THEORETICAL STATISTICS

Chairman : Prof. R. A. FISHER

7th January, 1938

[Joint Meeting with the Section of Physics and Mathematics of the Indian Science Congress.]

THE DISTRIBUTION OF THE STUDENTISED D^2-STATISTIC

BY R. C. BOSE AND S. N. ROY.

STATISTICAL LABORATORY.

INTRODUCTION.

The problem of discrimination and classification is insistent in sciences. If we pose to ourselves the question :—Can the two populations from which two given samples have been drawn, be reasonably (*i.e.* on a given probability level) supposed to be distinct? The answer then is provided by proper tests of significance, which so far as the univariate case is concerned, have long been known. For the multivariate case, Karl Pearson in 1921 suggested his well known Coefficient of Racial Likeness (C), which was first used by Tildesley in a paper in 1921[15]. Karl Pearson discussed this coefficient in his own paper in 1926[13]. With the same object in view Romanovsky in 1928[14] proposed another coefficient (H). Work in this line has been continued by Wilks[16], Pearson[12] and others.

But the problem of classification is not solved thereby. Suppose we have three samples Σ_1, Σ_2, Σ_3 and we have reasonably satisfied ourselves that the populations π_1, π_2, π_3 from which these have been drawn are distinct from one another. It usually becomes necessary (especially in the biological sciences) to go further and ask ourselves a question like this :— Is the population π_1 in some significant sense, closer to π_2 or to π_3? For this purpose we need a "measure" of divergence between two different populations. This concept (as distinguished from a "test") of a measure of divergence between two populations was first introduced by P. C. Mahalanobis in 1925 in his presidential address to the Anthropological Section of the Indian Science Congress and was used elsewhere[7,8]. This idea was further developed on the theoretical side in a paper presented before the Indian Science Congress in 1928[9] in which he found, for the uncorrelated case, the first four moments of what may be called the classical form of D^2 (the measure of divergence devised by him) and also some approximate results for the "Studentized" form of the Statistic. The exact distribution as well as the moments of the classical D^2 for the correlated case was obtained by R. C. Bose in 1935[1], and a series of elegant properties of these moment functions were brought out in two mathematically powerful papers by S. N. Bose[2,3].

The Classical D^2 involved the population variances, which in actual practice had to be estimated from the data itself, the results thus obtained being only approximate. To avoid all approximation we must construct the statistic D^2 in such a way, that it involves only the sample readings. The need for this was all along recognised by P. C. Mahalanobis who had given approximate results for the uncorrelated case in 1928[9], and who explicitly defined in a note in 1936[10] the Studentised form (*i.e.* involving only the sample readings) of the D^2-statistic when the populations under consideration are different.

If we denote by Δ^2, the measure of divergence between the two populations, then the sample statistic D^2 is identical in form, except for some factors involving sample sizes, with the generalised T^2 of Hotelling, the distribution of which given in 1931[6] by Hotelling, provides us at once with the distribution of the Studentised D^2 in the special case.

The problem of distinguishing between two multivariate normal populations was also considered independently by R. A. Fisher in 1936[3]. Starting from a linear compound with arbitrary coefficients, of the variates, he chooses the compounding coefficients, so as to make a maximum, the ratio between the square of the difference of the sample means (for the compound character) and the within variance of the samples, for the same character. It should be noted that the compounding coefficients obtained after maximisation are really functions of the sample variances and covariances. Fisher has not explicitly calculated these coefficients, but a little algebra shows that after maximisation the ratio spoken of above, is proportional to Hotelling's[6] T^2, and Mahalanobis's Studentised D^2. The distribution of this ratio implied in Fisher's paper is in agreement with the one obtained by Hotelling[6].

The object of the present paper is to obtain the distribution of the Studentised D^2, for the general case $\Delta^2 \neq 0$, that is when the populations are different, and incidentally to verify the distribution of Hotelling[6] and Fisher[3].

It will be seen that the distribution of D^2 does not involve any population parameter except the function Δ^2 of the parameters. It is also interesting to note that the distribution found by Fisher[4] by an entirely different method for another statistic, namely a certain kind of multiple correlation coefficient, happens after a little transformation to be similar in form, to the distribution found by us.*

§1. DEFINITION OF THE STUDENTISED D²-STATISTIC.

Consider two samples Σ and Σ' of sizes n and n' from two multivariate normal populations π and π' with the same set of variances and covariances α_{ij} $(i, j=1, 2, \ldots p)$ where $\alpha_{ij}=\rho_{ij}\, \sigma_i\, \sigma_j$, σ_i and σ_j being the standard deviations for the i-th and j-th characters respectively, and ρ_{ij}, the correlation between the i-th and j-th characters. The matrix $\| \alpha_{ij} \|$ will be said to be the common dispersion matrix for the two populations. Let a_{ij}, a'_{ij} $(i, j=1, 2, \ldots\ldots\ldots p)$ denote the respective variances and covariances of the samples Σ and Σ', so that $\| a_{ij} \|$ and $\| a'_{ij} \|$ are their respective dispersion matrices. Let α_i, α'_i $(i=1, 2, \ldots\ldots p)$ be the means for the i-th character for the populations π and π' and let a_i, a'_i denote the corresponding quantities for the samples Σ and Σ'. Let us set

$$c_{ij}=\frac{n\, a_{ij}+n'\, a'_{ij}}{n+n'} \qquad \ldots \quad (1\cdot1)$$

Let c^{ij} as usual denote the minor of c_{ij} in the determinant $| c_{ij} |$, divided by the determinant itself. A like definition holds for α^{ij}. Then the Studentised D² is defined by

$$p\, D^2=c^{11}(a_1-a'_1)^2+c^{22}(a_2-a'_2)^2+\ldots\ldots+c^{pp}(a_p-a'_p)^2$$
$$+2c^{12}(a_1-a'_1)(a_2-a'_2)+\ldots\ldots+2c^{p-1,\,p}(a_{p-1}-a'_{p-1})(a_p-a_p') \ldots (1\cdot2)$$

Likewise if Δ^2 is the population value of D² then

$$p\Delta^2=\alpha^{11}(\alpha_1-\alpha'_1)^2+\alpha^{22}(\alpha_2-\alpha_2')^2+\ldots\ldots+\alpha^{pp}(\alpha_p-\alpha'_p)^2$$
$$+2\alpha^{12}(\alpha_1-\alpha_1')(\alpha_2-\alpha_2')+\ldots\ldots+2\alpha^{p-1,\,p}(\alpha_{p-1}-\alpha'_{p-1})(\alpha_p-\alpha_p') \qquad (1\cdot3)$$

*The introduction has been added by us after the conference. R. C. Bose and S. N. Roy.

§2. Invariance of D^2.

We shall show, that if from the primary p statistical variates, we obtain a new set of p variates, by a linear transformation with matrix $\|\lambda_{ij}\|$ of rank p, then D^2 remains unchanged.

If a_{ij} goes over to b_{ij}, it is easy to verify that

$$b_{ij} = \lambda_{i1}\lambda_{j1}a_{11} + \ldots\ldots\lambda_{ip}\lambda_{jp}\ a_{pp} + (\lambda_{i1}\lambda_{j2} + \lambda_{i2}\ \lambda_{j1})a_{12} + \ldots\ldots + (\lambda_{i,\ p-1}\lambda_{j,\ p} + \lambda_{i,\ p}\ \lambda_{j,\ p-1})a_{p-1,\ p} \quad (2\cdot1)$$

Thus

$$\| b_{ij} \| = \| \lambda_{ij} \|\ \| a_{ij} \|\ \| \lambda_{ji} \| \quad \ldots \quad (2\cdot2)$$

where $\|\lambda_{ji}\|$ is the matrix obtained from $\|\lambda_{ij}\|$ by interchanging the rows and columns.

Similarly if a'_{ij} goes over to b'_{ij}, then we get a formula corresponding to (2·1). Hence if c_{ij} goes over to d_{ij} we clearly have, remembering (1·1)

$$\| d_{ij} \| = \| \lambda_{ij} \|\ \| c_{ij} \|\ \| \lambda_{ji} \| \quad \ldots \quad (2\cdot3)$$

The same relation holds among corresponding determinants. Thus

$$| d_{ij} | = | c_{ij} |\ | \lambda_{ij} |^2 \quad \ldots \quad (2\cdot4)$$

Now it is easy to verify that

$$D^2 = \begin{vmatrix} c_{11} & c_{12} & \ldots\ldots\ldots & c_{1p} & L_1 \\ c_{21} & c_{22} & \ldots\ldots\ldots & c_{2p} & L_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{p1} & c_{p2} & \ldots\ldots\ldots & c_{pp} & L_p \\ L_1 & L_2 & \ldots\ldots\ldots & L_p & 0 \end{vmatrix} \div | c_{ij} | \quad \ldots \quad (2\cdot5)$$

where $L_i = (a_i - a_i')$ $(i = 1, 2, \ldots..p)$. ... (2·51)

Now if L_i goes over to M_i, then clearly

$$M_i = \lambda_{i1}\ L_1 + \lambda_{i2}L_2 + \ldots\ldots\lambda_{ip}L_p \quad \ldots \quad (2\cdot6)$$

Hence it is easy to see that

$$\begin{Vmatrix} d_{11} & d_{12} & \ldots & \ldots & d_{1p} & M_1 \\ d_{21} & d_{22} & \ldots & \ldots & d_{2p} & M_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ d_{p1} & d_{p2} & \ldots & \ldots & d_{pp} & M_p \\ M_1 & M_2 & \ldots & \ldots & M_p & 0 \end{Vmatrix} = S. \begin{Vmatrix} c_{11} & c_{12} & \ldots & \ldots & c_{1p} & L_1 \\ c_{21} & c_{22} & \ldots & \ldots & c_{2p} & L_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{p1} & c_{p2} & \ldots & \ldots & c_{pp} & L_p \\ L_1 & L_2 & \ldots & \ldots & L_p & 0 \end{Vmatrix} . \bar{S} \quad (2.7)$$

where S denotes the matrix

$$\begin{Vmatrix} \lambda_{11} & \lambda_{12} & \dots & \dots & \lambda_{1p} & 0 \\ \lambda_{21} & \lambda_{22} & \dots & \dots & \lambda_{2p} & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \lambda_{p1} & \lambda_{p2} & \dots & \dots & \lambda_{pp} & 0 \\ 0 & 0 & \dots & \dots & 0 & 1 \end{Vmatrix} \quad \dots (2\cdot71)$$

and S the matrix obtained from this by interchanging the rows and columns. Hence there exists the corresponding determinantal relation

$$\begin{vmatrix} d_{11} & d_{12} & \dots & \dots & d_{1p} & M_1 \\ d_{21} & d_{22} & \dots & \dots & d_{2p} & M_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{p1} & d_{p2} & \dots & \dots & d_{pp} & M_p \\ M_1 & M_2 & \dots & \dots & M_p & 0 \end{vmatrix} = \begin{vmatrix} c_{11} & c_{12} & \dots & \dots & c_{1p} & L_1 \\ c_{21} & c_{22} & \dots & \dots & c_{2p} & L_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_{p1} & c_{p2} & \dots & \dots & c_{pp} & L_p \\ L_1 & L_2 & \dots & \dots & L_p & 0 \end{vmatrix} \times |\lambda_{ij}|^2 \quad \dots (2\cdot8)$$

Remembering the value of D^2 given in (2·5), its invariance follows from the relations (2·4) and (2·8).

It is worthwhile to note that if a linear transformation is performed on all the variates of a multivariate normally distributed population, then the density factor in the joint distribution of the sample readings remains unaltered in form.

This density factor is

$$e^{-\frac{n}{2}\sum_{j=1}^{p}\sum_{i=1}^{p}\alpha^{ij}\{a_{ij}+(a_i-\alpha_i)(a_j-\alpha_j)\}} \quad \dots (2\cdot9)$$

$$\text{Now } \sum_{j=1}^{p}\sum_{i=1}^{p}\alpha^{ij}(a_i-\alpha_i)(a_j-\alpha_j) = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \dots & \dots & \alpha_{1p} & a_1-\alpha_1 \\ \alpha_{21} & \alpha_{22} & \dots & \dots & \alpha_{2p} & a_2-\alpha_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{p1} & \alpha_{p2} & \dots & \dots & \alpha_{pp} & a_p-\alpha_p \\ a_1-\alpha_1 & a_2-\alpha_2 & \dots & \dots & a_p-\alpha_p & 0 \end{vmatrix} \div -|\alpha_{ij}|$$

Its invariance is established exactly in the same way as that of D^2.

Next let us consider $\sum_{j=1}^{p}\sum_{i=1}^{p}\alpha^{ij}a_{ij}$. To prove its invariance we resort to the following consideration :—

From (2·1) it is clear that

$\|\beta_{ij}+k\,b_{ij}\|$ transforms to $S_* \|\alpha_{ij}+k\alpha_{ij}\| \bar{S}_*$ and $\|\beta_{ij}\|$ transforms to $S_* \|\alpha_{ij}\| \bar{S}_*$ where

S_e denotes the transformation matrix $\|\lambda_{ij}\|$ and $\overline{S}_e$ the matrix obtained from it by interchanging the rows and columns, while $\|\beta_{ij}\|$ and $\|b_{ij}\|$ are the transforms of $\|\alpha_{ij}\|$ and $\|a_{ij}\|$.

Hence clearly
$$\frac{|\, \alpha_{ij} + k a_{ij} \,|}{|\, \alpha_{ij} \,|}$$

is invariant. Since k is arbitrary, if we expand the above in powers of k, the coefficient of every power of k is invariant. But the coefficient of the first power of k is exactly

$$\sum_{j=1}^{p} \sum_{i=1}^{p} \alpha^{ij} a_{ij}$$

Hence its invariance.

Taking this together with the already established invariance of

$$\sum_{j=1}^{p} \sum_{i=1}^{p} \alpha^{ij} (a_i - \alpha_i)\ (a_j - \alpha_j)$$

the invariance of the density factor (2·9) follows.

§3. The first reduction of the distribution problem.

We know that the joint distribution of the sample readings x_{ik}, $x'_{ik'}$ ($i=1, 2, \ldots\ldots p$ $k=1, 2, \ldots\ldots n$, $k'=1, 2, \ldots\ldots n'$) is

$$\text{const. } e^{-\frac{1}{2}\sum_{j=1}^{p}\sum_{i=1}^{p} \alpha^{ij}\{n(a_i-\alpha_i)(a_j-\alpha_j)+n'(a'_i-\alpha'_i)(a'_j-\alpha'_j)+(na_{ij}+n'a'_{ij})\}} \times \Pi dx_{ik}\ \Pi dx'_{ik'} \qquad (3\cdot1)$$

where $\Pi\ dx_{ik}$ stands for $dx_{11}\ dx_{12}\ dx_{13} \ldots\ldots dx_{pn}$ and a similar notation holdsfor $\Pi\ dx'_{ik'}$.

Sample Σ can be represented in the usual Fisherian space S_n of n dimensions, by the points with coordinates

$$(x_{i1}, x_{i2}, \ldots\ldots\ldots x_{in}) \qquad i=1, 2, \ldots\ldots\ldots p \qquad \ldots\ (3\cdot11)$$

or what is the same thing, by p vectors x_i joining the points to the origin. We take another space $S'_{n'}$ of n' dimensions absolutely orthogonal to the former space, and represent in it the sample Σ' by p other similar vectors. Let

$$y_{ik} = x_{ik} - a_i,\ y'_{ik'} = x'_{ik'} - a'_i \qquad \ldots\ (3\cdot12)$$

where $i=1, 2, \ldots\ldots\ldots p$, $\quad k=1, 2, \ldots\ldots\ldots n$, $\quad k'=1, 2, \ldots\ldots\ldots n'$.

Let y_i, y'_i denote the vectors, with components $(y_{i1}, y_{i2}, \ldots\ldots y_{in})$ and $(y'_{i1}, y'_{i2} \ldots\ldots y'_{in'})$ lying in the space S_n and $S'_{n'}$ respectively. Then the vectors y_i lie in a flat S_{n-1} of $n-1$ dimensions, perpendicular to the equiangular line in S_n . Similar considerations apply to the vectors y'_i.

Let O be the origin of coordinates and let M_i be the point on the equiangular line in S_n such that $OM_i = \frac{1}{\sqrt{n}}$ times the projection of x_i on the equiangular line. Then $OM_i = a_i$. In the same way we can find M'_i on the other equiangular line such that $OM_i' = a_i'$. Also if $y_i \cdot y_j$ is the scalar product of the vectors y_i and y_j then clearly

$$a_{ij} = (y_i \cdot y_j)/n, \qquad a'_{ij} = (y'_i \cdot y'_j)/n'. \qquad \dots \quad (3.13)$$

Let us now take a new set of $(n+n')p$ variables a_i, a_i', z_{ik}, $z'_{ik'}$ $(i=1, 2, \dots\dots\dots p;$ $k=1, 2, \dots\dots n-1$, $k'=1, 2 \dots\dots n'-1)$ such that

$$\text{(i)} \quad a_i = \frac{1}{n} \sum_{k=1}^{n} x_{ik} \quad (i=1, 2, \dots\dots p)$$

$$\text{(ii)} \quad a'_i = \frac{1}{n'} \sum_{k'=1}^{n'} x'_{ik'} \quad (i=1, 2, \dots\dots p)$$

(iii) z_{ik} $(k=1, 2, \dots\dots n-1)$ are the components of y_i along any $(n-1)$ mutually orthogonal lines in S_{n-1}. Then z_{ik} is naturally a linear function of

$$x_{i1}, x_{i2} \dots\dots\dots\dots\dots x_{in}$$

(iv) Similar considerations apply to $z'_{ik'}$.

The distribution (3·1) can now be written in the form

$$\text{const.} \times e^{-\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} a^{ij} \{n\,(a_i - \alpha_i)\,(a_j - \alpha_j) + n'\,(a'_i - \alpha'_i)\,(a'_j - \alpha'_j) + (n\,a_{ij} + n' a'_{ij})\}}$$

$$\times \prod_{i=1}^{p} da_i \prod_{i=1}^{p} da'_i \; \Pi dz_{ik} \; \Pi dz'_{ik'}. \qquad \dots \quad (3·2)$$

It should be noted that a_{ij}'s are expressible purely in terms of z_{ik}'s and a'_{ij}'s are expressible purely in terms of $z'_{ik'}$'s.

Next we introduce the new variables $a_i - a'_i$ and $a_i + a'_i$ in the place of a_i and a'_i and integrating for $a_i + a'_i$ we get the distribution in the form

$$\text{const.} \times e^{-\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} a^{ij} \left[\frac{\bar{n}}{2} \{(a_i - a_i') - (\alpha_i - \alpha_i')\} \{(a_j - a_j') - (\alpha_j - \alpha_j')\} + N c_{ij} \right]}$$

$$\times \prod_{i=1}^{p} d(a_i - a'_i) \; \Pi z_{ik} \; \Pi z'_{ik'}. \qquad \dots \quad (3·3)$$

where c_{ij} is given by (1·1) and

$$\frac{2}{\bar{n}} = \frac{1}{n} + \frac{1}{n'}. \qquad N = n + n' \qquad \dots \quad (3·31)$$

In the plane of the equiangular lines of S_n and $S'_{n'}$, let R_i be the point, whose projections on the equiangular lines coincide with M_i and M'_i. Then if Q_i is the projection of R_i on OY, the external bisector of the equiangular lines,

$$OQ_i = \frac{1}{\sqrt{2}} (a_i - a'_i) \qquad (i = 1, 2, \ldots p) \qquad \ldots \quad (3\cdot4)$$

Let the vector v_i be the resultant of the vectors y_i and y'_i. Then it is easily seen that

$$Nc_{ij} = v_i \cdot v_j \qquad \ldots \quad (3\cdot5)$$

where the dot denotes the scalar product.

We may note that the spaces S_{n-1}, $S'_{n'-1}$ containing the vectors y_i, y'_i respectively $(i = 1, 2, \ldots\ldots p)$ are orthogonal to one another, as also to OY. Hence the new vectors v_i are also orthogonal to OY, and lie in the space $S_{n+n'-2}$, which comprises both S_{n-1} and $S'_{n'-1}$.

Let t_i be the resultant of the vectors v_i and the vector $\bar{n}^{\frac{1}{2}}$ OQ_i, where $\bar{n}$ is given by (3·31).

Then

$$t_i \cdot t_j = \frac{\bar{n}}{2} (a_i - a'_i)(a_j - a'_j) + (na_{ij} + n'a'_{ij})$$

$$= \frac{\bar{n}}{2} (a_i - a'_i)(a_j - a'_j) + Nc_{ij} = g_{ij} \text{ (say)} \qquad \ldots \quad (3\cdot6)$$

The distribution (3·3) now takes the form

$$e^{-\frac{1}{2}p\Delta^2 - \frac{1}{2} \sum_{j=1}^{p} \sum_{i=1}^{p} \alpha^{ij} \{g_{ij} - \bar{n} (a_i - a'_i)(a_j - a'_j)\}}$$

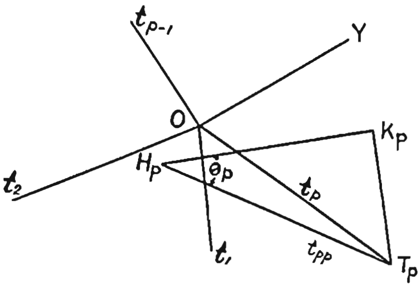
$$\prod_{i=1}^{p} d(a_i - a'_i) \, \Pi dz_{ik} \, \Pi dz'_{ik'} \qquad \ldots \quad (3\cdot7)$$

§4. Further transformation of the volume element

Let T_j denote the extremity of the vector t_j. Take now the rectangular coordinates t_{ij} $(i = 1, 2, \ldots j, j = 1, 2, \ldots p)$ for the system of points T_j, developed earlier by the authors [11]. They are really coordinates of the points relative to the orthogonal system of axes OZ_1, OZ_2, …… OZ_p, where OZ_1 lies along t_1, OZ_2 is perpendicular to t_1, in the plane (t_1, t_2), OZ_3 is perpendicular to the plane (t_1, t_2) in the hyperplane (t_1, t_2, t_3) and so on. Finally OZ_p is perpendicular to the space $(t_1, t_2, \ldots\ldots t_{p-1})$ and lies in the space $(t_1, t_2, \ldots t_p)$

Let θ_1 be the angle between t_1 and OY, θ_2 the angle between the plane (OY, t_1) and the plane (t_1, t_2), θ_3 the angle between the hyperplane (OY, t_1, t_2) and the hyperplane (t_1, t_2, t_3) Finally let θ_p be the angle between the spaces $(OY, t_1, t_2, \ldots\ldots t_{p-1})$ and $(t_1, t_2, \ldots\ldots t_p)$.

Let H_p be the foot of the perpendicular from T_p on the space $(t_1, t_2, \ldots\ldots t_{p-1})$ and K_p the foot of the perpendicular on the space $(OY, t_1, t_2, \ldots\ldots t_{p-1})$. Then the angle $H_p K_p T_p$ is a right angle, and the angle $T_p K_p H_p$ is θ_p, $H_p K_p$ being perpendicular to the space $(t_1, t_2, \ldots\ldots t_{p-1})$. Thus $H_p T_p = t_{pp}$ and $K_p T_p = t_{pp} \sin \theta_p$.



Let the vectors $t_1, t_2 \ldots\ldots t_{p-1}$ remain fixed in position. Consider now the vector t_p. Then with fixed values of $t_{1p}, t_{2p}, \ldots\ldots t_{pp}, \theta_p$ the point T_p describes a sphere of radius $t_{pp} \sin \theta_p$, and is constrained to lie in a space of $(n+n'-1)-p$ dimensions orthogonal to the space $(OY, t_1, t_2, \ldots . t_{p-1})$ and immersed in the space comprising S_{n-1}, $S'_{n'-1}$ and OY. If now we give a slight latitude to the coordinates t_{ip} and θ_p so that they lie between t_{ip} and $t_{ip}+dt_{ip}$, and θ_p and $\theta_p+d\theta_p$ respectively $(i=1, 2, \ldots\ldots p)$, then the point T_p describes a volume element proportional to

$$(t_{pp} \sin \theta_p)^{n+n'-2-p}\, dt_{pp}.\ t_{pp}\, d\theta_p.\ dt_{p-1,\,p}\, dt_{p-2,\,p} \ldots\ldots dt_{1,\,p} \qquad \ldots \quad (4\cdot1)$$

If we successively free the vectors $t_{p-1}, t_{p-2}, \ldots\ldots t_1$ then it is easy to see that the complete volume element in (3.7), is proportional to

$$\prod_{i=1}^{p} \{ (t_{ii} \sin \theta_i)^{n+n'-2-i}\, dt_{ii}.\ t_{ii}\, d\theta_i.\ dt_{i-1,\,i}\ \ dt_{i-2,\,i} \ldots\ldots dt_{1i} \}$$

$$= t_{pp}^{\,N-p-1}\ t_{p-1,p-1}^{\,N-p} \ldots\ldots t_{11}^{\,N-2}\ \Pi\, dt_{ij}$$

$$\times\ (\sin\theta_p)^{N-p-2} (\sin \theta_{p-1})^{N-p-1} \ldots\ldots (\sin \theta_1)^{N-3} \prod_{i=1}^{p} d\theta_i \qquad \ldots \quad (4\cdot2)$$

Hence the distribution (3·7) now assumes the form

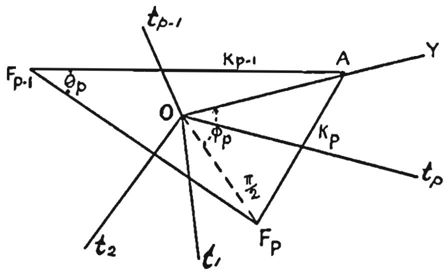
$$\text{Const} \times e^{-\frac{1}{2}\bar{n}\, p\Delta^2 - \frac{1}{2} \sum_{j=1}^{p} \sum_{i=1}^{p} \alpha^{ij} \{ g_{ij} - \bar{n}\, (a_i - a'_i)(a_j - a'_j) \}}$$

$$\times \prod_{i=1}^{p} (t_{ii})^{N-i-1}\ \Pi\, dt_{ij} \prod_{i=1}^{p} (\sin \theta_i)^{N-i-2} \prod_{i=1}^{p} d\theta_i \qquad \ldots \quad (4\cdot3)$$

It is worth noting that g_{ij}'s are expressible purely in terms of t_{ij}'s while $(a_i - a'_i)$ are expressible in terms of t_{ij}'s and θ_i's combined.

§ 5. Geometrical interpretation of D².

Let OA be the unit vector along OY. We shall denote it by i. Let F_i denote the foot of the perpendicular from A to the space $(t_1, t_2, \ldots\ldots t_i)$. Let the length AF_i be



denoted by k_i. Then from geometry it is clear that the angle $AF_{i-1}F_i$ is equal to θ_i (the angle θ_p is shown in Fig. (2), and the angle $AF_i F_{i-1}$ is a right angle. Hence

$$\sin\theta_i = \frac{k_i}{k_{i-1}} \quad (i=1, 2, \ldots\ldots p) \qquad \ldots \quad (5\cdot1)$$

Of course $k_0=1$ being the length OA itself.

Let us now consider the length k_p. We shall show that it is very closely connected with D².

$$\text{Now } k_p = \frac{\text{Vol}\,(i, t_1, t_2, \ldots\ldots t_p)}{\text{Vol}\,(t_1, t_2, \ldots\ldots t_p)} \qquad \ldots \quad (5\cdot2)$$

But Vol $(i, t_1, t_2, \ldots\ldots t_p)$ is the same as the volume formed by the unit vector OA, and the projections of the vectors $t_1, t_2, \ldots\ldots t_p$ on the space perpendicular to OA, which we have earlier called $S_{n+n'-2}$. But these projections, from the way in which $t_1, t_2, \ldots\ldots t_p$ have been derived, are at once seen to be $v_1, v_2, \ldots\ldots v_p$. Hence the numerator in (5.2) is numerically equal to Vol $(v_1, v_2, \ldots\ldots v_p)$.

$$\therefore \quad k_p^2 = \frac{|\, v_i \cdot v_j \,|}{|\, t_i \cdot t_j \,|}$$

$$= \frac{|\, Nc_{ij} \,|}{|\, Nc_{ij} + \frac{\bar{n}}{2}(a_i - a_i')(a_j - a_j') \,|} \quad \text{from (3·5) and (3·6)}$$

$$= \frac{|\, Nc_{ij} \,|}{|\, Nc_{ij} \,| + \frac{\bar{n}}{2} N^{p-1} C_{ij}(a_i - a_i')(a_j - a_j') \,|}$$

C_{ij} denoting the minor of c_{ij} in $|\, c_{ij} \,|$.

Hence from the definition of D² given in (1.2), it follows that

$$k_p^2 = \frac{1}{1 + \frac{p\bar{n}}{2N} D^2} \qquad \ldots (5\cdot3)$$

Denoting by φ_p the angle AOP_p, *i.e.*, the angle between the unit vector $\mathbf{i}_*$ and the space $(\mathbf{t}_1, \mathbf{t}_2, \ldots\ldots \mathbf{t}_p)$ we have, since $k_p = \sin \varphi_p$,

$$D^2 = \frac{2N}{p\bar{n}} \cot^2 \varphi_p \qquad \ldots \quad (5\cdot4)$$

If a statistic similar to D^2 were constructed for the first i variates only, then we could denote it by D_i^2; we should than have had

$$k_i^2 = \frac{1}{1 + \dfrac{i\bar{n}}{2N} D_i^2} \qquad \ldots \quad (5\cdot5)$$

nd

$$D_i^2 = \frac{2N}{i\bar{n}} \cot^2 \varphi_i \qquad \ldots \quad (5\cdot6)$$

It should be noted that consistently with this notation D^2 should be denoted by D_p^2.

§ 6. Transformation of the volume element resumed.

From (5.1) it follows that

$$k_1 = \sin \theta_1, \; k_2 = \sin \theta_1 \sin \theta_2, \;\ldots\ldots\ldots\ldots\; k_p = \sin \theta_1 \sin \theta_2 \ldots \sin \theta_p \qquad \ldots \quad (6\cdot1)$$

The Jacobian

$$\frac{\partial(\theta_1, \theta_2, \ldots \theta_p)}{\partial(k_1, k_2 \ldots k_p)} = \frac{1}{(\sin\theta_1)^{p-1}(\sin\theta_2)^{p-2} \ldots\ldots (\sin\theta_{p-2})^2 (\sin\theta_{p-1}) (\cos\theta_1 \cos\theta_2 \ldots\ldots \cos\theta_p)}$$

$$= \frac{k_1 \, k_2 \ldots\ldots k_{p-1}}{\sqrt{\{(1-k_1^2)(k_1^2-k_2^2)\ldots\ldots(k_{p-1}^2-k_p^2)\}}}$$

$$\times \frac{1}{(\sin\theta_1)^{p-1}(\sin\theta_2)^{p-2}\ldots\ldots(\sin\theta_{p-1})} \qquad \ldots \quad (6\cdot2)$$

The volume element (4.2) is now transformed to

$$\prod_{i=1}^{p} (t_{ii})^{N-i-1} \, \Pi \, dt_{ij} . (\sin\theta_1 \sin\theta_2 \ldots\ldots \sin\theta_p)^{N-2-p}$$

$$\times \frac{k_1 \, k_2 \ldots\ldots k_p}{\sqrt{\{(1-k_1^2)(k_1^2-k_2^2)\ldots\ldots(k_{p-1}^2-k_p^2)\}}} \prod_{i=1}^{p} dk_i$$

$$= \prod_{i=1}^{p} (t_{ii})^{N-i-1} \, \Pi \, dt_{ij}$$

$$\times \frac{k_p^{N-p-2} \, k_1 \, k_2 \ldots\ldots k_{p-1}}{\sqrt{\{(1-k_1^2)(k_1^2-k_2^2)\ldots\ldots(k_{p-1}^2-k_p^2)\}}} \prod_{i=1}^{p} dk_i \qquad \ldots \quad (6\cdot3)$$

Making the substitution $\quad k_i^2 = z_i, \qquad \ldots \quad (6\cdot4)$

the volume element becomes

$$\prod_{i=1}^{p} (t_{ii})^{N-i-1} \, \Pi \, dt_{ij} . \, z^{(N-p-2)/2} dz_1 \, dz_2 \ldots\ldots dz_p / \sqrt{\{(1-z_1)(z_1-z_2)\ldots\ldots(z_{p-1}-z_p)\}} \ldots \quad (6\cdot5)$$

Since $\sin^2 \theta_i = z_i/z_{i-1}$,

$$1 \geq z_1 \geq z_2 \geq z_3 \;\ldots\ldots\ldots\; \geq z_{p-1} \geq z_p \geq 0$$

Hence if we wish to integrate out in (6.5) the variables $z_1, z_2, \ldots\ldots z_{p-1}$ in this order then z_1 varies from z_2 to 1, z_2 varies from z_3 to 1, and so on till we come to z_{p-1} which varies from z_p to 1. Of course z_p itself varies from 0 to 1.

§ 7. The distribution of D² for the case $\Delta^2 = 0$

In this case the two populations are identical which we shall rigorously prove in section 8, so that $\alpha_j = \alpha'_j$ $(j = 1, 2, \ldots\ldots p)$. Remembering (6·5) the distribution (4·3) now becomes

$$\text{Const} \times e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha^{ij} g_{ij}} \prod_{i=1}^{p} (t_{ii})^{N-i-1} \quad \Pi\, dt_{ij}$$

$$\times \frac{z_p^{(N-p-3)/2}\, dz_1\, dz_2, \ldots\ldots dz_{p-1}\, dz_p}{\sqrt{\{(1-z_1)(z_1-z_2)\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (7·1)$$

Since g_{ij}'s are expressible purely in terms of t_{ij}'s, we can integrate out for t_{ij}'s and obtain the joint distribution of $z_1, z_2 \ldots\ldots z_p$ in the form

$$\text{Const} \times z_p^{(N-p-3)/2}\, dz_p \frac{dz_1\, dz_2 \ldots\ldots dz_{p-1}}{\sqrt{\{(1-z_1)(z_1-z_2)\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (7·2)$$

where the ranges of variation of $z_1, z_2, \ldots\ldots z_p$ are as mentioned after the formula (6·5).

Put $z_1 = a_1 x_1 + b_1$, adjusting a_1 and b_1 so as to make $x_1 = 0$ when $z_1 = z_2$ and $x_1 = 1$ when $z_1 = 1$. Then clearly $z_1 = (1-z_2)\, x_1 + z_2$.

$$\therefore \int_{z_2}^{1} \frac{dz_1}{\sqrt{\{(1-z_1)(z_1-z_2)\}}} = (1-z_2)$$

In the same way, after integrating out for z_1, let us set

$$z_2 = (1 - z_3)\, x_2 + z_3$$

$$\int_{z_3}^{1} \frac{dz_2}{\sqrt{(z_2 - z_3)}} = \int_{0}^{1} \frac{(1-z_3)\, dx_2}{\sqrt{(1-z_3)}\sqrt{x_2}} = \text{const} \sqrt{(1-z_3)}$$

After integrating out for z_2, the portion of the integral involving z_3 i

$$\int_{z_4}^{1} \frac{\sqrt{(1 - z_3)}\, dz_3}{\sqrt{(z_3 - z_4)}}$$

Setting now $z_3 = (1 - z_4)\, z_3 + x_4$ this integral reduces to

$$\int_{0}^{1} (1 - z_4) \frac{\sqrt{(1 - x_3)}\, dx_3}{\sqrt{x_3}} = \text{const} \times (1 - z_4)$$

Proceeding in this way if we successively integrate out for $z_1, z_2, \ldots\ldots z_{p-1}$, we obtain the distribution of z_p in the form

$$\text{Const} \times z_p^{(N-p-1)/2} (1-z_p)^{(p-2)/2} \quad dz_p$$

Remembering that

$$z_p = k_p^2 = \frac{1}{1+\frac{p\bar{n}}{2N} D^2} \qquad \ldots \quad (7\cdot3)$$

we have the distribution of D^2 in the form

$$\text{Const} \times \frac{(D^2)^{(p-2)/2}}{\left\{1+\frac{p\bar{n}}{2N}D^2\right\}^{(N-1)/2}} dD^2 \qquad \ldots \quad (7\cdot4)$$

Since the total frequency is unity, the value of the constant in (7.4) is easily seen to be

$$\left(\frac{p\bar{n}}{2N}\right)^{p/2} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p-1}{2}\right)} \qquad \ldots \quad (7\cdot5)$$

This is in agreement with the distribution implied in Fisher's paper[4] already referred to.

§8. CONSTRUCTION OF A SPECIAL TRANSFORMATION.

Next we have to turn our attention to the problem of finding the distribution of D^2 for the case $\Delta^2 \neq 0$. It has been shown already (cf §2), that the form of the multivariate normal frequency distribution, is invariant under a linear transformation. The same holds for and D^2 and Δ^2. We can thus at the very outset simplify our problem, by performing on our variates, a suitable linear transformation, without in any way altering the distribution results we seek.

We shall now show, that a linear transformation can be constructed, so that in the two populations Π and Π' (for which the variances and covariances of the original variates were identical), the covariances of the new variates are all zero, variances all unity, and all the mean differences except one (that of the first variate, say) are zero. Since Δ^2 is invariant the square of the non-vanishing mean difference is identical with $p\Delta^2$ (cf 1·3).

The linear transformations form a group, so that the resultant of two linear transformations, is again a linear transformation. Hence to achieve our purpose we can proceed with linear transformations in stages.

We shall first show that their exists a linear transformation of the variates, so that the new variates have (in the population) vanishing covariances and unit variances. Let us take a space of p dimensions, and in it coinitial vectors

$$\omega_1, \omega_2, \ldots\ldots\ \ldots\ldots \omega_p$$

such that $\omega_i . \omega_j = a_{ij}$, the dot denoting the scalar product, and $i, j = 1, 2, \ldots\ldots\ldots\ldots p$.

It is of course taken for granted that none of the vectors $\omega_1, \omega_2, \ldots\ldots\ldots \omega_p$ is a linear combination of the others, so that statistically speaking no single variate is (for all possible individuals) predictable solely on the basis of the others. In this space let us take p mutually orthogonal unit vectors

$$i_1, i_2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots i_p$$

so that we can express each of these vectors as a linear combination of $\omega_1, \omega_2, \ldots\ldots\ldots \omega_p$.

Let

$$i_k = \lambda_{k1}\, \omega_1 + \lambda_{k2}\, \omega_2 + \ldots\ldots \ldots\ldots \lambda_{kp}\, \omega_p \qquad \ldots \quad (8{\cdot}1)$$

Then clearly the transformation with matrix

$$\| \lambda_{kl} \|$$

serves our purpose, *i.e.* the new variates after transformation are uncorrelated in the population, and have unit variances. Before proceeding to the second stage of our transformation we shall pause to deduce a useful fundamental result.

The means α_k, α'_k $(k=1, 2, \ldots\ldots p)$ of the two populations of course suffer the same transformation. Let them become β_k, β'_k. Then from the invariance of Δ^2 it follows, that

$$p\,\Delta^2 = \sum_{k=1}^{p} (\beta_k - \beta'_k)^2 \qquad \ldots \quad (8{\cdot}2)$$

Hence Δ^2 can vanish when and only when $\beta_k = \beta'_k$ $(k=1, 2, \ldots\ldots\ldots p)$. But $\beta_k - \beta'_k$ is connected with $\alpha_k - \alpha'_k$ by the linear transformation

$$\beta_k - \beta'_k = \sum_{l=1}^{p} \lambda_{kl}\,(\alpha_l - \alpha'_l) \qquad (k = 1, 2, \ldots\ldots\ldots p) \qquad \ldots \quad (8{\cdot}3)$$

with matrix $\|\lambda_{kl}\|$ of rank p. Hence the necessary and sufficient condition for the vanishing of the set $(\beta_k - \beta'_k)$, $(k=1, 2, \ldots\ldots\ldots p)$ is the vanishing of the set $\alpha_k - \alpha'_k$ $(k=1, 2, \ldots\ldots p)$.

Hence Δ^2 for the two populations II and II′ is zero when and only when the population means are identical, *i.e.* $\alpha_k = \alpha'_k$, $(k=1, 2, \ldots\ldots p)$. It is evident that what is true of Δ^2 is also true of D^2, the sample means now being considered.

We now proceed to the second stage of our transformation. Now let us apply to our variates an orthogonal transformation, possessing the matrix $\|\nu_{ij}\|$ where

$$\nu_{1j} = \frac{\beta_j - \beta'_j}{\Delta\sqrt{p}} \quad (j = 1, 2, \ldots\ldots\ldots\ldots\ldots p) \qquad \ldots \quad (8{\cdot}4)$$

and the other elements are arbitrary, excepting that the whole set satisfies the relations

$$\sum_{k=1}^{p} \nu_{ik}\, \nu_{jk} = \delta_{ij}, \qquad \delta_{ij} = 0 \text{ when } i \neq j, \qquad \delta_{ii} = 1 \qquad \ldots \quad (8{\cdot}5)$$

The population mean differences then suffer the same linear transformation. The square of the mean differences for the first variate becomes $p\Delta^2$, while the other mean differences vanish. It should be noted that the variances and covariances are unaltered, all variances remaining as before unity, and all covariances zero.

The resultant of the two linear transformations constructed just before *viz.*, of those with matrices $\|\lambda_{ij}\|$ and $\|\nu_{ij}\|$ gives us the desired transformation.

§9. SIMPLIFICATION OF THE DISTRIBUTION PROBLEM OF D^2, WHEN $\Delta^2 \neq 0$ BY THE USE OF OUR SPECIAL TRANSFORMATIONS.

From the considerations of the last paragraph it follows that in our populations Π_1 and Π_2, we can without any loss of generality consider

$$\left.\begin{aligned} a_{ij} &= 0 \text{ if } i \neq j, \quad a_{ii} = 1 \\ (a_1 - a'_1)^2 &= p\Delta^2 \\ (a_i - a'_i) &= 0 \text{ for } i = 2, 3, \ldots\ldots p \end{aligned}\right\} \quad \ldots (9\cdot1)$$

$$\therefore \quad a^{ij} = 0 \text{ if } i \neq j, \quad a^{ii} = 1 \quad \ldots (9\cdot2)$$

We may consider Δ to be the positive root of Δ^2 and $\sqrt{p}$, the positive root of p. Then by suitably naming the populations as first end second, we can write

$$a_1 - a'_1 = \Delta\sqrt{p} \quad \ldots (9\cdot21)$$

The distribution (4·3) can then be written as

$$\text{const} \times e^{-\frac{1}{2}\bar{n}p\Delta^2 - \frac{1}{2}\left(\sum_{i=1}^{p} g_{ii}\right) - \frac{1}{2}\bar{n}\sqrt{p}\,\Delta(a_1 - a'_1)}$$

$$\times \prod_{i=1}^{p} (t_{ii})^{(N-i-1)} \Pi\, d\, t_{ij} \prod_{i=1}^{p} (\sin\theta_i)^{N-i-2} \prod_{i=1}^{p} (d\,\theta_i)$$

where from (3·6).

$$g_{ii} = \frac{\bar{n}}{2}(a_i - a'_i)^2 + Nc_{ii}$$

Remembering the manner in which the chain of rectangular coordinates was constructed we have

$$\left.\begin{aligned} &g_{11} = t_{11}^2 \\ &g_{22} = t_{12}^2 + t_{22}^2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &g_{pp} = t_{1p}^2 + t_{2p}^2 + \ldots\ldots t_{pp}^2 \end{aligned}\right\} \quad \ldots (9\cdot31)$$

Also from geometry

$$\sqrt{g_{11}}.\cos\theta = \sqrt{\bar{n}/2}.\ (a_1 - a'_1) \quad \ldots (9\cdot32)$$

$$i.e. \quad \sqrt{\bar{n}/2}.\ (a_1 - a_1') = t_{11}\cos\theta \quad \ldots (9\cdot33)$$

Making these substitutions in (9·3) we can forthwith integrate out for all the t_{ii}'s and t_{ij}'s except t_{11}. We thus get our distribution in the form

$$\text{const} \times e^{-\frac{1}{2}(c^2\Delta^2 - 2c\Delta t_{11}\cos\theta_1 + t_{11}^2)} \times (t_{11})^{N-2}\, dt_{11}$$

$$\times \prod_{i=1}^{p} (\sin\theta_i)^{N-i-2} \prod_{i=1}^{p} d\cdot\theta_i \quad \ldots (9\cdot4)$$

where $c = \sqrt{(p\bar{n}/2)}$

Here we note that t_{ii} varies from o to ∞ and $\theta_1, \theta_2, \ldots\ldots\ldots\ldots \theta_p$ each vary from o to π.

If now we want to let θ_i $(i=1, 2, \ldots\ldots p)$ vary only from o to $\pi/2$, then the distribution will assume the form

$$\text{const} \times e^{-\frac{1}{2}(t^2_{11}+c^2\Delta^2)} \cosh(c\Delta t_{11}\cos\theta_1)\times(t_{11})^{N-2}\, dt_{11}$$

$$\times \prod_{i=1}^{p} (\sin\theta_i)^{N-i-2} \prod_{i=1}^{p} d\theta_i \qquad \ldots \quad (9\cdot5)$$

Making now as in §6, the substitutions

$k_1=\sin\theta_1 \quad k_2=\sin\theta_1\sin\theta_2, \ \ldots k_p=\sin\theta_1\sin\theta_2\ldots\ldots\sin\theta_p,\ k_i^2=z_i \quad (i=1, 2, \ldots\ldots p)$

we can write (9·5) in the form

$$\text{const} \times e^{-\frac{1}{2}(t^2_{11}+c^2\Delta^2)} \cosh\{c\,\Delta\, t_{11}\sqrt{(1-z_1^2)}\}\times t_{11}^{N-2}\, dt_{11}$$

$$\times \frac{z_p^{(N-p-3)/2}\, dz_1\, dz_2\ldots\ldots dz_p}{\sqrt{\{(1-z_1)(z_1-z_2)\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (9\cdot6)$$

It is worth noting that after θ_i's have been restricted to vary from o to $\pi/2$ there is a (1, 1) correspondence between the sets $(\theta_1, \theta_2 \ldots\ldots \theta_p)$, $(k_1, k_2, \ldots\ldots k_p)$ and $(z_1, z_2, \ldots\ldots z_p)$.

Also the formula (7·3) connecting z_p and D^2 holds, *i.e.*,

$$z_p = \frac{1}{1+\frac{p\bar{n}}{2N}D^2} = \frac{1}{1+\frac{c^2}{N}D^2}$$

The range of variation of $z_1, z_2, \ldots\ldots\ldots z_p$ is the same as remarked at the end of §6.

§10. Distribution of D² for the case $\Delta^2 \neq 0$.

Consider now the joint distribution of t_{11} and $z_1, z_2, \ldots\ldots\ldots z_p$ given by (9·6). In virtue of the relation (7·3), the problem of finding the distribution of D^2 is equivalent to that of finding the distribution of z_p. Remembering now the range of variation of t_{11}, $z_1, z_2 \ldots\ldots\ldots z_p$ we easily see that the distribution of z_p is given by

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} \times z_p^{(N-p-3)/2}\, dz_p$$

$$\times \int_0^\infty \int_{z_p}^1 \int_{z_{p-1}}^1 \ldots\ldots \int_{z_3}^1 \int_{z_2}^1 e^{-\frac{1}{2}t_{11}^2}\, t_{11}^{N-2} \cosh\{c\,\Delta\, t_{11}\sqrt{(1-z_1)}\}$$

$$\times \frac{dz_1\, dz_2\ldots\ldots\ldots\ldots\ldots\ldots dz_{p-1}\, dt_{11}}{\sqrt{\{(1-z_1)(z_1-z_2)\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (10\cdot1)$$

The order of integration meant is clearly indicated, *i.e.* we first integrate over z_1, then over z_2 and so on, till we come to z_{p-1}. We now integrate over z_{p-1} and then over t_{11}.

To integrate over z_1 we put

$$z_1 = (1-z_2)\sin^2\psi_1 + z_2 \qquad \ldots \quad (10\cdot11)$$

Then

$$\left.\begin{aligned} (1-z_1) &= (1-z_2)\cos^2\psi_1 \\ (z_1-z_2) &= (1-z_2)\sin^2\psi_1 \\ dz_1 &= 2(1-z_2)\sin\psi_1\cos\psi_1 d\psi_1 \end{aligned}\right\} \qquad \ldots \quad (10\cdot12)$$

It is clear that limits of ψ_i are from 0 to $\pi/2$.

The integral for z_1 is now transformed to

$$\text{const} \times \int_0^{\pi/2} \cosh \{c\,\Delta\, t_{11} (\cos \psi_1) \sqrt{(1-z_2)}\}\, d\psi_1 \qquad \ldots \quad (10{\cdot}13)$$

Using the well known relation[16]

$$I_\nu (z) = \frac{2\,(z/2)^\nu}{\Gamma(\nu+\frac{1}{2})\,\Gamma(\frac{1}{2})} \int_0^{\pi/2} \cosh (z \cos \theta) \sin^{2\nu} \theta \, d\theta \qquad \ldots \quad (10{\cdot}14)$$

which is valid for $R\,(\nu+\frac{1}{2})>0$, for the special case $\nu=0$ the integral (10·13) is seen to be

$$\text{const} \times I_0\, |c\,\Delta\, t_{11}\, \sqrt{(1-z_2)}| \qquad \ldots \quad (10{\cdot}15)$$

The distribution (10·1) now reduces to

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} \times z_p^{(N-p-2)/2}$$

$$\times \int_0^\infty \int_{z_p}^1 \int_{z_{p-1}}^1 \ldots\ldots \int_{z_4}^1 \int_{z_3}^1 e^{-\frac{1}{2}t_{11}^2}\; t_{11}^{N-2}\; I_0 \{(c\,\Delta\, t_{11}\, \sqrt{(1-z_2)}\}$$

$$\times \frac{dz_2\, dz_3 \ldots\ldots\ldots\ldots\ldots\ldots dz_{p-1}\, dt_{11}}{\sqrt{\{(z_2-z_3)(z_3-z_4)\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (10{\cdot}2)$$

Setting now $z_2 = (1-z_3)\cos^2\psi_2 + z_3$, the integral for z_2 i.e.

$$\text{const} \times \int_{z_3}^1 I_0 \{c\,\Delta\, t_{11}\, \sqrt{(1-z_2)}\}\, \frac{dz_2}{\sqrt{(z_2-z_3)}} \qquad \ldots \quad (10{\cdot}21)$$

now transforms to

$$\text{const} \times \int_0^{\pi/2} (1-z_3)^{1/2} \sin \psi_2\, I_0 \{c\,\Delta\, t_{11}\, \sqrt{(1-z_3)} \sin \psi_2\}\, d\psi_2 \qquad \ldots \quad (10{\cdot}22)$$

We now have the relation (which is a slightly modified form of Sonine's first integral)[17]

$$J_{\mu+\nu+1} (z) = \frac{z^{\nu+1}}{2^\nu\, \Gamma(\nu+1)} \int_0^{\pi/2} J_\mu (z \sin \theta) \sin^{\mu+1} \theta \cos^{2\nu+1} \theta \, d\theta \qquad (10{\cdot}23)$$

which is valid when both R (μ) and R (ν) exceed -1. Here there is no-restriction on the complex number z. We are however only interested in J-functions of pure imaginary arguments, *i.e.* in I-functions of real arguments. Hence remembering the relation $I_n(z) = i^{-n} J_n(iz)$ we shall write (10·23) in the form.

$$I_{\mu+\nu+1} (x) = \frac{x^{\nu+1}}{2^\nu\, \Gamma(\nu+1)} \int_0^{\pi/2} I_\mu (x \sin \theta) \sin^{\mu+1} \theta \cos^{2\nu+1} \theta \, d\theta \qquad \ldots \quad (10{\cdot}24)$$

Using now the relation (10·24) for the case $\mu = 0$, $\nu = -\frac{1}{2}$, the integral (10·22) reduces to

$$\text{const} \times (c\,\Delta\, l_{11})^{-\frac{1}{2}} (1-z_3)^{\frac{1}{4}}\, I_{\frac{1}{2}}\{c\,\Delta\, l_{11}\, \sqrt{(1-z_3)}\} \qquad \ldots \quad (10\cdot25)$$

The distribution (10·2) now reduces to

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-2)/2} \times \int_0^\infty \int_{z_p}^1 \int_{z_{p-1}}^1 \ldots\ldots \int_{z_4}^1 e^{-\frac{1}{2}l_{11}^2} l_{11}^{N-2}\, I_{\frac{1}{2}}\{c\,\Delta\, l_{11}\, \sqrt{(1-z_3)}\}$$

$$\times \frac{(c\,\Delta\, l_{11})^{-\frac{1}{2}} (1-z_3)^{1/4}\, dz_3\, dz_4 \ldots\ldots\ldots dz_{p-1}}{\sqrt{\{(z_3-z_4)\,(z_4-z_5)\ldots\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (10\cdot3)$$

Setting now

$$z_3 = (1-z_4)\cos^2 v_3 + z_4 \qquad \ldots \quad (10\cdot31)$$

the integral for z_3 ie

$$\int_{z_4}^1 \frac{(1-z_3)^{1/4}}{\sqrt{(z_3-z_4)}}\, I_{\frac{1}{2}}\{c\,\Delta\, l_{11}\, \sqrt{(1-z_3)}\}\, dz_3 \qquad \ldots \quad (10\cdot32)$$

reduces to

$$\text{const} \times \int_0^{\pi/2} (1-z_4)^{3/4} \sin^{3/2} v_3\, I_{\frac{1}{2}}\{c\,\Delta\, l_{11} \sin v_3\, \sqrt{(1-z_4)}\}\, d\,v_3 \qquad \ldots \quad (10\cdot33)$$

Using now (10·24) for the values $\mu = \frac{1}{2}$, $\nu = -\frac{1}{2}$ the integral (10·33) is seen to be equal o

$$\text{const} \times (c\,\Delta\, l_{11})^{-1} (1-z_4)^{\frac{1}{2}}\, I_1\{c\,\Delta\, l_{11}\, \sqrt{(1-z_4)}\} \qquad \ldots \quad (10\cdot34)$$

So that the distribution (10·3) becomes

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-2)/2}$$

$$\times \int_0^\infty \int_{z_p}^1 \int_{z_{p-1}}^1 \ldots\ldots \int_{z_5}^1 e^{-\frac{1}{2}l_{11}^2}\, l_{11}^{N-2}\, I_1\{c\,\Delta\, l_{11}\, \sqrt{(1-z_4)}\}$$

$$\times \frac{(c\,\Delta\, l_{11})^{-1} (1-z_4)^{\frac{1}{2}}\, dz_4\, dz_5 \ldots\ldots\ldots dz_{p-1}\, dl_{11}}{\sqrt{\{(z_4-z_5)\,(z_5-z_6)\ldots\ldots(z_{p-1}-z_p)\}}} \qquad \ldots \quad (10\cdot4)$$

Proceeding on in this way, and successively integrating out for $z_4, z_5, \ldots\ldots\ldots z_{p-1}$ we ultimately have the distribution of z_p in the form

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-2)/2}\, dz_p$$

$$\times \int_0^\infty . e^{-\frac{1}{2}l_{11}^2}\, l_{11}^{N-2}\, I_{\frac{p-3}{2}}\{c\,\Delta\, l_{11}\, \sqrt{(1-z_p)}\}.\ (c\,\Delta\, l_{11})^{-(p-3)/2} (1-z_p)^{(p-3)/4}\, dl_{11}$$

or after absorbing powers of c into the constant in the form

$$\text{const} \times \Delta^{-(p-2)/2} e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-1)/2}(1-z_p)^{(p-3)/2}\, dz_p$$

$$\times \int_0^\infty e^{-\frac{1}{2}t_{11}^2} t_{11}^{(2N-p-3)/2} I_{\frac{p-2}{2}}\{c\,\Delta\, t_{11}\sqrt{(1-z_p)}\}\, dt_{11} \qquad \dots \quad (10{\cdot}5)$$

It now remains only to integrate out for t_{11}

For this we have to remember the relation[18]

$$\int_0^\infty J_\nu(at)\, e^{-p^2t^2} t^{\mu-1}\, dt = \frac{\Gamma(\frac{1}{2}\mu+\frac{1}{2}\nu)\,(a/2p)^\nu}{2p^\mu\,\Gamma(\nu+1)}\ {}_1F_1\left(\tfrac{1}{2}\nu+\tfrac{1}{2}\mu,\ \nu+1,\ -\frac{a}{4p^2}\right)$$

$$\dots \quad (10{\cdot}51)$$

where

$${}_1F_1(\alpha, \rho, z) = 1 + \frac{\alpha}{1\,!\,\rho}z + \frac{\alpha(\alpha+1)}{2\,!\,\rho(\rho+1)}z^2 + \frac{\alpha(\alpha+1)(\alpha+2)}{3\,!\,\rho(\rho+1)(\rho+2)}z^3 + \dots \quad (10{\cdot}52)$$

Writing it in terms of the I function

$$\int_0^\infty I_\nu(at)\, e^{-p^2t^2}\, t^{\mu-1}\, dt$$

$$= \frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2}\mu)\,(a/2p)^\nu}{2p^\mu\,\Gamma(\nu+1)}\ {}_1F_1\left(\tfrac{1}{2}\nu+\tfrac{1}{2}\mu,\ \nu+1,\ \frac{a^2}{4p^2}\right) \qquad \dots \quad (10{\cdot}53)$$

Both the formulæ (10·51) and (10·53) are valid when

$$R(\mu+\nu) > 0 \qquad \dots \quad (10{\cdot}54)$$

Let us now reduce the integral in (10·5), by using (10·53), for the special case when

$$\nu = \frac{p-2}{2},\quad \mu = n + n' - \frac{p}{2},\quad p^2 = \tfrac{1}{2},\quad a = c\,\Delta\sqrt{(1-z_p)}$$

Since each of the two samples is at least of size one

$$R(\mu+\nu) = n + n' - 1 > 0$$

Hence finally the distribution of z_p comes in the form

$$\text{const} \times \Delta^{-(p-2)/2} e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-3)/2}(1-z_p)^{(p-3)/2}$$

$$\times [c\,\Delta\sqrt{(1-z_p)}]^{(p-2)/2}\ {}_1F_1\left(\frac{N-1}{2},\ \frac{p}{2},\ \tfrac{1}{2}c^2\Delta^2\,\overline{1-z_p}\right)$$

or after simplification and absorbing of Δ in the constant we get

$$\text{const} \times e^{-\frac{1}{2}c^2\Delta^2} z_p^{(N-p-3)/2}(1-z_p)^{(p-3)/2}$$

$$\times\ {}_1F_1\left(\frac{N-1}{2},\ \frac{p}{2},\ \tfrac{1}{2}c^2\Delta^2\,\overline{1-z_p}\right) dz_p \qquad \dots \quad (10{\cdot}6)$$

Applying Kummer's first transformation *viz.*,

$$ {}_1F_1(\alpha, \rho, z) = e^{z} \; {}_1F_1(\rho - \alpha, \rho, -z) \qquad \ldots \quad (10{\cdot}61) $$

to (10·6), we can write it in the alternative form

$$ \text{const} \times e^{-\frac{1}{2}c^2\Delta^2 z_p} \, z_p^{(N-p-1)/2}(1-z_p)^{(p-2)/2} $$

$$ \times \; {}_1F_1\left(\frac{1+p-N}{2}, \frac{p}{2}, -\tfrac{1}{2}c^2\Delta^2\,\overline{1-z_p}\right)dz_p \qquad \ldots \quad (10{\cdot}7) $$

Finally remembering that

$$ z_p = \frac{1}{1+\dfrac{p\bar{n}}{2N}D^2} \quad \text{and } c^2 = \frac{p\bar{n}}{2} $$

we get the distribution of D^2 in the two alternative forms

$$ \text{const} \times e^{-\frac{1}{2}p\bar{n}\Delta^2}(D^2)^{(p-2)/2}\left(1+\frac{p\bar{n}}{2N}D^2\right)^{-(N-1)/2} $$

$$ \times \; {}_1F_1\left(\frac{N-1}{2}, \frac{p}{2}, \frac{p^2\bar{n}^2\Delta^2D^2}{8N+4p\bar{n}D^2}\right)d\,D^2 \qquad \ldots \quad (10{\cdot}8) $$

and

$$ \text{const} \times e^{-\frac{(p\bar{n}\Delta^2)/4}{1+(p\bar{n}/2N)D^2}} \, (D^2)^{\frac{p-2}{2}}\left(1+\frac{p\bar{n}}{2N}D^2\right)^{-(N-1)/2} $$

$$ \times \; {}_1F_1\left(\frac{1+p-N}{2}, \frac{p}{2}, -\frac{p^2\bar{n}^2\Delta^2D^2}{8N+4p\bar{n}D^2}\right)d\,D^2 \qquad \ldots \quad (10{\cdot}9) $$

Putting $\Delta^2=0$, the exponential and hypergeometric portions become unity in both (10·8) and (10·9), so that both the forms of distribution reduce as they should to the form (7·4). This further shows that the constant involved in (10·8) as well as (10·9) is the same as that involved in (7·4) *viz.*

$$ \left(\frac{p\bar{n}}{2N}\right)^{p/2} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p-1}{2}\right)} \qquad \ldots \quad (10{\cdot}91) $$

Bibliography.

1. Bose, R. C. On the Exact Distribution of the D²-Statistic *Sankhyā*: Vol. 2 (2), 1936, pp. 143-154.
2. Bose, S. N. On the Complete Moment-Coefficients of the D²-Statistic. *Sankhyā*: Vol. 2 (4), 1936, pp. 385-396.
3. Bose, S. N. On the Moment-Coefficients of the D²-Statistic, and certain Integral and Differential Equations connected with the Multivariate Normal Population. *Sankhyā*: Vol. 3 (2), 1937, pp. 105-124.

4. FISHER, R. A. The general Sampling Distribution of the Multiple Correlation Coefficient *Proc. Roy Soc. A.* Vol. 121, 1928, pp. 654-673.

5. FISHER, R. A. The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics.* Vol. 7 (2), 1936, pp. 179-188.

6. HOTELLING, H. The generalisation of Student's Ratio. *Annal. Math. Stat.* Vol. 2 (3), 1931, pp. 360-378.

7. MAHALANOBIS, P. C. Analysis of Race-Mixture in Bengal. *Jour. and Proc. Asiatic Soc. of Bengal* (New Series). Vol. 23 (3), 1927, pp. 301-333.

8. MAHALANOBIS, P. C. A Statistical Study of the Chinese head. *Man in India.* Vol. 8 (2 and 3), 1928, pp. 107-122.

9. MAHALANOBIS, P. C. On Tests and Measures of Group Divergence. *Jour. Proc. Asiatic Soc. of Bengal* (New Series). Vol. 26 (4), 1930, pp. 541-588.

10. MAHALANOBIS, P. C. On the Generalised Distance in Statistics. *Proc. Nat. Inst. of Sciences of India.* Vol. 2 (1), 1936, pp. 49-55.

11. MAHALANOBIS, P. C., BOSE, R. C. AND ROY, S. N. Normalisation of Statistical Variates, and the use of Rectangular Co-ordinates in the theory of Sampling Distributions. *Sankhyā*: Vol. 3 (1), 1937, pp. 1-40.

12. PEARSON, E. S. AND WILKS, S. S. Methods of Statistical Analysis appropriate for samples of two variables. *Biom.* Vol. 25, 1933, pp. 353-378.

13. PEARSON, KARL. On the Coefficient of Racial Likeness. *Biom.* Vol. 18, 1926. pp. 105-117.

14. ROMANOVSKY, V. On the Criteria that two given samples belong to the same Normal Population. *Metron.* Vol. 7 (3), 1928, pp. 3-46.

15. TILDESLEY, M. L. A first study of the Burmese Skull. *Biom.* Vol. 13, 1921, pp. 176-262.

16. WATSON, G. N. *Theory of Bessel Functions*, p. 70.

17. WATSON, G. N. *Loc. Cit.*, p. 373.

18. WATSON, G. N. *Loc. Cit.*, p. 100 and p. 393.

19. WILKS, S. S. Certain Generalisations in the Analysis of Variance. *Biom.* Vol. 24, 1932, pp. 471-494.

DISCUSSION ON R. C. BOSE AND S. N. ROY'S PAPER.

PROFESSOR FISHER remarked that he, and Professors Hotelling and Mahalanobis had been unwittingly treading the same ground. He was glad to avail himself of the present opportunity to clear up this point. Messrs. R. C. Bose and S. N. Roy's paper has carried the work a distinct step forward.

PROF. MAHALANOBIS pointed out the fundamental distinction between tests and measures of divergence, and mentioned that in dealing with anthropological material he had approached the problem primarily from a quantitative point of view. The present paper supplied the necessary mathematical tool to use the D^2-statistic when only the sample values of the dispersion were known.