 - k_p^2)\ldots(k_{p-1}^2 - k_p^2)$$

$$\times\, E^{n+n'-p-2} \left\{ \prod_{ij=1}^{p} \frac{\zeta_{ij}}{(\xi_i + \lambda^2 k_j^2)^{\frac{1}{2}}} \right\} \times \left\{ \prod_{i=1}^{p} k_i^{n-p-1}\, dk_i \right\} \qquad (5)$$

where E is the determinant

$$\left| \frac{\zeta_{ij}}{(\xi_i^2 + \lambda^2 k_j^2)^{\frac{1}{2}}} \right|$$

It is easily seen that if the two populations have the same dispersion matrix i.e., if

$$\| \alpha_{ij} \| = \| \alpha'_{ij} \|, \text{ then } \xi_i\ (i=1, 2, \ldots p) = 1$$

and the joint distribution (5) is easily seen to reduce to (3) as it should.

The k_i's are supposed to be arranged in descending order of magnitude starting from k_1. It should be noticed with reference to the form (3) that in it we can make the substitutions for ζ_{ij}^m etc. given by (2.1) only after we have broken up $E^{n+n'-p-2}$ as a multinomial and multiplied it with the other factors. We cannot start with the substitutions. There is much work to be done on this distribution function before we can get at the separate distribution of the individual k_i's and some other suitable functions of k_i's which may be more appropriate for testing certain classes of hypothesis than the k_i's themselves. Work is at present being continued on this.

S. N. Roy.

Statistical Laboratory,
Presidency College,
Calcutta, 20-1-1940.

[1] Science and Culture, *5*, 131, 1939.

[2] *Ibid.*, *5*, 197, 1939.

[3] *Sankhya*, *4*, 3, 1939.