RELATIONS BETWEEN TWO SETS OF VARIATES*.

By HAROLD HOTELLING, Columbia University.

CONTENTS.

1. *The Correlation of Vectors. The Most Predictable Criterion and the Tetrad Difference.* Concepts of correlation and regression may be applied not only to ordinary one-dimensional variates but also to variates of two or more dimensions. Marksmen side by side firing simultaneous shots at targets, so that the deviations are in part due to independent individual errors and in part to common causes such as wind, provide a familiar introduction to the theory of correlation; but only the correlation of the horizontal components is ordinarily discussed, whereas the complex consisting of horizontal and vertical deviations may be even more interesting. The wind at two places may be compared, using both components of the velocity in each place. A fluctuating vector is thus matched at each moment with another fluctuating vector. The study of individual differences in mental and physical traits calls for a detailed study of the relations between sets of correlated variates. For example the scores on a number of mental tests may be compared with physical measurements on the same persons. The questions then arise of determining the number and nature of the independent relations of mind and body shown by these data to exist, and of extracting from the multiplicity of correlations in the system suitable characterizations of these independent relations. As another

* Presented before the American Mathematical Society and the Institute of Mathematical Statisticians at Ann Arbor, September 12, 1935.

example, the inheritance of intelligence in rats might be studied by applying not one but s different mental tests to N mothers and to a daughter of each. Then $\frac{s(s-1)}{2}$ correlation coefficients could be determined, taking each of the mother-daughter pairs as one of the N cases. From these it would be possible to obtain a clearer knowledge as to just what components of mental ability are inherited than could be obtained from any single test.

Much attention has been given to the effects of the crops of various agricultural commodities on their respective prices, with a view to obtaining demand curves. The standard errors associated with such attempts, when calculated, have usually been found quite excessive. One reason for this unfortunate outcome has been the large portion of the variance of each commodity price attributable to crops of other commodities. Thus the consumption of wheat may be related as much to the prices of potatoes, rye, and barley as to that of wheat. The like is true of supply functions. It therefore seems appropriate that studies of demand and supply should be made by groups rather than by single commodities*. This is all the more important in view of the discovery that demand and supply curves provide altogether inadequate foundation for the discussion of related commodities, the ignoring of the effects of which has led to the acceptance as part of classical theory of results which are wrong not only quantitatively but qualitatively. It is logically as well as empirically necessary to replace the classical one-commodity type of analysis, relating for example to the incidence of taxation, utility, and consumers' surplus, by a simultaneous treatment of a multiplicity of commodities†.

The relations between two sets of variates with which we shall be concerned are those that remain invariant under internal linear transformations of each set separately. Such invariants are not affected by rotations of axes in the study of wind or of hits on a target, or by replacing mental test scores by an equal number of independently weighted sums of them for comparison with physical measurements. If measurements such as height to shoulder and difference in height of shoulder and top of head are replaced by shoulder height and stature, the invariant relations with other sets of variates will not be affected. In economics there are important linear transformations corresponding for example to the mixing of different grades of wheat in varying proportions‡. Both prices and quantities are then transformed linearly.

In this case, besides the invariants to be discussed in this paper, there will be others resulting from the fact that the transformation of quantities is not independent of that of the prices, but is contragredient to it. (Cf. Section 16 below.)

* The only published study known to the writer of groups of commodities for which standard errors were calculated is the paper of Henry Schultz, "Interrelations of Demand," in *Journal of Political Economy*, Vol. XLI. pp. 468—512, August, 1933. Some at least of the coefficients obtained are significant.

† Harold Hotelling, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions" in *Journal of Political Economy*, Vol. XL. pp. 577—616, October, 1932, and "Demand Functions with Limited Budgets" in *Econometrica*, Vol. III. pp. 66—78, January, 1935.

‡ Harold Hotelling, "Spaces of Statistics and their Metrization" in *Science*, Vol. LXVII. pp. 149—150, February 10, 1928.

Sets of variates, which may also be regarded as many-dimensional variates, or as vectors possessed of frequency distributions, have been investigated from several different standpoints. The work of Gauss on least squares and that of Bravais, Galton, Pearson, Yule and others on multivariate distributions and multiple correlation are early examples. In "The Generalization of Student's Ratio*," the writer has given a method of testing in a manner invariant under linear transformations, and with full statistical efficiency, the significance of sets of means, of regression coefficients, and of differences of means or regression coefficients. A procedure generalizing the analysis of variance to vectors has been applied to the study of the internal structure of cells by means of Brownian movements, for which the vectors representing displacements in consecutive fifteen-second intervals were compared with their resultants to demonstrate the presence of invisible obstructions restricting the movement†. Finally, S. S. Wilks has published important work on relations among two or more sets of variates which are invariant under internal linear transformations‡. Denoting by A, B and D respectively the determinants of sample correlations within a set of s variates, within a set of t variates, and in the set consisting of all these $s+t$ variates, the distribution of the statistic,

$$z = \frac{D}{AB} \qquad \text{(1·1)},$$

was determined exactly by Wilks under the hypothesis that the distribution is normal, with no population correlation between any variate in one set and any in the other. Wilks also found distributions of analogous functions of three or more sets, and of other related statistics.

The statistic (1·1) is invariant under internal linear transformations of either set, as will be proved in Section 4. Another example of such a statistic is provided by the maximum multiple correlation with either set of a linear function of the other set, which has been the subject of a brief study§. This problem of finding, not only a best predictor among the linear functions of one set, but at the same time the function of the other set which it predicts most accurately, will be solved in Section 3 in a more symmetrical manner. When the influence of these two linear functions is eliminated by partial correlation, the process may be repeated with the residuals. In this way we may obtain a sequence of pairs of variates, and of correlations between them, which in the aggregate will fully characterize the invariant relations between the sets, in so far as these can be represented by correlation coefficients. They will be called *canonical variates* and *canonical correlations*. Every invariant under general linear internal transformations, such for example as z, will be seen to be a function of the canonical correlations.

* *Annals of Mathematical Statistics*, Vol. II. pp. 360—378, August, 1931.

† L. G. M. Baas-Becking, Henriette van de Sande Bakhuyzen, and Harold Hotelling, "The Physical State of Protoplasm" in *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, Second Section, Vol. v. (1928).

‡ "Certain Generalizations in the Analysis of Variance" in *Biometrika*, Vol. XXIV. pp. 471—494, November, 1932.

§ Harold Hotelling, "The Most Predictable Criterion" in *Journal of Educational Psychology*, Vol. XXVI. pp. 139—142, February, 1935.

Observations of the values taken in N cases by the components of two vectors constitute two matrices, each of N columns. If each vector has s components, then each matrix has s rows. In this case we may consider the correlation coefficient between the C_s^N-rowed determinants in one matrix and the corresponding determinants in the other. Since a linear transformation of the variates in either set effects a linear transformation of the rows of the matrix of observations, which merely multiplies all these determinants by the same constant, it is evident that the correlation coefficient thus calculated is invariant in absolute value. We shall call it the *vector correlation* or *vector correlation coefficient*, and denote it by q. When $s=2$, if we call the variates of one set x_1, x_2, and those of the other x_3, x_4, and r_{ij} the correlation of x_i with x_j, then it is easy to deduce with the help of the theorems stated in Section 2 below that

$$q=\frac{r_{13}r_{24}-r_{14}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{34}^2)}} \quad \text{..........................(1·2).}$$

For larger values of s, q will have as its numerator the determinant of correlations of variates in one set with variates in the other, and as its denominator the geometric mean of the two determinants of internal correlations. A generalization of q for sets with unequal numbers of components will be given in Section 4.

Corresponding to the correlation coefficient r between two simple variates, T. L. Kelley has defined the *alienation coefficient* as $\sqrt{1-r^2}$. The square of the correlation coefficient between x and y is the fraction of the variance of y attributable to x, while the square of the alienation coefficient is the fraction independent of x. If we adopt this apportionment of variance as a basis of generalization, we shall be consistent in calling $\sqrt{z}$ the *vector alienation coefficient.*

If there exists a linear function of one set equal to a linear function of the other—if for example x_1 is identically equal to x_3—the expression (1·2) for q reduces to a partial correlation coefficient. If one set consists of a single variate and the other of two or more, the vector correlation coincides with the multiple correlation. If each set contains only one variate, q is the simple correlation between the two. Thus the vector correlation coefficient provides a generalization of several familiar concepts.

The numerator of (1·2), known as the tetrad difference or tetrad, has been of much concern to psychologists. The vanishing in the population of all the tetrads among a set of tests is a necessary condition for the theory, propounded by Spearman, that scores on the tests are made up of a component common in varying degrees to all of them, and of independent components specific to each. The vanishing of some but not all of the tetrads is a condition for certain variants of the situation *. The sampling errors of the tetrad have therefore received much attention. In dealing with them it has been thought necessary to ignore at least three types of error:

(1) The standard error formulae used are only asymptotically valid for very large samples, with no means of determining how large a sample is necessary.

* Truman L. Kelley, *Crossroads in the Mind of Man*, Stanford University Press, 1928. This book, in addition to relevant test data and discussion, contains references to the extensive literature, a standard error formula for the tetrad, and other mathematical developments.

(2) The assumption is made implicitly that the distribution of the tetrad is normal, though this cannot possibly be the case, since the range is finite*.

(3) Since the standard error formulae involve unknown population values, these are in practice replaced by sample values. No limit is known for the errors committed in this way.

Now it is evident that to test whether the population value of the tetrad is zero—the only value of interest—is the same thing as to test the vanishing of any multiple of the tetrad by a finite non-vanishing quantity. Wishart† considered the tetrad of covariances, which is simply the product of the tetrad of correlations by the four standard deviations. For this function he found exact values of the mean and standard error, thus eliminating the first source of error mentioned above.

The exact distribution of q found in Section 8 below may be used to test the vanishing of the tetrad, eliminating the first and second sources of error. Unfortunately even this distribution involves a parameter of the population, one of the canonical correlations, which must usually be estimated from the sample, introducing again an error of the third type. However there may be cases in which this one parameter will be known from theory or from a larger sample.

Now it will be shown that q is the product of the canonical correlations. Hence at least one of these correlations is zero if the tetrad is. Thus still another test of the same hypothesis may be made in this way. Now we shall obtain for a canonical correlation vanishing in the population the extremely simple standard error formula $\frac{1}{\sqrt{n}}$, involving no unknown parameter. Thus this test evades errors of the third kind, but is subject to those of the first two, although the second is somewhat mitigated by an ultimate approach to normality, since the canonical correlations satisfy the criterion for approach to normality derived by Doob in the article cited. Further research is needed to find an exact test involving no unknown parameter. The question of whether this is possible raises a very fundamental problem in the theory of statistical inference. We shall, however, find exact distributions appropriate for testing a variety of hypotheses.

2. *Theorems on Determinants and Matrices.* We shall have frequent occasion to refer to the following well-known theorem, the proofs of which parallel those of the multiplication theorem for determinants, and which might advantageously be used in expounding many parts of the theory of statistics:

Given two arrays, each composed of m *rows and* n *columns* (m $\leq$ n). *The determinant formed by multiplying the rows of one array by those of the other equals the*

* The first proof that the distribution of the tetrad approaches normality for large samples was given by J. L. Doob in an article, "The Limiting Distributions of Certain Statistics," in the *Annals of Mathematical Statistics*, Vol. VI. pp. 160—169 (September, 1935). The proof is applicable only if the population value of z is different from unity, i.e. if the sets x_1, x_2 and x_3, x_4 are not completely independent. If they are completely independent, the limiting distribution is of the form $\frac{1}{2}ce^{-c|t|}dt$, as Doob showed. What the distribution of the tetrad is for any finite number of cases no one knows.

† "Sampling Errors in the Theory of Two Factors" in *British Journal of Psychology*, Vol. XIX. pp. 180—187 (1928).

sum of the products of the m*-rowed determinants in the first array by the corresponding* m*-rowed determinants in the second.*

When the two arrays are identical, we have the corollary that *the symmetrical determinant of the products of rows by rows of an array of* m *rows and* n *columns* ($\mathrm{m} \leqslant \mathrm{n}$) *equals the sum of the squares of the* m*-rowed determinants in the array, and is therefore not negative.*

3. *Canonical Variates and Canonical Correlations. Applications to Algebra and Geometry.* If $x_1, x_2, \ldots$ are variates having zero expectations and finite covariances, we denote these covariances by

$$\sigma_{\alpha\beta} = E x_\alpha x_\beta,$$

where E stands for the mathematical expectation of the quantity following. If new variates $x_1', x_2', \ldots$ are introduced as linear functions of the old, such that

$$x_\gamma' = \sum_\alpha c_{\gamma\alpha} x_\alpha,$$

then the covariances of the new variates are expressed in terms of those of the old by the equations

$$\sigma_{\gamma\delta}' = \sum_{\alpha\beta}\sum c_{\gamma\alpha} c_{\delta\beta} \sigma_{\alpha\beta} \quad \text{.............................(3·1)},$$

obtained by substituting the equations above directly in the definition

$$\sigma_{\gamma\delta}' = E x_\gamma' x_\delta',$$

and taking the expectation term by term.

Now (3·1) gives also the formula for the transformation of the coefficients of a quadratic form $\Sigma\Sigma\sigma_{\alpha\beta} x_\alpha x_\beta$ when the variables are subjected to a linear transformation. Hence the problem of standardizing the covariances among a set of variates by linear transformations is algebraically equivalent to the canonical reduction of a quadratic form. The transformation of a quadratic form into a sum of squares corresponds to replacing a set of variates by uncorrelated components. It is to be observed that the fundamental nature of covariances implies that $\Sigma\Sigma\sigma_{\alpha\beta} x_\alpha x_\beta$ is a positive definite quadratic form, and that only real transformations are relevant to statistical problems.

Considering two sets of variates $x_1, \ldots, x_s$ and $x_{s+1}, \ldots, x_{s+t}$, we shall denote the covariances, in the sense of expectations of products, by $\sigma_{\alpha\beta}$, $\sigma_{\alpha i}$, and σ_{ij}, using Greek subscripts for the indices $1, 2, \ldots, s$ and Latin subscripts for $s+1, \ldots, s+t$. Determination of invariant relations between the two sets by means of the correlations or covariances among the $s+t$ variates is associated with the algebraic problem, which appears to be new, of determining the invariants of the system consisting of two positive definite quadratic forms

$$\sum_{\alpha\beta}\sum \sigma_{\alpha\beta} x_\alpha x_\beta, \quad \sum_{ij}\sum \sigma_{ij} x_i x_j,$$

in two separate sets of variables, and of a bilinear form

$$\sum_{\alpha i}\sum \sigma_{\alpha i} x_\alpha x_i$$

in both sets, under real linear non-singular transformations of the two sets separately.

Sample covariances are also transformed by the formula (3·1). The ensuing analysis might therefore equally well be carried out for a sample instead of for the population. Correlations might be used instead of covariances, either for the sample or for the population, by introducing appropriate factors, or by assuming the standard deviations to be unity.

We shall assume that there is no fixed linear relation among the variates, so that the determinant of their covariances or correlations is not zero. This implies that there is no fixed linear relation among any subset of them; consequently every *principal* minor of the determinant of $s+t$ rows is different from zero.

If we consider a function u of the variates in the first set and a function v of those in the second, such that

$$u = \sum_{\alpha} a_\alpha x_\alpha, \quad v = \sum_{i} b_i x_i,$$

the conditions

$$\Sigma\Sigma\sigma_{\alpha\beta} a_\alpha a_\beta = 1, \quad \Sigma\Sigma\sigma_{ij} b_i b_j = 1 \qquad \text{(3·2)}$$

are equivalent to requiring the standard deviations of u and v to be unity. The correlation of u with v is then

$$R = \sum_{\alpha i}\Sigma \sigma_{\alpha i} a_\alpha b_i \qquad \text{(3·3).}$$

If u and v are chosen so that this correlation is a maximum, the coefficients a_α and b_i will satisfy the equations obtained by differentiating

$$\Sigma\Sigma\sigma_{\alpha i} a_\alpha b_i - \tfrac{1}{2}\lambda\Sigma\Sigma\sigma_{\alpha\beta} a_\alpha a_\beta - \tfrac{1}{2}\mu\Sigma\Sigma\sigma_{ij} b_i b_j,$$

namely

$$\sum_i \sigma_{\alpha i} b_i - \lambda \sum_\beta \sigma_{\alpha\beta} a_\beta = 0 \qquad \text{(3·4),}$$

$$\sum_\alpha \sigma_{\alpha i} a_\alpha - \mu \sum_j \sigma_{ij} b_j = 0 \qquad \text{(3·5).}$$

Here λ and μ are Lagrange multipliers. Their interpretation will be evident upon multiplying (3·4) by a_α and summing with respect to α, then multiplying (3·5) by b_i and summing with respect to i. With (3·2) and (3·3), this process gives

$$\lambda = \mu = R.$$

The $s+t$ homogeneous linear equations (3·4) and (3·5) in the $s+t$ unknowns a_α and b_i will determine variates u and v making R a maximum, a minimum, or otherwise stationary, if their determinant vanishes. Since $\lambda = \mu$, this condition is

$$\begin{vmatrix} -\lambda\sigma_{11} \cdots -\lambda\sigma_{1s} & \sigma_{1,s+1} \cdots\cdots \sigma_{1,s+t} \\ \cdots\cdots & \cdots\cdots \\ -\lambda\sigma_{s1} \cdots -\lambda\sigma_{ss} & \sigma_{s,s+1} \cdots\cdots \sigma_{s,s+t} \\ \sigma_{s+1,1} \cdots \sigma_{s+1,s} & -\lambda\sigma_{s+1,s+1} \cdots -\lambda\sigma_{s+1,s+t} \\ \cdots\cdots & \cdots\cdots \\ \sigma_{s+t,1} \cdots \sigma_{s+t,s} & -\lambda\sigma_{s+t,s+1} \cdots -\lambda\sigma_{s+t,s+t} \end{vmatrix} = 0 \qquad \text{(3·6).}$$

This symmetrical determinant is the discriminant of a quadratic form $\phi - \lambda\psi$, where

$$\phi = 2\Sigma\Sigma_{\alpha i}\sigma_{\alpha i} z_\alpha z_i, \quad \psi = \Sigma\Sigma_{\alpha\beta}\sigma_{\alpha\beta} z_\alpha z_\beta + \Sigma\Sigma_{ij}\sigma_{ij} z_i z_j.$$

Here ψ is positive definite because it is the sum of two positive definite quadratic forms. Consequently* all the roots of (3·6) are real. Moreover the elementary divisors are all of the first degree†. This means that the matrix of the determinant in (3·6) is reducible, by transformations which do not affect either its rank or its linear factors, to a matrix having zeros everywhere except in the principal diagonal, while the elements in this diagonal are polynomials

$$E_1(\lambda), \quad E_2(\lambda), \quad \ldots, \quad E_{s+t}(\lambda),$$

none of which contains any linear factor of the form $\lambda - \rho$ raised to a degree higher than the first. Therefore, if a simple root of (3·6) is substituted for λ, the rank is $s+t-1$; but substitution of a root of multiplicity m for λ makes the rank $s+t-m$. Consequently if a simple root is substituted for λ and μ in (3·4) and (3·5) these equations will determine values of $a_1, a_2, \ldots, a_s, b_{s+1}, \ldots, b_{s+t}$, uniquely except for constant factors whose absolute values are determinable from (3·2). Not all these quantities are zero; from this fact, and the form of (3·4) and (3·5), it is evident that at least one a_α and at least one b_i differ from zero, provided the value put for λ is not zero. The variates u and v will then be fully determinate except that they may be replaced by the pair $-u, -v$. But for a root of multiplicity m there will be m linearly independent solutions instead of one in a complete set of solutions. From these may be obtained m different pairs of variates u and v.

The coefficient of the highest power of λ in (3·6) is the product of two principal minors, both of which differ from zero because the variates have been assumed algebraically independent. The equation is therefore of degree $s+t$. We assume as a mere matter of notation, if $s \neq t$, that $s < t$. Then of the $s+t$ roots at least $t-s$ vanish; for the coefficients of λ^{t-s-1} and lower powers of λ are sums of principal minors of $2s+1$ or more rows, in which λ is replaced by zero, and every such minor vanishes, as can be seen by a Laplace expansion. Also, the sign of λ may be changed in (3·6) without changing the equation, for this may be accomplished by multiplying each of the first s rows and last t columns by -1. Therefore the negative of every root is also a root. The $s+t-(t-s) = 2s$ roots that do not necessarily vanish consist therefore of s positive or zero roots $\rho_1, \rho_2, \ldots, \rho_s$, and of the negatives of these roots. These s roots which are positive or zero we shall call the *canonical correlations* between the sets of variates; the corresponding linear functions u, v whose coefficients satisfy (3·2), (3·4) and (3·5) we call *canonical variates*‡. It is clear that every canonical correlation is the correlation coefficient between a pair of canonical variates. Hence no canonical correlation can exceed unity. The greatest canonical correlation is the maximum multiple correlation

* Maxime Bôcher, *Introduction to Higher Algebra*, New York, 1931, p. 170, Theorem 1.

† Bôcher, p. 305, Theorem 4; p. 267, Theorem 2; p. 271, Definition 3.

‡ The word "canonical" is used in the algebraic theory of invariants with a meaning consistent with that of this paper.

with either set of a disposable linear function of the other set. If u, v are canonical variates corresponding to ρ_γ, then the pair $u, -v$ or $v, -u$ is associated with the root $-\rho_\gamma$.

If a pair of canonical variates corresponding to a root ρ_γ is

$$u_\gamma = \sum_\alpha a_{\alpha\gamma} x_\alpha, \quad v_\gamma = \sum_i b_{i\gamma} x_i \quad \text{..........(3·7)},$$

the coefficients must satisfy (3·4) and (3·5), so that

$$\sum_i \sigma_{\alpha i} b_{i\gamma} = \rho_\gamma \sum_\beta \sigma_{\alpha\beta} a_{\beta\gamma} \quad \text{..........(3·8)},$$

$$\sum_\alpha \sigma_{\alpha i} a_{\alpha\gamma} = \rho_\gamma \sum_j \sigma_{ij} b_{j\gamma} \quad \text{..........(3·9)}.$$

Also let

$$u_\delta = \sum_\beta a_{\beta\delta} x_\beta, \quad v_\delta = \sum_i b_{i\delta} x_i \quad \text{..........(3·10)}$$

be canonical variates associated with a canonical correlation ρ_δ. Among the four variates (3·7) and (3·10) there are six correlations. Apart from ρ_γ and ρ_δ these are obviously

$$\left.\begin{array}{ll} Eu_\gamma u_\delta = \Sigma\Sigma\sigma_{\alpha\beta} a_{\beta\gamma} a_{\alpha\delta}, & Eu_\gamma v_\delta = \Sigma\Sigma\sigma_{\alpha i} a_{\alpha\gamma} b_{i\delta} \\ Ev_\gamma u_\delta = \Sigma\Sigma\sigma_{\alpha i} b_{i\gamma} a_{\alpha\delta}, & Ev_\gamma v_\delta = \Sigma\Sigma\sigma_{ij} b_{i\gamma} b_{j\delta} \end{array}\right\} \quad \text{..........(3·11)}.$$

We shall prove that the last four are all zero. Multiply (3·8) by $a_{\alpha\delta}$ and sum with respect to α. The result, with the help of (3·11), may be written

$$Ev_\gamma u_\delta = \rho_\gamma Eu_\gamma u_\delta \quad \text{..........(3·12)}.$$

Multiplying (3·9) by $b_{i\delta}$ and summing with respect to i, we get

$$Eu_\gamma v_\delta = \rho_\gamma Ev_\gamma v_\delta \quad \text{..........(3·13)}.$$

Interchanging γ and δ in this and then using (3·12), we obtain

$$\rho_\gamma Eu_\gamma u_\delta = \rho_\delta Ev_\gamma v_\delta \quad \text{..........(3·14)}.$$

Again interchanging γ and δ, we have

$$\rho_\delta Eu_\gamma u_\delta = \rho_\gamma Ev_\gamma v_\delta.$$

If $\rho_\gamma^2 \neq \rho_\delta^2$, the last two equations show that $Eu_\gamma u_\delta = Ev_\gamma v_\delta = 0$. Hence, by (3·12) and (3·13), $Ev_\gamma u_\delta$ and $Eu_\gamma v_\delta$ vanish. Thus all the correlations among canonical variates are zero except those between the canonical variates associated with the same canonical correlation.

If ρ_α is a root of multiplicity m, it is possible by well-known processes to obtain m solutions of the linear equations such that, if

$$a_{1\gamma}, \ \ldots \ a_{s\gamma}, \ b_{s+1,\gamma}, \ \ldots, \ b_{s+t,\gamma},$$

$$a_{1\delta}, \ \ldots, \ a_{s\delta}, \ b_{s+1,\delta}, \ \ldots, \ b_{s+t,\delta},$$

are any two of these solutions, they will satisfy the orthogonality condition

$$\sum_\alpha a_{\alpha\gamma} a_{\alpha\delta} + \sum_i b_{i\gamma} b_{i\delta} = 0 \quad \text{..........(3·15)}.$$

There is no loss of generality in supposing that each of the original variates was uncorrelated with the others in the same set and had unit variance. In this case (3·15) is equivalent to

$$Eu_\gamma u_\delta + Ev_\gamma v_\delta = 0,$$

where u_γ, v_γ, u_δ, v_δ are given by (3·7) and (3·10). For this case of equal roots we have also from (3·14),

$$\rho_\alpha (Eu_\gamma u_\delta - Ev_\gamma v_\delta) = 0.$$

If $\rho_\alpha \neq 0$, the last two equations show that $Eu_\gamma u_\delta = Ev_\gamma v_\delta = 0$, and then from (3·12) and (3·13) we have that $Ev_\gamma u_\delta = Eu_\gamma v_\delta = 0$. These correlations also vanish if $\rho_\alpha = 0$, for then the right-hand members of (3·8) and (3·9) vanish, leaving two distinct sets of equations in disjunct sets of unknowns. The solutions may therefore be chosen so that the two sums in (3·15) vanish separately.

A double zero root determines uniquely, if $s = t$, a pair of canonical variates. If $s < t$, such a root determines a canonical variate for the less numerous set, and leaves $t - s$ degrees of freedom for the choice of the other.

The reduction of our sets of variates to canonical form may be completed by the choice of new variates $v_{s+1}, v_{s+2}, \ldots, v_t$ as linear functions of the second and more numerous set (unless the numbers in the two sets are equal), uncorrelated with each other and with the canonical variates v_γ previously determined, and having unit variance. This may be done in infinitely many ways, as is well known. These variates will also be uncorrelated with the canonical variates u_γ. Indeed, if

$$v_k = \Sigma b_{jk} x_j$$

is one of them, its correlation with u_γ is, by (3·7) and (3·9),

$$Eu_\gamma v_k = \Sigma\Sigma \sigma_{\alpha j} a_{\alpha\gamma} b_{jk} = \rho_\gamma \Sigma\Sigma \sigma_{ij} b_{i\gamma} b_{jk} = \rho_\gamma Ev_\gamma v_k,$$

which vanishes because v_k was defined to be uncorrelated with v_γ.

The normal form of two sets of variates under internal linear transformations is thus found to consist of linear functions $u_1, u_2, \ldots, u_s$ of one set, and $v_1, v_2, \ldots, v_t$ of the other, such that all the correlations among these linear functions are zero, except that the correlation of u_γ with v_γ is a positive number ρ_γ ($\gamma = 1, 2, \ldots, s$). Therefore *the only invariants of the system under internal linear transformations are* $\rho_1, \rho_2, \ldots, \rho_s$, *and functions of these quantities.*

The solution of the algebraic problem mentioned at the beginning of this section, by steps exactly parallel to those just taken with the statistical problem, is the following:

The positive definite quadratic forms $\Sigma\Sigma\sigma_{\alpha\beta} x_\alpha x_\beta$, *and* $\Sigma\Sigma\sigma_{ij} x_i x_j$, *and the bilinear form* $\Sigma\Sigma\sigma_{\alpha i} x_\alpha x_i$ *with real coefficients, where the Latin subscripts are summed from* 1 *to* s *and the Greek subscripts from* $s+1$ *to* $s+t$, *and* $s \leqslant t$, *may be reduced by a real linear transformation of* $x_1, \ldots, x_s$ *and a real linear transformation of* $x_{s+1}, \ldots, x_{s+t}$ *simultaneously to the respective forms* $x_1^2 + \ldots + x_s^2$, $x_{s+1}^2 + \ldots + x_{s+t}^2$, *and* $\rho_1 x_1 x_{s+1} + \rho_2 x_2 x_{s+2} + \ldots + \rho_s x_s x_{2s}$. *A fundamental system of invariants under such transformations consists of* $\rho_1, \ldots, \rho_s$.

This algebraic theorem holds also if the quadratic forms are not restricted to be positive definite, provided (3·6) has no multiple roots and the forms are non-singular.

The normalization process we have defined may also be carried out for a sample, yielding canonical correlations $r_1, r_2, \ldots, r_s$, which may be regarded as estimates of $\rho_1, \rho_2, \ldots, \rho_s$, and associated canonical variates. With sampling problems raised in this way we shall largely be concerned in the remainder of this paper.

A further application is to geometry. In a space of N dimensions a sample of N values of a variate may be represented by a point whose coordinates are the observed values. The sample correlation between two variates is the cosine of the angle between lines drawn from the origin to the representing points, with the proviso, since deviations from means are used in the expression for a correlation, that the sum of all the coordinates of each point be zero. A sample of $s+t$ variates determines a flat space of s and one of t dimensions, intersecting at the origin, and containing the points representing the two sets of variates. In typical cases these two flat spaces do not intersect except at this one point. A complete set of metrical invariants of a pair of flat spaces is easily seen from the foregoing analysis to consist of s angles whose cosines are $r_1, \ldots, r_s$. Indeed, like all correlations, they are invariant under rotations of the N-space about the origin, and they do not depend on the particular points used to define the two flat spaces. Each of these invariants is the angle between a line in one flat space and a line in the other. One of the invariants is the *minimum* angle of this kind, and the others are in a sense stationary. The condition that the two flat spaces intersect in a line is that one of the invariant quantities $r_1, \ldots, r_s$ be unity. They intersect in a plane if two of these quantities equal unity. For two planes through a point in space of four or more dimensions, there will be two invariants r_1, r_2, of which one is the cosine of the minimum angle. If $r_1 = r_2$, the planes are *isocline*. Every line in each plane then makes the minimum angle with some line in the other. If $r_1 = r_2 = 0$, the planes are completely perpendicular; every line in one plane is then perpendicular to every line in the other. If one of these invariants is zero and the other is not, the planes are semi-perpendicular; every line in each plane is perpendicular to a certain line in the other.

The determinant of the correlations among canonical variates is

$$\Delta = \begin{vmatrix} 1 & 0 & \ldots & 0 & \rho_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \rho_2 & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & \rho_s & \ldots & 0 \\ \rho_1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \rho_2 & \ldots & 0 & 0 & 1 & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \rho_s & 0 & 0 & \ldots & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 1 \end{vmatrix}$$

$$= (1-\rho_1^2)(1-\rho_2^2)\ldots(1-\rho_s^2) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3{\cdot}16).$$

The rank of the matrix

$$\begin{Vmatrix} \rho_{1,s+1} & \cdots\cdots\cdots & \rho_{1,s+t} \\ \cdots & \cdots & \cdots \\ \rho_{s,s+1} & \cdots\cdots\cdots & \rho_{s,s+t} \end{Vmatrix}$$

of correlations between the two sets is invariant under non-singular linear transformations of either set. Transformation to canonical variates reduces this matrix to

$$\begin{Vmatrix} \rho_1 & 0 & \dots & 0 & \dots\dots\dots & 0 \\ 0 & \rho_2 & \dots & 0 & \dots\dots\dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \rho_s & \dots\dots\dots & 0 \end{Vmatrix}.$$

The rank is therefore the number of canonical correlations that do not vanish. This is the number of independent components common to the two sets. In the parlance of mental testing, the number of "common factors" of two sets of tests (e.g. mental and physical, or mathematical and linguistic tests) is the number of non-vanishing canonical correlations.

4. *Vector Correlation and Alienation Coefficients.* In terms of the covariances among the variates in the two sets $x_1, \dots, x_s$ and $x_{s+1}, \dots, x_{s+t}$, we define the following determinants, maintaining the convention that Greek subscripts take values from 1 to s, and Latin subscripts take values from $s+1$ to $s+t$. It will be assumed throughout that $s \leqslant t$. A is the determinant of the covariances among the variates in the first set, arranged in order: that is, the element in the αth row and βth column of A is $\sigma_{\alpha\beta}$. B is the determinant of the covariances among variates in the second set, likewise ordered. D is the determinant of $s+t$ rows containing in order all the covariances among all the variates of both sets. C is obtained from D by replacing the covariances among the variates of the first set, including their variances, by zeros. Symbolically,

$$A = |\sigma_{\alpha\beta}|, \quad B = |\sigma_{ij}|, \quad C = \left|\begin{array}{c|c} 0 & \sigma_{\alpha i} \\ \hline \sigma_{i\alpha} & \sigma_{ij} \end{array}\right|, \quad D = \left|\begin{array}{c|c} \sigma_{\alpha\beta} & \sigma_{\alpha i} \\ \hline \sigma_{i\alpha} & \sigma_{ij} \end{array}\right|.$$

Suppose now that new variates $x_1', \dots, x_s'$ are defined in terms of the old variates in the first set by the s equations

$$x_\gamma' = \Sigma c_{\gamma\alpha} x_\alpha.$$

The new covariances are then expressed in terms of the old by (3·1). The determinant of these new covariances, which we shall denote by A', may by (3·1) and the multiplication theorem of determinants be expressed as the product of three determinants, of which two equal the determinant $c = |c_{\gamma\alpha}|$ of the coefficients of the transformation, while the third is A. If the variates of the second set are subjected to a transformation of determinant d, the determinants of covariances among the new variates analogous to those defined above are readily seen in this way to equal

$$A' = c^2 A, \quad B' = d^2 B, \quad C' = c^2 d^2 C, \quad D' = c^2 d^2 D \quad \dots\dots\dots\dots (4{\cdot}1).$$

Thus A, B, C, D are *relative invariants* under internal transformations of the two sets of variates.

The ratios $$Q^2 = \frac{(-1)^s C}{AB} \text{ and } Z = \frac{D}{AB} \qquad \text{.........................(4·2)}$$

we shall call respectively the squares of the *vector correlation coefficient* or *vector correlation*, and of the *vector alienation coefficient*. It is evident that both are absolute invariants under internal transformations of the two sets, since their values computed from transformed variates have numerators and denominators multiplied by the same factor c^2d^2, in accordance with (4·1).

The notation just used is appropriate to a population, but the same definitions and reasoning may be applied to a sample. We denote by q^2 and z the same functions of the sample covariances that Q^2 and Z, respectively, have been defined to be of the population covariances.

A particularly simple linear transformation consists of dividing each variate by its standard deviation. The covariances among the new variates are then the same as their correlations, which are also the correlations among the old variates. Hence, in the definitions of the vector correlation and alienation coefficients, the covariances may be replaced by the correlations. For example, if $s = t = 2$, the squared vector correlation in a sample may be written

$$q^2 = \frac{\begin{vmatrix} 0 & 0 & r_{13} & r_{14} \\ 0 & 0 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & r_{34} \\ r_{34} & 1 \end{vmatrix}} = \frac{(r_{13}r_{24} - r_{14}r_{23})^2}{(1 - r_{12}^2)(1 - r_{34}^2)} \qquad \text{.........(4·3).}$$

The vector correlation coefficient will always be taken as the positive square root of q^2 or of Q^2 (which are seen below to be positive) when $s < t$, and usually also when $s = t$. However, if in accordance with (4·3) we write

$$q = \frac{r_{13}r_{24} - r_{14}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{34}^2)}} \qquad \text{...........................(4·4),}$$

it is evident that q may be positive for some samples of a particular set of variates, and negative for other samples. It may sometimes be advantageous, as in testing whether two samples arose from the same population, to retain the sign of q for each sample, since this provides evidence in addition to that given by the absolute value of q. But unless otherwise stated we shall always regard q as the positive root of q^2. Likewise, Q, $\sqrt{z}$ and $\sqrt{Z}$ will denote the positive roots unless otherwise specifically indicated in each case. A transformation of either set will reverse the sign of the algebraic expression (4·4) if the determinant of the transformation is negative. This will be true of a simple interchange of two variates; for example, $x_1' = x_2$, $x_2' = x_1$ has the determinant -1. On the other hand, the sign is conserved if the determinant of the transformation is positive. Such considerations apply whenever $s = t$.

Since the vector correlation and alienation coefficients are invariants, they may be computed on the assumption that the variates are canonical. In this case $A = B = 1$, and D is given by (3·16). To obtain C we replace the first s 1's in the principal diagonal of (3·16) by 0's. It then follows that

$$C = (-1)^s \rho_1^2 \rho_2^2 \dots \rho_s^2.$$

This confirms that the value of Q^2 given in (4·2) is positive. In this way the vector correlation and alienation coefficients are expressible in terms of the canonical correlations by the equations

$$Q = \pm\, \rho_1 \rho_2 \dots \rho_s, \quad Z = (1-\rho_1^2)(1-\rho_2^2)\dots(1-\rho_s^2) \quad \dots\dots\dots(4·5),$$

$$q = \pm\, r_1 r_2 \dots r_s, \quad z = (1-r_1^2)(1-r_2^2)\dots(1-r_s^2) \quad \dots\dots\dots(4·6).$$

From these results it is obvious that both the vector correlation and vector alienation coefficients are confined to values not exceeding unity. Also Z and z are necessarily positive, except that they vanish if, and only if, all the variates in one set are linear functions of those in the other.

Since the denominator of (4·4) is obviously less than unity, and since we have just shown that $q \leqslant 1$, the tetrad must be still less. This simple proof that the tetrad is between -1 and $+1$ shows the falsity of the idea that the range of the tetrad is from -2 to $+2$, which has gained some currency. An equivalent proof in vector notation was communicated to the writer by E. B. Wilson.

The only case in which Z can attain its maximum value unity is that in which all the canonical correlations vanish. In this case no variate in either set is correlated with any variate in the other, so that the two sets are completely independent, at least if the distribution is normal. Moreover, $Q = 0$. On the other hand, the only case in which Q can be unity is that in which all the canonical correlations are unity. In this event, $Z = 0$; also, the variates in the first set are linear functions of those in the second. Thus either z, $1-q$, or $1-q^2$ might be used as an index of independence, while we might use q, q^2 or $1-z$ as a measure of relationship between the two sets. The work of Wilks alluded to in Section 1 provides an exact distribution of z on the hypothesis of complete independence, a distribution which may thus be used to test this hypothesis.

If we regard the elements of A, B and C as sample covariances, we have in case $s = t$ a simple interpretation of q. Consider the two matrices of observations on the two sets of variates in N individuals, in which each row corresponds to a variate and each column to an individual observed. From Section 2 it is evident that the square of the sum of the products of corresponding s-rowed determinants in the two matrices is $(-1)^s N^{2s} C$; also that the sums of squares of the s-rowed determinants in the two matrices are $N^s A$ and $N^s B$. Therefore q is simply the product-moment correlation coefficient between corresponding s-rowed determinants.

The generalized variance of a set of variates may be defined as the determinant of their ordered covariances, such as A or B. Let $\xi_1, \xi_2, \dots, \xi_s$ be estimates respectively of $x_1, x_2, \ldots, x_s$ obtained from $x_{s+1}, \ldots, x_{s+t}$ by least squares, and let the regression equations be

$$\xi_a = \sum_i b_{ai} x_i \qquad (4\cdot7).$$

The appropriateness of Q as a generalization of the correlation coefficient, and of $\sqrt{Z}$ as a generalization of the alienation coefficient, will be apparent from the following theorem:

The ratio of the generalized variance of $\xi_1, \ldots, \xi_s$ to that of $x_1, \ldots, x_s$ is Q^2. The ratio of the generalized variance of $x_1 - \xi_1, x_2 - \xi_2, \ldots, x_s - \xi_s$ to that of $x_1, \ldots, x_s$ is Z.

This theorem is expressed in terms of the population, but an exactly parallel one holds for a sample.

Proof: If $x_1, \ldots, x_s$ be subjected to a linear transformation of determinant c, and if $\xi_1, \ldots, \xi_s$ be subjected to the same transformation (i.e. a transformation with the same coefficients), then $x_1 - \xi_1, \ldots, x_s - \xi_s$ will also be subjected to this transformation. The generalized variances of all three of these sets of variates will be multiplied by the same constant c^2, just as in (4·1) we found that $A' = c^2 A$. Ratios among the three determinants will therefore be absolutely invariant; consequently our theorem is true if it is true when the original variates are canonical. Suppose, then, that this is the case. Since each canonical variate is correlated with only one of the other set, the regression equations (4·7) reduce simply to

$$\xi_a = \rho_a x_{a+s} \quad (a = 1, \ldots, s).$$

Since the variance of x_{a+s} is unity, that of ξ_a is ρ_a^2; that of the deviation $x_a - \xi_a$ is $1 - \rho_a^2$. Since the canonical variates $x_1, \ldots, x_s$ are mutally uncorrelated, the same is true of the ξ_a, and also of the $x_a - \xi_a$. The generalized variance of the canonical variates is unity; that of the ξ_a is the product of the elements in its principal diagonal, namely $\rho_1^2 \rho_2^2 \ldots \rho_s^2$; and the generalized variance of the $x_a - \xi_a$ is $(1 - \rho_1^2) \ldots (1 - \rho_s^2)$. In view of (4·5), this proves the theorem.

A further property of the vector correlation is obvious from the final paragraph of Section 3:

A necessary and sufficient condition that the number of components in an uncorrelated set of components common to two sets of variates be less than the number of variates in either set is that the vector correlation be zero.

When $s = 2$ the canonical correlations not only determine the vector correlation and alienation coefficients but are determined by them. If as usual we take q positive, (4·6) becomes $q = r_1 r_2$, $z = (1 - r_1^2)(1 - r_2^2)$, whence

$$r_1^2 + r_2^2 = 1 - z + q^2, \quad r_1^2 r_2^2 = q^2 \qquad (4\cdot8).$$

Solving, and denoting the greater canonical correlation by r_1, we have

$$\left.\begin{aligned} r_1 &= \tfrac{1}{2}\left[\sqrt{(1+q)^2 - z} + \sqrt{(1-q)^2 - z}\right] \\ r_2 &= \tfrac{1}{2}\left[\sqrt{(1+q)^2 - z} - \sqrt{(1-q)^2 - z}\right] \end{aligned}\right\} \qquad (4\cdot9).$$

Since the canonical correlations are real, $(r_1 - r_2)^2$ is positive; therefore

$$z \leqslant (1-q)^2 \quad \ldots\ldots(4\cdot10).$$

In like manner, the vector correlation and alienation coefficients in the population are subject not only to the inequalities $0 \leqslant Q^2 \leqslant 1$, $0 \leqslant Z \leqslant 1$, but also, when $s=2$, to

$$Z \leqslant (1-Q)^2.$$

These inequalities become equalities when the roots are equal.

The fundamental equation (3·6), regarded as an equation in λ^2, has as roots the squares of the canonical correlations. Hence, by (4·8), it reduces it to the form

$$\lambda^4 - (1 - z + q^2)\lambda^2 + q^2 = 0 \quad \ldots\ldots(4\cdot11),$$

where $s=2$.

5. *Standard Errors.* The canonical correlations and the coefficients of the canonical variates are defined in Section 3 in such a way that they are continuous functions of the covariances, with continuous derivatives of all orders, except for certain sets of values corresponding to multiple or zero roots, within the domain of variation for which the covariances are the coefficients of a positive definite quadratic form. This is true for the canonical reduction of a sample as well as for that of a population. The probability that a random sample from a continuous distribution will yield multiple roots is zero; and sample covariances must always be the coefficients of a positive definite form.

We shall in this section derive asymptotic standard errors, variances and covariances for the canonical correlations on the assumption that those in the population are unequal, and that the population has the multiple normal distribution. From these we shall derive standard errors for the vector correlation and alienation coefficients q and z. The deviation of sample from population values in these as in most cases have variances of order n^{-1}, and distributions approaching normality of form as n increases*.

Let $x_1, \ldots, x_p$ be a normally distributed set of variates of zero means and covariances

$$\sigma_{ij} = Ex_i x_j \quad \ldots\ldots(5\cdot1).$$

For a sample of N in which x_{if} is the value of x_i observed in the fth individual, the sample covariance of x_i and x_j is

$$s_{ij} = \frac{\sum\limits_f (x_{if} - \bar{x}_i)(x_{jf} - \bar{x}_j)}{N-1} = \frac{\Sigma x_{if}x_{jf} - N\bar{x}_i\bar{x}_j}{N-1} \quad \ldots\ldots(5\cdot2),$$

where $\bar{x}_i$ and $\bar{x}_j$ are the sample means. To simplify the later work, we introduce the *pseudo-observations*, x_{if}', defined in terms of the observations by the equations

$$x_{if}' = \sum_{g=1}^{N} c_{fg} x_{ig} \quad \ldots\ldots(5\cdot3),$$

* For a proof of approach to normality for a general class of statistics including those with which we deal, cf. Doob, *op. cit.*

where the quantities c_{fg}, independent of i and therefore the same for all the variates x_i, are the coefficients of an orthogonal transformation, such that

$$c_{N1} = c_{N2} = \ldots = c_{NN} = \frac{1}{\sqrt{N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot4).$$

Since the transformation is orthogonal we must have

$$\sum_h c_{fh} c_{gh} = \delta_{fg} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot5),$$

where δ_{fg} is the Kronecker delta, equal to unity if $f = g$, but to zero if $f \neq g$. The coefficients c_{fg} may be chosen in an infinite variety of ways consistently with these requirements, but will be held fixed throughout the discussion. Since linear functions of normally distributed variates are normally distributed, the pseudo-observations are normally distributed. Their population means are, from (5·3),

$$Ex_{if}' = \Sigma c_{fg} Ex_{ig} = 0,$$

since the original variates were assumed to have zero means. Also, since the expectation of the product of independent variates is zero, and since the different individuals in a sample are assumed independent, so that, by (5·1),

$$Ex_{ih} x_{jk} = \delta_{hk} \sigma_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot6),$$

we have, from (5·3), (5·6) and (5·5),

$$\left.\begin{aligned} Ex_{if}' x_{jg}' &= \sum_{hk}\sum c_{fh} c_{gk} Ex_{ih} x_{jk} \\ &= \sum_{hk}\sum c_{fh} c_{gk} \delta_{hk} \sigma_{ij} \\ &= \sum_h c_{fh} c_{gh} \sigma_{ij} \\ &= \delta_{fg} \sigma_{ij} \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot7).$$

From (5·4) and (5·3) it is clear that

$$x_{iN}' = \frac{\sum_g x_{ig}}{\sqrt{N}} = \sqrt{N}\, \bar{x}_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot8).$$

The equations (5·3) may, on account of their orthogonality, be solved in the form

$$x_{if} = \pm\, \Sigma c_{gf} x_{ig}'.$$

Therefore, by (5·5),

$$\sum_f x_{if} x_{jf} = \Sigma\Sigma\Sigma c_{gf} c_{hf} x_{ig}' x_{jh}' = \Sigma\Sigma \delta_{gh} x_{ig}' x_{jh}' = \sum_g x_{ig}' x_{jg}'.$$

Substituting this result and (5·8) in (5·2), we find that the final term of the sum cancels out. Introducing therefore the symbol S for summation from 1 to $N-1$ with respect to the second subscript, and putting also

$$n = N - 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot9),$$

we have the compact result

$$s_{ij} = \frac{Sx_{ig}' x_{jg}'}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5\cdot10).$$

Since the pseudo-observations are normally distributed with the covariances (5·7) and zero means, they have exactly the same distribution as the observations in a random sample of n from the original population. The equivalence of the mean product (5·10) with the sample covariance (5·2) establishes the important principle that *the distribution of covariances in a sample of* $n+1$ *is exactly the same as the distribution of mean products in a sample of* n, if the parent population is normally distributed about zero means. Use of this principle will considerably simplify the discussions of sampling.

An important extension of this consideration lies in the use of deviations, not merely from sample means, but from regression equations based on other variates. In such cases the number of degrees of freedom n to be used is the difference between the sample number and the number of constants in each of the regression equations, which number must be the same for all the deviations. The estimate of covariance of deviations in the ith and jth variates to be used is then the sum of the products of corresponding deviations, divided by n. This may also be regarded as the mean product of the values of x_i and x_j in n pseudo-observations, as above, without elimination of the means or of the extraneous variates. The sampling distributions with which we shall be concerned will all be expressed in terms of the number of degrees of freedom n, rather than in terms of the number of observations N. This will permit immediately of the extension, which is equivalent to replacing all the correlations, in terms of which our statistics may be defined, by partial correlations representing the elimination of a particular set of variates, the same in all cases.

A variance is of course the covariance of a variate with itself, so that this whole discussion of covariances is equally applicable to variances.

The characteristic function of a multiple normal distribution with zero means is well known to be

$$M(t_1, t_2, \dots) = Ee^{\Sigma t_i x_i} = e^{\Sigma\Sigma\sigma_{ij}t_it_j/2}.$$

The moments of the distribution are the derivatives of the characteristic function, evaluated for $t_1 = t_2 = \dots = 0$. From the fourth derivative with respect to t_i, t_j, t_k and t_m it is easy to show in this way that

$$Ex_ix_jx_kx_m = \sigma_{ij}\sigma_{km} + \sigma_{im}\sigma_{jk} + \sigma_{ik}\sigma_{jm} \qquad (5\cdot11).$$

From (5·10) we have

$$Es_{ij}s_{km} = \frac{1}{n^2} SSEx_{ig}'x_{jg}'x_{kf}'x_{mf}' \qquad (5\cdot12).$$

Now if $f \neq g$,

$$Ex_{ig}'x_{jg}'x_{kf}'x_{mf}' = (Ex_{ig}'x_{jg}')(Ex_{kf}'x_{mf}') = \sigma_{ij}\sigma_{km} \qquad (5\cdot13),$$

since the expectation of the product of *independent* quantities is the product of their expectations. Of the n^2 terms in the double sum in (5·12), $n^2 - n$ are equal to (5·13). The remaining n terms are those for which $f = g$, and each of them equals (5·11). Hence

$$Es_{ij}s_{km} = \sigma_{ij}\sigma_{km} + \frac{1}{n}(\sigma_{im}\sigma_{jk} + \sigma_{ik}\sigma_{jm}).$$

Inasmuch as

$$Es_{ij} = \sigma_{ij},$$

we have, if we put $$d\sigma_{ij} = s_{ij} - \sigma_{ij},$$

for the deviation of sample from population value, that the sampling covariance of two covariances is

$$Ed\sigma_{ij}d\sigma_{km} = Es_{ij}s_{km} - \sigma_{ij}\sigma_{km},$$

whence $$Ed\sigma_{ij}d\sigma_{km} = \frac{1}{n}(\sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk}) \quad \text{.....................(5·14)}.$$

This is a fundamental formula from which may be derived directly a number of more familiar special cases. For example, to obtain the variance of a variance, merely put $i = j = k = m$, which gives

$$\sigma_{sii}^2 = E(d\sigma_{ii})^2 = \frac{2\sigma_{ii}^2}{n}.$$

Returning from these general considerations to the problem of canonical correlations, we recall from (3·2) and (3·3) that for any particular canonical correlation ρ_1,

$$\Sigma\Sigma\sigma_{\alpha\beta}a_\alpha a_\beta = 1, \quad \Sigma\Sigma\sigma_{ij}b_ib_j = 1, \quad \rho_1 = \Sigma\Sigma\sigma_{\alpha i}a_\alpha b_i \quad \text{...........(5·15)},$$

where α and β are summed from 1 to s, and i and j from $s+1$ to $s+t$. Any particular set of sampling errors $d\sigma_{AB}$ in the covariances determines a corresponding set of sampling errors in the a_α and b_i and in ρ_1, for these quantities are definite analytic functions of the covariances except when ρ_1 is a multiple or zero root of (3·6), cases which we now exclude from consideration. In terms of the derivatives of these functions we define

$$da_\alpha = \Sigma\Sigma\frac{\partial a_\alpha}{\partial\sigma_{AB}}d\sigma_{AB}, \quad db_i = \Sigma\Sigma\frac{\partial b_i}{\partial\sigma_{AB}}d\sigma_{AB}, \quad d\rho_1 = \Sigma\Sigma\frac{\partial\rho_1}{\partial\sigma_{AB}}d\sigma_{AB} \quad (5·16),$$

where $d\sigma_{AB} = s_{AB} - \sigma_{AB}$, and the summations are over all values of A and B from 1 to $s+t$. Then differentiating (5·15) we have

$$\left.\begin{aligned}&\Sigma\Sigma(2\sigma_{\alpha\beta}a_\alpha da_\beta + a_\alpha a_\beta d\sigma_{\alpha\beta}) = 0, \quad \Sigma\Sigma(2\sigma_{ij}b_idb_j + b_ib_jd\sigma_{ij}) = 0,\\ &d\rho_1 = \Sigma\Sigma(\sigma_{\alpha i}a_\alpha db_i + \sigma_{\alpha i}b_ida_\alpha + a_\alpha b_id\sigma_{\alpha i})\end{aligned}\right\} \quad (5·17).$$

Let us now suppose that the variates are in the population canonical. This assumption does not entail any loss of generality as regards ρ_1, since ρ_1 is an invariant under transformations of the variates of either set. Since a_α is the coefficient of x_α in the expression for one of the canonical variates, which we take to be x_1, we have in the population $a_1 = 1$, $a_2 = a_3 = \ldots = a_s = 0$. In the same way,

$$b_{s+1} = 1, \quad b_{s+2} = \ldots = b_{s+t} = 0.$$

Also, since the covariances among canonical variates are the elements of the determinant in (3·16), we have

$$\sigma_{\alpha\beta} = \delta_{\alpha\beta}, \quad \sigma_{ij} = \delta_{ij}, \quad \sigma_{\alpha i} = \delta_{\alpha+s,\, i}\rho_\alpha \quad \text{.................(5·18)},$$

the Kronecker deltas being equal to unity if the two subscripts are equal, and otherwise vanishing. When these special values of the a's, b's and σ's are substituted in (5·17) most of the terms drop out, leaving the simple equations

$$\left.\begin{aligned}&2da_1 + d\sigma_{11} = 0, \quad 2db_{s+1} + d\sigma_{s+1,\, s+1} = 0,\\ &d\rho_1 = \rho_1 db_{s+1} + \rho_1 da_1 + d\sigma_{1,\, s+1}\end{aligned}\right\} \text{..............(5·19)}.$$

Substituting from the first two in the third of these equations, we get

$$d\rho_1 = d\sigma_{1,s+1} - \tfrac{1}{2}\rho_1 (d\sigma_{11} + d\sigma_{s+1,s+1}) \quad \text{.................(5·20).}$$

For any other simple root ρ_2 we have in the same way

$$d\rho_2 = d\sigma_{2,s+2} - \tfrac{1}{2}\rho_2 (d\sigma_{22} + d\sigma_{s+2,s+2}) \quad \text{.................(5·21).}$$

Squaring (5·20), taking the expectation, using the fundamental formula (5·14), and finally substituting the canonical values (5·18), we have

$$\left.\begin{aligned} nE(d\rho_1)^2 &= \sigma_{11}\sigma_{s+1,s+1} + \sigma_{1,s+1}^2 - \rho_1 (2\sigma_{11}\sigma_{1,s+1} + 2\sigma_{s+1,s+1}\sigma_{1,s+1}) \\ &\quad + \tfrac{1}{4}\rho_1^2 (2\sigma_{11}^2 + 4\sigma_{1,s+1}^2 + 2\sigma_{s+1,s+1}^2) \\ &= 1 + \rho_1^2 - \rho_1 (2\rho_1 + 2\rho_1) + \tfrac{1}{4}\rho_1^2 (2 + 4\rho_1^2 + 2) \\ &= (1 - \rho_1^2)^2 \end{aligned}\right\} \quad \text{(5·22).}$$

Treating the product of (5·20) and (5·21) in the same way we obtain

$$E d\rho_1 d\rho_2 = 0 \quad \text{...................................(5·23).}$$

A sample canonical correlation r_1 may be expanded about ρ_1 in a Taylor series of the form

$$r_1 = \rho_1 + \Sigma\Sigma \frac{\partial \rho_1}{\partial \sigma_{AB}} d\sigma_{AB} + \tfrac{1}{2}\Sigma\Sigma\Sigma\Sigma \frac{\partial^2 \rho_1}{\partial \sigma_{AB}\partial \sigma_{CD}} d\sigma_{AB}\, d\sigma_{CD} + \dots \quad \text{...(5·24),}$$

or, by the last of (5·16),

$$r_1 - \rho_1 = d\rho_1 + \dots \quad \text{..........................(5·25).}$$

The expectation of the product of any number of the sampling deviations $d\sigma_{AB}$ is a fixed function of the σ's divided by a power of n whose exponent increases with the number of the quantities $d\sigma_{AB}$ in the product. Since $E d\sigma_{AB} = 0$, we have from (5·24) and (5·14) that $E(r_1 - \rho_1)$ is of order n^{-1}. Hence squaring (5·25) and using (5·22), we find that the sampling variance of r_1 is given by $\frac{(1-\rho_1^2)^2}{n}$, apart from terms of higher order in n^{-1}. If by the standard error of r_1 we understand the leading term in the asymptotic expansion of the square root of the variance, we have for this standard error

$$\sigma_{r_1} = \frac{1 - \rho_1^2}{\sqrt{n}} \quad \text{................................(5·26).}$$

It is remarkable that this standard error of a canonical correlation is of exactly the same form as that of a product-moment correlation coefficient calculated directly from data, at least so far as the leading term is concerned.

The covariance of two statistics or their correlation would ordinarily be of order n^{-1}; but from (5·23) it appears that the covariance of r_1 and r_2 is of order n^{-2} at least. All these results hold as between any pair of simple non-vanishing roots. To summarize:

Let $\rho_1, \rho_2, \dots, \rho_p$ be any set of simple non-vanishing roots of (3·6). *For sufficiently large samples these will be approximated by certain of the canonical correlations $r_1, r_2, \dots, r_p$ of the samples in such a way that, when $r_\gamma - \rho_\gamma$ is divided by the standard error*

$$\sigma_{r_\gamma} = \frac{1 - \rho_\gamma^2}{\sqrt{n}} \quad (\gamma = 1, 2, \dots, p) \dots \quad \text{.............(5·27),}$$

the resulting variates have a distribution which, as n increases, approaches the normal distribution of p independent variates of zero means and unit standard deviations.

For small samples there will be ambiguities as to which root of the determinantal equation for the sample is to be regarded as approximating a particular canonical correlation of the population. As n increases, the sample roots will separately cluster more and more definitely about individual population roots.

If a canonical correlation ρ_γ is zero, and if $s = t$, the foregoing result is applicable with the qualification that sample values r_γ approximating ρ_γ must not all be taken positive, but must be assigned positive and negative values with equal probabilities. Alternatively, if we insist on taking all the sample canonical correlations as positive, the distribution will be that of absolute values of a normally distributed variate.

To prove this, suppose that the determinantal equation has zero as a double root. For sample covariances sufficiently near those in the population, there will be a root r close to zero, which will be very near the value of λ obtained by dropping from the equation all but the term in λ^2 and that independent of λ. The latter is for $s = t$ a perfect square, and the former does not vanish, since the zero root is only a double one. Hence r is the ratio of a polynomial in the s_{AB}'s to a non-vanishing regular function in the neighbourhood. This means that the differential method applicable to non-vanishing roots is also valid here, and that, since the derivatives are continuous, (5·27) holds even when $\rho_\gamma = 0$.

Since a tetrad difference is proportional to a vector correlation, which is the product of the canonical correlations, the question whether the tetrad differs significantly from zero is equivalent to the question whether a canonical correlation is significantly different from zero. This may be tested by means of the standard error (5·27), which reduces in this case to $\frac{1}{\sqrt{n}}$. Since this is independent of unknown parameters, we have here a method of meeting the third of the difficulties mentioned in Section 1 in connection with testing the significance of the tetrad.

For $s = 2$, a zero root is of multiplicity t at least. From the final result in § 9 below it may be deduced that if zero is a root of multiplicity exactly t, if r is the corresponding sample canonical correlation, and if $s = 2$, then nr^2 has the χ^2 distribution with $t - 1$ degrees of freedom. This provides a means of testing the significance of a sample canonical correlation in all cases in which $s = 2$.

We shall conclude this section by deriving standard error formulae for the vector correlation and vector alienation coefficients, assuming the canonical correlations in the population all distinct. Differentiating (4·5) and supposing all canonical correlations positive we have

$$dQ = \Sigma \frac{d\rho_\gamma}{\rho_\gamma}, \quad dZ = -2Z\Sigma \frac{\rho_\gamma d\rho_\gamma}{1 - \rho_\gamma^2}.$$

Taking the expectations of the squares and products of these expressions and using

(5·22) and (5·23), we obtain for the variances and covariance, apart from terms of higher order in n^{-1},

$$\sigma_q = Q\sqrt{\frac{1}{n}\sum_{\gamma=1}^{s}\frac{(1-\rho_\gamma^2)^2}{\rho_\gamma^2}}, \quad \sigma_z = 2Z\sqrt{\frac{\rho_1^2+\ldots+\rho_s^2}{n}},$$

$$EdQ\,dZ = -\frac{2}{n}QZ\Sigma(1-\rho_\gamma^2) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5{\cdot}28).$$

For the case $s=2$ these formulae reduce with the help of (4·5) to

$$\sigma_q = \sqrt{\frac{(1-Q^2)^2 - Z(1+Q^2)}{n}}, \quad \sigma_z = 2Z\sqrt{\frac{1-Z+Q^2}{n}},$$

$$EdQ\,dZ = -\frac{2}{n}QZ(1+Z-Q^2).$$

6. *Examples, and an Iterative Method of Solution.* The correlations obtained by Truman L. Kelley* among tests in (1) reading speed, (2) reading power, (3) arithmetic speed, and (4) arithmetic power are given by the elements of the following determinant, in which the rows and columns are arranged in the order given:

$$D = \begin{vmatrix} 1{\cdot}0000 & {\cdot}6328 & {\cdot}2412 & {\cdot}0586 \\ {\cdot}6328 & 1{\cdot}0000 & -{\cdot}0553 & {\cdot}0655 \\ {\cdot}2412 & -{\cdot}0553 & 1{\cdot}0000 & {\cdot}4248 \\ {\cdot}0586 & {\cdot}0655 & {\cdot}4248 & 1{\cdot}0000 \end{vmatrix} = {\cdot}4129.$$

These correlations were obtained from a sample of 140 seventh-grade school children. Let us inquire into the relations of arithmetical with reading abilities indicated by these tests.

The two-rowed minors of D in the upper left, lower right, and upper right corners are respectively

$$A = {\cdot}5996, \quad B = {\cdot}8195, \quad \sqrt{C} = {\cdot}01904.$$

Hence, by (4·2),

$$q^2 = {\cdot}0007377, \quad q = {\cdot}027161, \quad z = {\cdot}84036 \quad \ldots\ldots\ldots\ldots\ldots(6{\cdot}1).$$

By means of (4·9) or (4·11) these values give for the canonical correlations

$$r_1 = {\cdot}3945, \quad r_2 = {\cdot}0688 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6{\cdot}2).$$

In this case $n = N-1 = 139$, and the standard error (5·27) reduces, for the hypothesis of a zero canonical correlation in the population, to $\frac{1}{\sqrt{139}} = {\cdot}0848$. It is plain, therefore, that r_2 is not significant, so that we do not have any evidence here of more than one common component of reading and arithmetical abilities.

Whether we have convincing evidence of *any* common component is another question. It is tempting to compare the value of r_1 also with the standard error ·0848 for the purpose of answering this question, which would give a decidedly significant value. This however is not a sensitive procedure for testing the hypothesis

* *Op. cit.*, p. 100. These are the raw correlations, not corrected for attenuation.

that there is no common factor; for this hypothesis of complete independence would mean that both canonical correlations would in the population be zero; they would therefore be a quadruple root of the fundamental equation, to which the standard error is not applicable. Other tests for complete independence will be considered in Section 11; these have a sound basis, and one of them (discovered by Wilks) gives approximately ·0001 as the probability of a value of z as small as or smaller than the value found above. We conclude that reading and arithmetic involve one common mental factor but, so far as these data show, only one.

Linear functions $a_1x_1 + a_2x_2$ and $b_3x_3 + b_4x_4$ having maximum correlation with each other may be used either to predict arithmetical from reading ability or vice versa. The coefficients will satisfy (3·4) and (3·5); when in these equations we substitute $r_1 = {\cdot}3945$ for λ and μ, and the given correlations for the covariances, and divide by $-\lambda = -{\cdot}3945$, we have

$$\begin{aligned} a_1 + {\cdot}6328a_2 - {\cdot}6114b_3 - {\cdot}1485b_4 &= 0, \\ {\cdot}6328a_1 + a_2 + {\cdot}1402b_3 - {\cdot}1660b_4 &= 0, \\ -{\cdot}6114a_1 + {\cdot}1402a_2 + b_3 + {\cdot}4248b_4 &= 0, \\ -{\cdot}1485a_1 - {\cdot}1660a_2 + {\cdot}4248b_3 + b_4 &= 0. \end{aligned}$$

The fourth equation must be dependent on the preceding three, so we ignore it except for a final checking. Replacing b_4 by unity we may solve the first three equations, which are symmetrical, by the usual least-square method. Thus we write the coefficients, without repetition, in the form

1·0000	·6328	− ·6114	− ·1485	·8729
	1·0000	·1402	− ·1660	1·6070
		1·0000	·4248	·9536

the last column consisting of the sums of the elements written or understood in the respective rows. The various divisions, multiplications and subtractions involved in solving the equations are applied to the elements in the rows, including those in the check column, which at every stage gives the sum of the elements written or understood in a row. In the array above, the coefficients of each equation begin in the first row and proceed downward to the diagonal, then across to the right, and this scheme is followed with the reduced set of equations obtained by eliminating an unknown, which is done in such a way as to preserve symmetry. This process yields finally the ratios

$$a_1 : a_2 : b_3 : b_4 = -2{\cdot}7772 : 2{\cdot}2655 : -2{\cdot}4404 : 1.$$

Therefore the linear functions of arithmetical and reading scores that predict each other most accurately are proportional to $-2{\cdot}7772x_1 + 2{\cdot}2655x_2$ and $-2{\cdot}4404x_3 + x_4$, respectively. It is for these weighted sums that the maximum correlation ·3945 is attained.

From the same individuals, Kelley obtained the correlations in the following table, in which the first two rows correspond to the arithmetic speed and power tests cited above, while the others are respectively memory for words, memory for meaningful symbols, and memory for meaningless symbols:

$$\left\| \begin{array}{cc|ccc} 1{\cdot}0000 & {\cdot}4248 & {\cdot}0420 & {\cdot}0215 & {\cdot}0573 \\ {\cdot}4248 & 1{\cdot}0000 & {\cdot}1487 & {\cdot}2489 & {\cdot}2843 \\ \hline {\cdot}0420 & {\cdot}1487 & 1{\cdot}0000 & {\cdot}6693 & {\cdot}4662 \\ {\cdot}0215 & {\cdot}2489 & {\cdot}6693 & 1{\cdot}0000 & {\cdot}6915 \\ 0573 & {\cdot}2843 & {\cdot}4662 & {\cdot}6915 & 1{\cdot}0000 \end{array} \right\|$$

From this we find $q^2 = {\cdot}0003209, \quad q = {\cdot}01792, \quad z = {\cdot}902466,$

whence $r_1 = {\cdot}3073, \quad r_2 = {\cdot}0583.$

Since in this case $s \neq t$, we cannot say as before that the standard error of r_2 when $\rho = 0$ is $n^{-\frac{1}{2}} = {\cdot}0848$. But, putting $\chi^2 = nr_2^2 = {\cdot}472$, with two degrees of freedom, we find $P = {\cdot}79$, so that r_2 is far from significant. However r_1 is decidedly significant.

In view of the tests in Section 11, we conclude in this case also that there is evidence of one common component but not of two.

If each of the two sets contains more than two variates, the two invariants q and z do not suffice to determine the coefficients of the various powers of λ in the determinantal equation, so that its roots can no longer be calculated in the foregoing manner. The coefficients in the equation will involve other rational invariants in addition to q and z, but we shall not be concerned with these, and it is desirable to have a procedure that does not require their calculation, or the explicit determination and solution of the equation. It is also desirable to avoid the explicit solution of the sets of linear equations (3·4) and (3·5) when the variates are numerous, since the labour of the direct procedure then becomes excessive. These computational difficulties are analogous to those in the determination of the principal axes of a quadric in n-space, or of the principal components of a set of statistical variates, problems for which an iterative procedure has been found useful, and has been proved to converge to the correct values in all cases*. We shall now show how a process partly iterative in character may be applied to determine canonical variates and canonical correlations between two sets.

If in the s equations (3·4) we regard $\lambda a_1, \lambda a_2, \ldots, \lambda a_s$ as the unknowns, we may solve for them in terms of the b's by the methods appropriate for solving normal equations. Indeed, the matrix of the coefficients of the unknowns is symmetrical; and in the solving process it is only necessary to carry along, instead of a single column of right-hand members, t columns, from which the coefficients of $b_{s+1}, \ldots, b_{s+t}$ in the expressions for $a_1, \ldots, a_s$ are to be determined. The entries initially placed

* Harold Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components" in *Journal of Educational Psychology*, Vol. xxiv. pp. 417—441 and 498—520 (September and October, 1933), Section 4.

in these columns are of course the covariances between the two sets. Let the solution of these equations consist of the s expressions

$$\lambda a_\alpha = \sum_i g_{\alpha i} b_i \quad (\alpha = 1, 2, \ldots, s) \quad \ldots\ldots\ldots\ldots(6\cdot3).$$

In exactly the same way the t equations (3·5), with μ replaced by λ, may be solved for $\lambda b_{s+1}, \ldots, \lambda b_{s+t}$ in the form

$$\lambda b_i = \sum_\beta h_{i\beta} a_\beta \quad (i = s+1, \ldots, s+t) \ldots\ldots\ldots(6\cdot4).$$

If we substitute from (6·4) in (6·3) and set

$$k_{\alpha\beta} = \sum_i g_{\alpha i} h_{i\beta} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6\cdot5),$$

we have

$$\lambda^2 a_\alpha = \sum_\beta k_{\alpha\beta} a_\beta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6\cdot6).$$

Now if an arbitrarily chosen set of numbers be substituted for $a_1, \ldots, a_s$ in the right-hand members of (6·6), the sums obtained will be proportional to the numbers substituted only if they are proportional to the true values of $a_1, \ldots, a_s$. If, as will usually be the case, the proportionality does not hold, the sums obtained, multiplied or divided by any convenient constant, may be used as second approximations to solutions $a_1, \ldots, a_s$ of the equations. Substitution of these second approximations in the right-hand members of (6·6) gives third approximations which may be treated in the same way; and so on. Repetition of this process gives repeated sets of trial values, whose ratios will be seen below to approach as limits those among the true values of $a_1, \ldots, a_s$. The factor of proportionality λ^2 in (6·6) becomes r_1^2, the square of the largest canonical correlation. When the quantities $a_1', \ldots, a_s'$ eventually determined as sufficiently nearly proportional to $a_1, \ldots, a_s$ are substituted in the right-hand members of (6·4), there result quantities $b_{s+1}', \ldots, b_{s+t}'$ proportional to $b_{s+1}, \ldots, b_{s+t}$, apart from errors which may be made arbitrarily small by continuation of the iterative process. The factor of proportionality to be applied in order to obtain linear functions with unit variance is the same for the a's and the b's; from (3·2), (3·4), and (3·5) it may readily be shown that if from the quantities obtained we calculate

$$m = \frac{r_1}{\sqrt{\sum\sum \sigma_{\alpha i} a_\alpha' b_i'}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6\cdot7),$$

then the true coefficients of the first pair of canonical variates are $ma_1', \ldots, ma_s'$, $mb_{s+1}', \ldots, mb_{s+t}'$.

In the iterative process, if $a_1, \ldots, a_s$ represent trial values at any stage, those at the next stage will be proportional to

$$a_\alpha' = \sum k_{\alpha\beta} a_\beta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6\cdot8).$$

Another application of the process gives

$$a_\gamma'' = \sum k_{\gamma\alpha} a_\alpha',$$

whence, substituting, we have

$$a_\gamma'' = \sum k_{\gamma\beta}^{(2)} a_\beta,$$

provided we put

$$k_{\gamma\beta}^{(2)} = \sum_\alpha k_{\gamma\alpha} k_{\alpha\beta}.$$

The last equation is equivalent to the statement that the matrix K^2 of the coefficients $k_{\gamma\beta}^{(2)}$ is the square of the matrix K of the $k_{\alpha\beta}$. It follows therefore that one application of the iterative process by means of the squared matrix is exactly equivalent to two successive applications with the original matrix. This means that if at the beginning we square the matrix only half the number of steps will subsequently be required for a given degree of accuracy.

The number of steps required may again be cut in half if we square K^2, for with the resulting matrix K^4 one iteration is exactly equivalent to four with the original matrix. Squaring again we obtain K^8, with which one iteration is equivalent to eight, and so on. This method of accelerating convergence is also applicable to the calculation of principal components*. It embodies the root-squaring principle of solving algebraic equations in a form specially suited to determinantal equations.

After each iteration it is advisable to divide all the trial values obtained by a particular one of them, say the first, so as to make successive values comparable. The value obtained for a_1, if this is the one used to divide the rest at each step, will approach r_1^2 if the matrix K is used in iteration, but will approach r_1^4 if K^2 is used, r_1^8 if K^4 is used, and so forth. When stationary values are reached, they may well be subjected once to iteration by means of K itself, both in order to determine r_1^2 without extracting a root of high order, and as a check on the matrix-squaring operations.

If our covariances are derived from a sample from a continuous multivariate distribution, it is infinitely improbable that the equation in ω,

$$\begin{vmatrix} k_{11}-\omega & k_{12} & \ldots\ldots & k_{1s} \\ k_{21} & k_{22}-\omega & \ldots\ldots & k_{2s} \\ \ldots & \ldots & \ldots & \ldots \\ k_{s1} & k_{s2} & \ldots\ldots & k_{ss}-\omega \end{vmatrix} = 0,$$

has multiple roots. If we assume that the roots $\omega_1, \omega_2, \ldots, \omega_s$ are all simple, and regard $a_1, \ldots, a_s$ as the homogeneous coordinates of a point in $s-1$ dimensions which is moved by the collineation (6·8) into a point $(a_1', \ldots, a_s')$, we know† that there exists in this space a transformed system of coordinates such that the collineation is represented in terms of them by

$$\bar{a}_1' = \omega_1 \bar{a}_1, \quad \bar{a}_2' = \omega_2 \bar{a}_2, \quad \ldots, \quad \bar{a}_s' = \omega_s \bar{a}_s.$$

Another iteration yields a point whose transformed homogeneous coordinates are proportional to

$$\omega_1^2 \bar{a}_1, \quad \omega_2^2 \bar{a}_2, \quad \ldots, \quad \omega_s^2 \bar{a}_s.$$

Continuation of this process means, if ω_1 is the root of greatest absolute value, that

* Another method of accelerated iterative calculation of principal components is given by T. L. Kelley in *Essential Traits of Mental Life*, Cambridge, Mass., 1935. A method similar to that given above is applied to principal components by the author in *Psychometrika*, Vol. I. No. 1 (1936).

† Bôcher, p. 293.

the ratio of the first transformed coordinates to any of the others increases in geometric progression. Consequently the moving point approaches as a limit the invariant point corresponding to this greatest root. Therefore the ratios of the trial values of $a_1, \ldots, a_s$ will approach those among the coefficients in the expression for the canonical variate corresponding to the greatest canonical correlation. Thus the iterative process is seen to converge, just as in the determination of principal components.

After the greatest canonical correlation and the corresponding canonical variates are determined, it is possible to construct a new matrix of covariances of deviations from these canonical variates. When the iterative process is applied to this new matrix, the second largest canonical correlation and the corresponding canonical variates are obtained. This procedure may be carried as far as desired to obtain additional canonical correlations and variates, as in the method of principal components; but the later stages of the process will yield results which will usually be of diminishing importance. The modification of the matrix is somewhat more complicated than in the case of principal components, and we shall omit further discussion of this extension.

The process of obtaining iteratively the greatest canonical correlation, the most predictable criterion, and the best predicter may be illustrated if we imagine that, with three variates in each set, we have obtained from a sample the matrix of correlations

$$\left\| \begin{array}{ccc:ccc} 1\cdot0 & \cdot7 & \cdot1 & \cdot5 & \cdot4 & \cdot2 \\ \cdot7 & 1\cdot0 & \cdot1 & \cdot4 & \cdot3 & \cdot5 \\ \cdot1 & \cdot1 & 1\cdot0 & \cdot2 & \cdot2 & \cdot4 \\ \hdashline \cdot5 & \cdot4 & \cdot2 & 1\cdot0 & \cdot8 & \cdot6 \\ \cdot4 & \cdot3 & \cdot2 & \cdot8 & 1\cdot0 & \cdot7 \\ \cdot2 & \cdot5 & \cdot4 & \cdot6 & \cdot7 & 1\cdot0 \end{array} \right\|.$$

From the first three rows we obtain the set of normal equations indicated by

$$\begin{array}{ccccccc} 1\cdot0 & \cdot7 & \cdot1 & \cdot5 & \cdot4 & \cdot2 & 2\cdot9 \\ & 1\cdot0 & \cdot1 & \cdot4 & \cdot3 & \cdot5 & 3\cdot0 \\ & & 1\cdot0 & \cdot2 & \cdot2 & \cdot4 & 2\cdot0 \end{array}$$

Here the second and third rows are understood to be filled out with unwritten terms in such a way as to make the matrix consisting of the first three columns symmetric. The entries in the last column are the sums of those written or understood in the respective rows preceding them. By linear operations on the rows, equivalent to solving the equations, they are reduced to

$$\begin{array}{ccccccc} 1 & & & \cdot423 & \cdot362 & -\cdot316 & 1\cdot470 \\ & 1 & & \cdot089 & \cdot031 & \cdot685 & 1\cdot804 \\ & & 1 & \cdot149 & \cdot161 & \cdot362 & 1\cdot671 \end{array}$$

This is the numerical equivalent of (6·3). Hence g_{ai} is the element in the αth row and ith column of the matrix

$$G = \begin{Vmatrix} \cdot 423 & \cdot 362 & -\cdot 316 \\ \cdot 089 & \cdot 031 & \cdot 685 \\ \cdot 149 & \cdot 161 & \cdot 362 \end{Vmatrix}.$$

From the last three columns of the given matrix of correlations we obtain likewise the normal equations indicated by

$$\begin{matrix} 1\cdot 0 & \cdot 8 & \cdot 6 & \cdot 5 & \cdot 4 & \cdot 2 & 3\cdot 5 \\ & 1\cdot 0 & \cdot 7 & \cdot 4 & \cdot 3 & \cdot 2 & 3\cdot 4 \\ & & 1\cdot 0 & \cdot 2 & \cdot 5 & \cdot 4 & 3\cdot 4 \end{matrix}$$

The three columns before the check column appear in the same order as in the lower left corner of the matrix of correlations. The solutions of these equations, corresponding to (6·4), are the elements of the matrix

$$H = \begin{Vmatrix} \cdot 522 & \cdot 385 & \cdot 054 \\ \cdot 121 & -\cdot 385 & -\cdot 199 \\ -\cdot 198 & \cdot 539 & \cdot 507 \end{Vmatrix},$$

in which $h_{i\beta}$ is the element in the ith row and βth column. Upon multiplying the rows of G by the columns of H we find that $k_{\alpha\beta}$, defined by (6·5), is in the αth row and βth column of

$$K = GH = \begin{Vmatrix} \cdot 327 & -\cdot 147 & -\cdot 209 \\ -\cdot 085 & \cdot 392 & \cdot 346 \\ \cdot 026 & \cdot 190 & \cdot 160 \end{Vmatrix}.$$

The check columns are used to verify the calculation to this point, and may be used also at the next stage, which is to compute, by multiplying the rows of K by its columns,

$$K^2 = \begin{Vmatrix} \cdot 114 & -\cdot 145 & -\cdot 153 \\ -\cdot 052 & \cdot 232 & \cdot 209 \\ -\cdot 003 & \cdot 101 & \cdot 086 \end{Vmatrix},$$

and in the same way,

$$K^4 = \begin{Vmatrix} \cdot 021 & -\cdot 066 & -\cdot 061 \\ -\cdot 019 & \cdot 082 & \cdot 074 \\ -\cdot 006 & \cdot 033 & \cdot 029 \end{Vmatrix}.$$

The iteration process may now be begun with the trial values 1, 1, 1; when this set (which may be regarded as a vector) is multiplied by the rows of K^4 there results simply the set of sums of rows, namely

$$-\cdot 106 \qquad \cdot 137 \qquad \cdot 056.$$

Dividing all three of these by the first we have

$$1\cdot 0 \qquad -1\cdot 3 \qquad -\cdot 5.$$

Multiplying this vector by the rows of K^4 gives

$$\cdot 137 \quad -\cdot 163 \quad -\cdot 063,$$

which upon division by the first becomes

$$1\cdot 00 \quad -1\cdot 19 \quad -\cdot 46.$$

Multiplication of this vector by the rows of K^4 and division by the first resulting element gives

$$1\cdot 00 \quad -1\cdot 18 \quad -\cdot 46,$$

which upon another repetition of the process recurs exactly to the two decimal places. We therefore return to the matrix K with these trial values; multiplying them by the rows of K, dividing by the first, and then repeating the process once, we have the values

$$\cdot 5968 \quad -\cdot 707 \quad -\cdot 272,$$

which, divided by the first, become

$$a_1' = 1, \quad a_2' = -1\cdot 187, \quad a_3' = -\cdot 456,$$

are stationary under further iterations, and are correct to three decimal places. The last divisor, $\cdot 5968$, is the square of the greatest canonical correlation, also correct to three places; hence $r_1 = \sqrt{\cdot 5968} = \cdot 773$. Substitution of a_1', a_2', a_3', in the right members of (6·4), which comes to the same thing as multiplication by the rows of H, yields

$$b_4' = \cdot 040, \quad b_5' = \cdot 669, \quad b_6' = -1\cdot 069.$$

Then from (6·7) we have $m = 1\cdot 016$; when this is multiplied by the values of a_α' and b_i' just found, there result the coefficients in the expressions for the leading canonical variates, namely

$$u_1 = 1\cdot 016x_1 - 1\cdot 217x_2 - \cdot 463x_3,$$
$$v_1 = \cdot 041x_4 + \cdot 680x_5 - 1\cdot 086x_6,$$

which have unit variances and the correlation $\cdot 773$.

7. *The Vector Correlation as a Product of Correlations or of Cosines.* We shall in this section define certain linear functions of the variates in each set, forming two sequences, of which the product of the correlations between corresponding members is the vector correlation q. This result will be used in Section 8 to obtain an exact sampling distribution of q. The resolution, though valid with respect to the population, needs for our purposes to be made with reference to a sample. We shall use the pseudo-observations defined in Section 5, but shall write the sample covariance (5·10) in the form

$$s_{LM} = \frac{Sx_Lx_M}{n} \qquad \text{.................................(7·1)},$$

where S stands for summation from 1 to n, the number of degrees of freedom, and where L and M stand for an arbitrary pair, equal or unequal, of the subscripts $1, 2, \ldots, s+t$.

The sequences of variates which we shall consider may be defined as follows. First, let $x_1' = x_1$. Then let x_α' $(\alpha = 2, 3, \ldots, s)$ be the difference between x_α and a least-square estimate of x_α in terms of $x_1, \ldots, x_{\alpha-1}$, all divided by such a constant that the variance of x_α' is unity. To define the other sequence, let x_{s+1}' be a linear function of $x_{s+1}, \ldots, x_{s+t}$ having maximum correlation with x_1; and let $x_{s+\beta}'$ $(\beta = 2, \ldots, s)$ be a linear function uncorrelated with $x_{s+1}', \ldots, x_{s+\beta-1}'$, and having maximum correlation with x_β'. All these are to have unit variances. For a sample we may set aside as infinitely improbable the possibility that any of these new variates should be indeterminate. Putting R_β for the correlation of x_β' with $x_{s+\beta}'$, we shall find that

$$q = R_1 R_2 \ldots R_s \qquad \text{.................................(7·2).}$$

The process will be more perspicuous in geometrical than in algebraic language because of the simplicity of the geometry associated with samples from normal distributions, and the remainder of this section will be in geometrical terms. In the space of n dimensions in which the pseudo-observations of a variate are the coordinates of a point, there is for each variate a spherically symmetrical distribution of probability density centred at the origin. Let X_L denote the point whose coordinates are the pseudo-observations $x_{L1}, x_{L2}, \ldots, x_{Ln}$ on the variate x_L $(L = 1, 2, \ldots, s+t)$. Let P_t be the flat space of t dimensions containing the origin O and the points $X_{s+1}, \ldots, X_{s+t}$ determined by the second set of variates. Perpendicular to OX_1 will be a flat space of $n-1$ dimensions, whose intersection with P_t will in general be of $t-1$ dimensions. Denote this intersection by P_{t-1}. Let P_{t-2} be the flat space of $t-2$ dimensions contained in P_{t-1} and perpendicular to OX_2; and so forth.

Let X_1' be the point on OX_1 of unit distance from the origin. We further define points $X_2', \ldots, X_s'$, all at unit distance from the origin, such that $OX_1', OX_2', \ldots, OX_s'$ are mutually perpendicular, and such that OX_2' is coplanar with OX_1 and OX_2; OX_3' is in the same flat 3-space with OX_1, OX_2 and OX_3; and so forth. The coordinates

$$\begin{matrix} x_{11}' & x_{12}' & \ldots & x_{1n}' \\ \cdots & \cdots & \cdots & \cdots \\ x_{s1}' & x_{s2}' & \ldots & x_{sn}' \end{matrix}$$

of these s points will thus satisfy

$$S x_\alpha' x_\beta' = \delta_{\alpha\beta} \qquad \text{.................................(7·3),}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta, equal to unity if $\alpha = \beta$ but otherwise zero, and where S stands for summation from 1 to n.

Let us also rotate the n axes so that the first t of them lie in P_t; and let an internal transformation be performed upon the t variates of the second set such that, for this particular sample, the coordinates representing them become

$$\left\| \begin{matrix} 1 & 0 & 0 & & \ldots & 0 \\ 0 & 1 & 0 & & \ldots & 0 \\ 0 & 0 & 1 & & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 1 \ldots & 0 \end{matrix} \right\| \qquad \text{..............................(7·4).}$$

None of these transformations affects the value of q, and we have, adapting the definition (4·2) to this case, by replacing the covariances by the functions (7·1) of $x_{11}', \ldots, x_{sn}'$ and of the elements of (7·4), and then multiplying each row of each determinant by n,

$$q^2 = (-1)^s \frac{C'}{A'B'},$$

where now $A' = B' = 1$, while

$$C' = \begin{vmatrix} 0 \ldots 0 & x_{11}' \;\ldots\ldots\; x_{1t}' \\ \ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ 0 \ldots 0 & x_{s1}' \;\ldots\ldots\; x_{st}' \\ x_{11}' \ldots x_{s1}' & 1 \quad 0 \;\ldots\; 0 \\ \ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ x_{1t}' \ldots x_{st}' & 0 \quad 0 \;\ldots\; 1 \end{vmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·5).$$

Now letting Σ stand for summation from 1 to t, we introduce determinants

$$D_\beta = \begin{vmatrix} \Sigma x_1'^2 & \ldots\ldots & \Sigma x_1' x_\beta' \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \Sigma x_\beta' x_1' & \ldots\ldots & \Sigma x_\beta'^2 \end{vmatrix} \quad (\beta = 1, 2, \ldots, s) \quad \ldots\ldots\ldots\ldots(7·6).$$

Upon expanding (7·5) with respect to the first s rows and columns we find, with the help of Section 2,

$$q^2 = (-1)^s C' = D_s \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·7).$$

Now any line perpendicular to OX_1 and OX_2 is perpendicular to all the lines in the plane of these two, in particular to OX_2'. Hence P_{t-2}, which consists of lines perpendicular to OX_1 and OX_2, is perpendicular to OX_2'. In like manner, P_{t-3} is perpendicular to OX_1', OX_2' and OX_3'; and in general $P_{t-\beta}$ is perpendicular to OX_1', OX_2', ..., OX_β'.

Since $P_{t-\beta}$ lies entirely within P_t, the coordinates of any point U in $P_{t-\beta}$ will be linearly dependent on the rows of (7·4), and so of the form

$$u_1, u_2, \ldots, u_t, 0, 0, \ldots, 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·8).$$

The orthogonality of OU to $OX_1', \ldots, OX_\beta'$ means that

$$\Sigma u x_\alpha' = 0 \quad (\alpha = 1, 2, \ldots, \beta) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·9).$$

Now let $\theta_{\beta+1}$ denote the angle that $OX_{\beta+1}'$ makes with $P_{t-\beta}$; that is, $\theta_{\beta+1}$ is the minimum angle of $OX_{\beta+1}'$ with a line OU such that the coordinates of U are of the form (7·8) and satisfy (7·9). Without loss of generality we may also take U at unit distance from the origin, so that

$$\Sigma u^2 = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·10).$$

Since $S x_{\beta+1}'^2 = 1$ by (7·3), we then have

$$\cos\theta_{\beta+1} = \Sigma u x_{\beta+1}' \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7·11).$$

To determine the minimum angle we therefore differentiate with respect to $u_1, \ldots, u_t$ the expression

$$\Sigma u x_{\beta+1}' - \tfrac{1}{2}\gamma \Sigma u^2 - \lambda_1 \Sigma u x_1' - \ldots - \lambda_\beta \Sigma u x_\beta'$$

where $\gamma, \lambda_1, \ldots, \lambda_\beta$ are Lagrange multipliers. This gives

$$\lambda_1 x_{1h}' + \ldots + \lambda_\beta x_{\beta h}' = x_{\beta+1,h}' - \gamma u_h \quad (h = 1, \ldots, t) \qquad \text{.........(7·12)}.$$

Multiply (7·12) by u_h and sum with respect to h. The left member disappears by (7·9), and from (7·10) and (7·11) we have

$$\gamma = \cos\theta_{\beta+1} \qquad \text{.........(7·13)}.$$

Upon multiplying (7·12) by $x_{\alpha h}'$, summing with respect to h, and using (7·9), we have, for $\alpha = 1, 2, \ldots, \beta$,

$$\lambda_1 \Sigma x_1' x_\alpha' + \ldots + \lambda_\beta \Sigma x_\beta' x_\alpha' = \Sigma x_\alpha' x_{\beta+1}' \qquad \text{.........(7·14)}.$$

Eliminating $\lambda_1, \ldots, \lambda_\beta$ from the $\beta+1$ equations (7·14) and (7·12) we have

$$\begin{vmatrix} \Sigma x_1'^2 & \cdots\cdots & \Sigma x_1' x_\beta' & \Sigma x_1' x_{\beta+1}' \\ \cdots & \cdots & \cdots & \cdots \\ \Sigma x_\beta' x_1' & \cdots\cdots & \Sigma x_\beta'^2 & \Sigma x_\beta' x_{\beta+1}' \\ x_{1h}' & \cdots\cdots & x_{\beta h}' & x_{\beta+1,h}' - \gamma u_h \end{vmatrix} = 0 \qquad \text{.........(7·15)}.$$

Multiply the last row of this determinant by $x_{\beta+1,h}'$ and sum with respect to h from 1 to t. The last element, with the help of (7·11) and (7·13), reduces to

$$\Sigma x_{\beta+1}'^2 - \gamma^2,$$

and so, from (7·6), we have

$$\gamma^2 = \frac{D_{\beta+1}}{D_\beta} \qquad \text{.........(7·16)}.$$

Hence, from (7·13),

$$\cos\theta_{\beta+1} = \sqrt{\frac{D_{\beta+1}}{D_\beta}} \quad (\beta = 1, 2, \ldots, s-1) \qquad \text{.........(7·17)}.$$

The cosine of the angle which OX_1 makes with P_t is

$$\cos\theta_1 = \sqrt{\Sigma x_1'^2} = \sqrt{D_1} \qquad \text{.........(7·18)}.$$

Multiplying together all the equations (7·17) and (7·18) and recalling (7·7), we obtain

$$q = \cos\theta_1 \cos\theta_2 \ldots \cos\theta_s \qquad \text{.........(7·19)}.$$

It is obvious that the correlations R_β defined at the beginning of this section have the property that

$$R_\beta = \cos\theta_\beta,$$

so that (7·19) is equivalent to (7·2).

8. *An Exact Sampling Distribution of q.* We shall now deduce the exact distribution of q in samples from a multivariate normal population in which the vector correlation is zero, for the case in which one of the sets consists of exactly two variates. From (4·5) it follows in this case that at least one of the two canonical correlations is zero. If the numbers of variates in both sets are 2, we have essentially the case of the tetrad difference; the distribution will then be symmetrical, since the population value is assumed to be zero. Let $\rho_2 = 0$, and for brevity put ν for ρ_1^2, which will be a parameter of the distribution.

The angle θ_1 defined in Section 7 between the line OX_1 determined by the sample values of the first variate and the flat space P_t determined by those of the second set has the property that $R_1 = \cos\theta_1$ is the multiple correlation of x_1 with the second set of variates. The population value of this multiple correlation is ρ_1. We assume all the variates subject to random sampling. In this case R_1 will have the "A" distribution discovered by R. A. Fisher*. In our notation, with samples of $n+1$ from which the means have been eliminated, or in samples of $n+k$ from which k degrees of freedom have been removed by least-squares elimination of other variates, the distribution of R_1 is

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{t}{2}\right)\Gamma\left(\frac{n-t}{2}\right)}(1-\nu)^{\frac{n}{2}}(R_1^2)^{\frac{t-2}{2}}(1-R_1^2)^{\frac{n-t-2}{2}}F\left(\frac{n}{2},\frac{n}{2},\frac{t}{2},\nu R_1^2\right)d(R_1^2) \qquad \text{......(8·1)},$$

with F denoting the hypergeometric function.

The points X_2' of Section 7 corresponding to an infinity of samples form a globular cluster having spherical symmetry with centre at the origin, in the flat space of $n-1$ dimensions perpendicular to OX_1. In this flat space is the space P_{t-1}, which makes with OX_2' the angle θ_2. Hence $R_2 = \cos\theta_2$ has the distribution of a multiple correlation coefficient in samples from an *uncorrelated* normal population, with $t-1$ "independent" variates. We replace n in (8·1) by $n-1$, t by $t-1$, R_1 by R_2, ν by zero, and have

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-t}{2}\right)}(R_2^2)^{\frac{t-3}{2}}(1-R_2^2)^{\frac{n-t-2}{2}}d(R_2^2) \qquad \text{.........(8·2)}.$$

From (7·2) we have $q = R_1R_2$. Hence put $R_2 = \frac{q}{R_1}$, $dR_2 = \frac{dq}{R_1}$ in (8·2), multiply by (8·1), and integrate with respect to R_1 from q to 1. This gives the distribution of q in the form

$$\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{t}{2}\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma^2\left(\frac{n-t}{2}\right)}(1-\nu)^{\frac{n}{2}}(q^2)^{\frac{t-3}{2}}d(q^2)$$

$$\times\int_{q^2}^1[(1-R^2)(R^2-q^2)]^{\frac{n-t-2}{2}}(R^2)^{\frac{-n+t+1}{2}}F\left(\frac{n}{2},\frac{n}{2},\frac{t}{2},\nu R^2\right)d(R^2) \qquad \text{...(8·3)},$$

where the subscript is dropped from the variable of integration. Now

$$\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{t}{2}\right)\Gamma\left(\frac{t-1}{2}\right)}=\frac{(n-2)!}{2^{n-t}(t-2)!}.$$

* "The General Sampling Distribution of the Multiple Correlation Coefficient" in *Proceedings of the Royal Society*, Vol. CXXI. A (1928), p. 660.

Making this substitution and changing the variable of integration to

$$x = \frac{1-R^2}{1-q^2},$$

we have therefore

$$\frac{(n-2)!}{2^{n-t}(t-2)!\,\Gamma^2\left(\frac{n-t}{2}\right)}(1-\nu)^{\frac{n}{2}}(1-q^2)^{n-t-1}(q^2)^{\frac{t-3}{2}}\,d(q^2)$$

$$\times\int_0^1 [x(1-x)]^{\frac{n-t-2}{2}}[1-(1-q^2)x]^{\frac{-n+t+1}{2}}F\left\{\frac{n}{2},\frac{n}{2},\frac{t}{2},\nu[1-(1-q^2)x]\right\}dx \quad\text{......(8·4).}$$

If it is supposed that the values of the variates in the second set are fixed, instead of varying normally from sample to sample, Fisher's distribution "C" of the multiple correlation coefficient should be used instead of his "A" distribution. This gives for q a new distribution, which for $s=2$ may be written in terms of a confluent hypergeometric function

$$\frac{\frac{1}{\sqrt{\pi}}\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-4}{2}\right)}e^{\frac{-\nu}{2}}(1-q^2)^{n-3}q\,dq\int_0^1[x(1-x)]^{\frac{n-4}{2}}[1-(1-q^2)x]^{\frac{-n+1}{2}}$$

$$\times F\left\{\frac{n}{2},1,\frac{1}{2}\frac{\nu q^2}{1-(1-q^2)x}\right\}dx \quad\text{......(8·5).}$$

However the conditions of sampling under which q is likely to be used are such that (8·4) appears to be the more important form, and we shall give no further consideration to (8·5).

By extension of the reasoning above the distribution of q may be found for larger values of s, provided all but one of the parameters $\rho_1, \dots, \rho_s$ vanish. Thus for $s=3$ we have, from (7·2),

$$q = R_1 R_2 R_3 = q' R_3,$$

where q' has the distribution (8·3), i.e. (8·4), while R_3 has the distribution obtained from (8·2) by replacing R_2 by R_3, n by $n-1$, and t by $t-1$. Combining this with (8·4) in the same manner that (8·2) was combined with (8·1) to produce (8·3), the new distribution of q is obtained. This process may be repeated to obtain the distribution for values of s as great as desired; but it must of course be remembered that $s \leqslant t$.

It is tempting to try to obtain the general distribution of q, without our assumption that all but one of the quantities $\rho_1, \dots, \rho_s$ are zero, by treating these as population values of the multiple correlations whose distributions are used successively in finding the distribution of q as above. However this suggested procedure appears to be incorrect. If $\rho_2 \neq 0$, the centre of the globular cluster formed by the projected X_2' points will have a centre which is not on P_{t-1}, where it should be if the multiple correlation distribution were to be amplified.

9. *Moments of q. The Distribution for Large Samples.* We shall derive the even moments of the distribution of Section 8, assuming that ν does not take

either of the extreme values 0 and 1. The latter is the case in which q becomes a partial correlation coefficient, the theory of which is well understood. The case $\nu=0$, corresponding to complete independence, is a very simple one, concerning which all information desired may be obtained from Section 11. The moments will be obtained by processes involving repeated interchanges of order of the processes of integration, differentiation, and summation of series. It will be observed that the uniform convergence and continuity required to justify these interchanges exist, provided ν is definitely *between* 0 and 1 without taking either of these values.

The odd moments of q about zero, which require no consideration unless $s=t$, vanish in this case when only one of the canonical correlations is different from zero, since the distribution of q must then be symmetrical. Let μ_{2k} be the $2k$th moment of q, which is also the kth moment of q^2, about zero. To determine its value, multiply (8·3) by q^{2k} and integrate with respect to q^2 from 0 to 1. In the double integral thus obtained, a reversal of the order of integration means that q^2 will vary from 0 to R^2, and then R^2 will vary from 0 to 1. The first integration in this new order may be effected at once, since upon putting $q^2=R^2z$, $d(q^2)=R^2dz$, we have

$$\int_0^{R^2}(q^2)^{\frac{t-3}{2}+k}(R^2-q^2)^{\frac{n-t-2}{2}}d(q^2)=(R^2)^{\frac{n-3}{2}+k}\frac{\Gamma\left(\frac{t-1}{2}+k\right)\Gamma\left(\frac{n-t}{2}\right)}{\Gamma\left(\frac{n-1}{2}+k\right)}.$$

Therefore

$$\mu_{2k}=\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t}{2}\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-t}{2}\right)\Gamma\left(\frac{n-1}{2}+k\right)}(1-\nu)^{\frac{n}{2}}$$

$$\times\int_0^1(R^2)^{\frac{t-2}{2}+k}(1-R^2)^{\frac{n-t-2}{2}}F\left(\frac{n}{2},\frac{n}{2},\frac{t}{2},\nu R^2\right)d(R^2).$$

Expanding the hypergeometric series and integrating term by term, since

$$\int_0^1(R^2)^{\frac{t-2}{2}+k+r}(1-R^2)^{\frac{n-t-2}{2}}d(R^2)=\frac{\Gamma\left(\frac{t}{2}+k+r\right)\Gamma\left(\frac{n-t}{2}\right)}{\Gamma\left(\frac{n}{2}+k+r\right)},$$

we obtain

$$\mu_{2k}=\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-1}{2}+k\right)}(1-\nu)^{\frac{n}{2}}\sum_{r=0}^{\infty}\frac{\Gamma^2\left(\frac{n}{2}+r\right)\Gamma\left(\frac{t}{2}+k+r\right)}{r!\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{t}{2}+r\right)\Gamma\left(\frac{n}{2}+k+r\right)}\nu^r \quad (9·1).$$

In this we make the substitution

$$\frac{\Gamma\left(\frac{n}{2}+r\right)}{\Gamma\left(\frac{n}{2}+k+r\right)}=\frac{1}{\Gamma(k)}\int_0^1 x^{\frac{n}{2}+r-1}(1-x)^{k-1}dx \quad \text{...........}(9·2);$$

upon reversing the order of integration and summation we have

$$\mu_{2k}=\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-1}{2}+k\right)\Gamma(k)}(1-\nu)^{\frac{n}{2}}$$

$$\times\int_0^1(1-x)^{k-1}\sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{n}{2}+r\right)\Gamma\left(\frac{t}{2}+k+r\right)}{r!\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{t}{2}+r\right)}x^{\frac{n}{2}+r-1}\nu^r dx.$$

Now make the further substitution

$$\frac{\Gamma\left(\frac{t}{2}+k+r\right)}{\Gamma\left(\frac{t}{2}+r\right)}=x^{-\frac{t}{2}-r+1}\frac{d^k}{dx^k}x^{\frac{t}{2}+k+r-1}\qquad\text{.................(9·3)}.$$

which gives, upon interchanging the order of summation and differentiation,

$$\mu_{2k}=\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-1}{2}+k\right)\Gamma(k)}(1-\nu)^{\frac{n}{2}}$$

$$\times\int_0^1 x^{\frac{n-t}{2}}(1-x)^{k-1}\frac{d^k}{dx^k}\left[x^{\frac{t}{2}+k-1}\sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{n}{2}+r\right)}{r!\,\Gamma\left(\frac{n}{2}\right)}x^r\nu^r\right]dx.$$

The sum is now a binomial expansion and

$$\mu_{2k}=\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n-1}{2}+k\right)\Gamma(k)}(1-\nu)^{\frac{n}{2}}$$

$$\times\int_0^1 x^{\frac{n-t}{2}}(1-x)^{k-1}\frac{d^k}{dx^k}\left[x^{\frac{t}{2}+k-1}(1-\nu x)^{-\frac{n}{2}}\right]dx\qquad\text{......(9·4)}.$$

From this form, by k successive integrations by parts, it is easy in any particular case to calculate μ_{2k}. Thus for $k=1$ we have

$$\mu_2=\frac{t-1}{n-1}(1-\nu)^{\frac{n}{2}}\int_0^1 x^{\frac{n-t}{2}}\frac{d}{dx}\left[x^{\frac{t}{2}}(1-\nu x)^{-\frac{n}{2}}\right]dx$$

$$=\frac{t-1}{n-1}(1-\nu)^{\frac{n}{2}}\left\{(1-\nu)^{-\frac{n}{2}}-\frac{n-t}{2}\int_0^1 x^{\frac{n}{2}-1}(1-\nu x)^{-\frac{n}{2}}dx\right\}$$

$$=\frac{t-1}{n-1}\left\{1-\frac{n-t}{n}(1-\nu)^{\frac{n}{2}}F\left(\frac{n}{2},\frac{n}{2},\frac{n}{2}+1,\nu\right)\right\}$$

Euler's transformation (10·1) of the hypergeometric reduces this to

$$\mu_2 = \frac{t-1}{n-1}\left\{1 - \frac{n-t}{n}(1-\nu)\, F\left(1, 1, \frac{n}{2}+1, \nu\right)\right\} \quad \text{............(9·5)},$$

in which the series can easily be calculated to any required accuracy. For large samples the convergence is extremely rapid. A series of powers of n^{-1} may also be obtained by expanding each term of the hypergeometric series:

$$\mu_2 = \frac{t-1}{n-1}\left\{\nu + (1-\nu)\left[\frac{t-2\nu}{n} + \frac{2(t+2)\nu - 8\nu^2}{n^2} + \dots\right]\right\} \quad \text{......(9·6)}.$$

This form brings out the manner in which, for large samples, μ_2 varies with ν.

The moments may alternatively be found by a slightly different method, giving a general result in which an integral does not appear explicitly. In the identity (9·3) let x be replaced by ν, and make both this substitution and (9·2) in (9·1). The result is

$$\mu_{2k} = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma(k)}\, \nu^{-\frac{t}{2}+1}\, (1-\nu)^{\frac{n}{2}}$$
$$\times \frac{d^k}{d\nu^k}\int_0^1 x^{\frac{n}{2}-1}(1-x)^{k-1}\sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{n}{2}+r\right)}{\Gamma\left(\frac{n}{2}\right) r!}\, x^r\, \nu^{\frac{t}{2}+k+r-1}\, dx.$$

The integral equals

$$\nu^{\frac{t}{2}+k-1}\int_0^1 x^{\frac{n}{2}-1}(1-x)^{k-1}(1-\nu x)^{-\frac{n}{2}}\, dx = \nu^{\frac{t}{2}+k-1}\left[\frac{\Gamma\left(\frac{n}{2}\right)\Gamma(k)}{\Gamma\left(\frac{n}{2}+k\right)}\right] F\left(\frac{n}{2}, \frac{n}{2}, \frac{n}{2}+k, \nu\right),$$

whence

$$\mu_{2k} = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{n}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)}\, \nu^{-\frac{t}{2}+1}\, (1-\nu)^{\frac{n}{2}}$$
$$\times \frac{d^k}{d\nu^k}\left[\nu^{\frac{t}{2}+k-1} F\left(\frac{n}{2}, \frac{n}{2}, \frac{n}{2}+k, \nu\right)\right] \text{......(9·7)}.$$

This may be made even more explicit by performing the differentiation with the help of Leibnitz' theorem; and Euler's transformation may be applied to each of the hypergeometric functions to give a rapidly convergent series. In this way we obtain

$$\mu_{2k} = \sum_{r=0}^{k}\frac{k!}{r!\,(k-r)!}\,\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)\Gamma\left(\frac{t}{2}+k\right)\Gamma^2\left(\frac{n}{2}+k-r\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{t}{2}+k-r\right)\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}+2k-r\right)}$$
$$\times\, \nu^{k-r}(1-\nu)^r\, F\left(k, k, \frac{n}{2}+2k-r, \nu\right) \text{......(9·8)}.$$

The expressions obtained from this by substituting particular values of k are different in form from those obtained directly from (9·4), but are reducible to them with the help of the Gauss relations between "neighbouring" hypergeometric functions.

From (9·7) it is easy to see that $\mu_0 = 1$, as it should; this checks a long chain of deductions.

The asymptotic value of μ_{2k} for large values of n will now be investigated. In the expression for the kth derivative obtained from (9·4) by Leibnitz' theorem, the term of highest order in n is

$$x^{\frac{t}{2}+k-1}\, \nu^k \frac{n}{2}\left(\frac{n}{2}+1\right)\ldots\left(\frac{n}{2}+k-1\right)(1-\nu x)^{-\frac{n}{2}-k},$$

whence

$$\mu_{2k} \sim \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma(k)}(1-\nu)^{\frac{n}{2}}\nu^k \frac{\Gamma\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$\times \int_0^1 x^{\frac{n}{2}+k-1}(1-x)^{k-1}(1-\nu x)^{-\frac{n}{2}-k}\,dx$$

$$= \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)\Gamma^2\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n}{2}+2k\right)\Gamma\left(\frac{n}{2}\right)}\nu^k(1-\nu)^{\frac{n}{2}} F\left(\frac{n}{2}+k, \frac{n}{2}+k, \frac{n}{2}+2k, \nu\right)$$

$$= \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{t-1}{2}+k\right)\Gamma^2\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}+k\right)\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{n}{2}+2k\right)\Gamma\left(\frac{n}{2}\right)}\nu^k F\left(k, k, \frac{n}{2}+2k, \nu\right)$$

$$\sim \frac{\Gamma\left(\frac{t-1}{2}+k\right)}{\Gamma\left(\frac{t-1}{2}\right)}\left(\frac{2\nu}{n}\right)^k \quad\ldots\ldots\ldots\ldots(9\cdot9).$$

Hence, for $s = t = 2$, the distribution of q approaches the normal form, with variance $\frac{\nu}{n}$, as is seen either from (9·9) or from (9·4), and mean value zero.

For $t \neq 2$, the distribution of q does not behave in this way. As in the case of multiple correlation, it is then confined to positive values. An approximation to the distribution is however suggested by the foregoing asymptotic values of the moments. These are in fact the moments of the χ^2 distribution with $t-1$ degrees of freedom, if we put

$$\chi^2 = \frac{nq^2}{\nu}.$$

The approximate distributions thus obtained may tentatively be used for testing the significance of q in large samples when ν has a value not too close to zero. For small samples and small values of ν, the methods of the next section are appropriate.

10. *The Distribution for Small Samples**. *Form of the Frequency Curve.* The distribution of Section 8 may in certain cases be expressed in elementary forms. If $n-t$ is even, the Euler transformation

$$F(a, b, c, x) = (1-x)^{c-a-b} F(c-a, c-b, c, x) \quad \ldots\ldots\ldots\ldots (10{\cdot}1)$$

may be applied to the hypergeometric function in the distribution to give a terminating series, and the integration can then be carried out for each term, the integrand in (8·3) being a rational function of R, or involving (if t is odd) a single quadratic surd.

Consider for example the simple case $t=2$, $n=4$. Since $t=s$, it is more convenient to work with the distribution of q than with that of q^2. We halve the numerical coefficient because negative and positive sample values are distinguishable. The distribution (8·3) becomes, for positive values of q,

$$(1-\nu)^2\, dq \int_q^1 F(2, 2, 1, \nu R^2)\, dR.$$

Using (10·1), expanding $F(-1, -1, 1, R^2)$, and putting $\nu = \rho^2$, this becomes

$$(1-\rho^2)^2\, dq \int_q^1 (1-\rho^2 R^2)^{-3} (1+\rho^2 R^2)\, dR,$$

or, carrying out the integration,

$$\frac{(1-\rho^2)^2}{8\rho}\left\{\log \frac{1-\rho q}{1-\rho}\, \frac{1+\rho}{1+\rho q} - \frac{2\rho q\,(3-\rho^2 q^2)}{(1-\rho^2 q^2)^2} + \frac{2\rho\,(3-\rho^2)}{(1-\rho^2)^2}\right\} dq.$$

Other special cases of the distribution function may be obtained by integrating the distributions obtained by R. A. Fisher for the multiple correlation coefficient for even values of n†.

A more systematic development, not depending on the oddness or evenness of n or t, and valuable when n and ν are not too great, is obtained by expanding the hypergeometric series in powers of ν and carrying out the integration term by term. Applying this procedure to (8·4) we obtain integrals of the form

$$\int_0^1 [x(1-x)]^{\frac{n-t-2}{2}} [1-(1-q^2)x]^{\frac{-n+t+1}{2}+k}\, dx \quad (k=0, 1, 2, \ldots).$$

But this is itself a hypergeometric integral, and equals

$$\frac{\Gamma^2\left(\frac{m}{2}\right)}{(m-1)!} F\left(\frac{m-1}{2} - k, \frac{m}{2}, m, 1-q^2\right),$$

where for brevity we put $\qquad m = n-t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10{\cdot}2).$

* The smallest samples to which the text is applicable are those for which $n=s+t$. For smaller samples, the matrix of pseudo-observations has more rows than columns; consequently there is a linear relation among the rows, i.e. among the sample variates, whose number is thus in effect reduced, so that a simpler theory is adequate. Thus, if $s=t=2$ and $n=3$, q reduces to a partial correlation coefficient, whose distribution is known.

† *Op. cit.*, p. 661.

Introducing the functions

$$v_k = \left(\frac{1+q}{2}\right)^{m-1} F\left(\frac{m-1}{2} - k,\ \frac{m}{2},\ m,\ 1-q^2\right) \quad \text{...........(10·3)},$$

the distribution (8·4) may thus be written, for $q > 0$,

$$\frac{(n-2)!}{(t-2)!\,(m-1)!}(1-\nu)^{\frac{n}{2}}(1-q)^{m-1}q^{t-2}dq \sum_{k=0}^{\infty} \frac{\Gamma^2\left(\frac{n}{2}+k\right)\Gamma\left(\frac{t}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)\Gamma\left(\frac{t}{2}+k\right)k!} v_k \nu^k$$
$$\text{......(10·4)}.$$

A factor $\frac{1}{2}$ must be applied to this expression in the case $t=2$ if we then distinguish negative from positive values of q.

The first of the "relationes inter functiones contiguas" of Gauss* is

$$F(a-1, b, c, x) = \frac{[c - 2a + (a-b)x]\,F(a, b, c, x) + a(1-x)\,F(a+1, b, c, x)}{c-a}.$$

Putting $\qquad a = \dfrac{m-1}{2} - k,\quad b = \dfrac{m}{2},\quad c = m,\quad x = 1-q^2,$

this shows that (10·3) satisfies the linear difference equation

$$v_{k+1} = \frac{(2k+1)(1+q^2)\,v_k + (m-2k-1)\,q^2 v_{k-1}}{m+2k+1} \quad \text{...........(10·5)},$$

which may be used as a recurrence relation for computing the successive v_k's as soon as we have determined two values whose indices differ by unity.

From the identity of Gauss (*ibid.*, p. 227)

$$F(\alpha, \beta, \alpha+\beta+\tfrac{1}{2}, x) = F\left(2\alpha, 2\beta, \alpha+\beta+\tfrac{1}{2}, \frac{1-\sqrt{1-x}}{2}\right)$$

we find at once

$$F\left(\frac{m-1}{2}, \frac{m}{2}, m, 1-q^2\right) = F\left(m-1, m, m, \frac{1-q}{2}\right) \quad \text{........(10·6)}.$$

In the series expansion of this last function, numerator and denominator factors cancel in each term in such a way as to leave a binomial expansion

$$\left(1 - \frac{1-q}{2}\right)^{-m+1} = \left(\frac{2}{1+q}\right)^{m-1}$$

In this way we have, from (10·3) and (10·6),

$$v_0 = 1 \quad \text{...................................(10·7)}.$$

In (10·3) put $k = -1$, and apply (10·1). The result reduces, with the help of (10·6), to

$$v_{-1} = q^{-1}.$$

We have thus obtained two consecutive values of v_k, from which the rest are successively determined by means of (10·5), a relation which may also be written

$$v_{k+1} = q^2 v_{k-1} + \frac{2k+1}{m+2k+1}\{(1+q^2)\,v_k - 2q^2 v_{k-1}\} \quad \text{...........(10·8)}.$$

* *Werke*, Vol. III. p. 130.

We thus find in turn

$$v_1 = q + \frac{(1-q)^2}{m+1}, \quad v_2 = q^2 + \frac{3}{m+3}(1-q)^2\left(q + \frac{1+q^2}{m+1}\right), \quad \ldots.$$

It is easy to show by means of the recurrence relation that the limit of v_k as m increases is q^k, and that the remaining terms of v_k constitute a polynomial in q having the factor $(1-q)^2$.

For a test of significance of q it is necessary to integrate the distribution from an arbitrary value to unity. For this purpose it is convenient to put $p = 1 - q$. In terms of p,

$$v_1 = 1 - p + \frac{p^2}{m+1},$$

$$v_2 = 1 - 2p + \left(1 + \frac{3}{m+1}\right)p^2 - \frac{3}{m+1}p^3 + \frac{3}{(m+1)(m+3)}p^4,$$

$$v_3 = 1 - 3p + 3\frac{m+3}{m+1}p^2 - \frac{m+13}{m+1}p^3 + \frac{3(2m+11)}{(m+1)(m+3)}p^4 - \frac{15p^5}{(m+1)(m+3)} + \frac{15p^6}{(m+1)(m+3)(m+5)}.$$

The distribution (10·4) may, for the leading case $t = 2$, be written

$$(n-2)(1-\nu)^{\frac{n}{2}}p^{n-3}\left\{1 + \left(\frac{n}{2}\right)^2 v_1\nu + \left(\frac{n}{2}\right)^2\left(\frac{n+1}{2}\right)^2\frac{v_2\nu^2}{(2\,!)^2} + \ldots\right\}dp \ldots (10\cdot9).$$

The series is uniformly convergent and may be integrated term by term, thus providing a test of significance for the tetrad difference. It is not however very convenient for computation unless n and ν are small. For large values of n, the method of the preceding section may be used: the standard error often gives a satisfactory test of significance, even when used with the crude inequality of Tchebycheff, which takes no account of the nature of the particular distribution.

Light is thrown on the form of the frequency curves by the expansions we have just obtained. The case $t = 2$ stands out as of a special character, different from the rest; this will be true in general where $s = t$. This special character is related to the fact that positive and negative sample values of q are distinguishable only if $s = t$. In other cases, just as in that of the multiple correlation (i.e. that of q when $s = 1$), the values must be taken as positive, and q^2 is in some respects a more natural variate to use.

The v_k, and therefore the convergent series in (10·4) and (10·9), and also the derivatives of the v_k and of the series, take definite finite values both for $q = 0$ and for $q = 1$. From (10·4) it is therefore evident that the frequency curve for q has, for $q = 1$, contact with the axis of order $m - 2$. For $q = 0$ the ordinate of the curve is zero for $t \geqslant 3$, but has a finite value for $t = 2$.

The derivative with respect to q of the integral in (8·3) has, if $\nu < 1$, a finite negative value for $q = 0$ as well as for every positive value of q. The ordinate of the distribution curve for $t = 2$ will therefore have these properties. This curve must be symmetrical about $q = 0$. Hence it is not flat-topped, but has a corner above the origin.

But if $\nu = 1$ the distribution of q for $s = t = 2$ does not have such a discontinuity in the middle. For in this case linear functions of the variates in the two sets exist which are perfectly correlated with each other, and are thus for our purposes identical. Taking these as x_1 and x_3, (1·2) shows that q is in every sample the partial correlation of the remaining two variates. Hence when $\nu = 1$ the distribution becomes identically that of the partial correlation coefficient. According to R. A. Fisher's work*, this is the same as the distribution of the simple correlation coefficient, with the sample number reduced by unity, a distribution having continuous derivatives of all orders throughout its range.

11. *Tests for Complete Independence.* If $s = 2$ and *both* canonical correlations vanish, the normal distribution of the population implies complete independence between the two sets. No linear function of the first set is correlated with any of the second. In this case $\nu = 0$, and the distribution of q reduces, as is at once evident from the form (10·4), together with (10·2) and (10·7), to the extremely simple form,

$$\frac{(n-2)!}{(t-2)!\,(n-t-1)!} q^{t-2} (1-q)^{n-t-1}\, dq \quad\text{.................(11·1)},$$

for positive values of q. Thus q has in this case the same distribution as the square of the multiple correlation coefficient in samples of n $(= N - 1)$ from an uncorrelated normal population, with $t - 1$ variates.

The question whether complete independence exists between two sets of variates for which we have sample correlations may be investigated by computing q and determining from (11·1) whether the probability of so great a value of q is negligible. This requires the integral of (11·1), which is easy to compute for any moderate value of t. For large values of t it may be obtained from the *Tables of the Incomplete Beta Function*†. For $t = 2$ the probability of a greater value of $|q|$ if complete independence really exists is simply

$$P = (1 - |q|)^{n-2} = (1 - |q|)^{N-3} \quad\text{....................(11·2)},$$

where N is the number in the sample. In this way a very simple test for complete independence may be applied.

But this is not by any means the only possible test of complete independence between two sets. Indeed, the distribution of the vector alienation coefficient (Section 4),

$$z = \frac{D}{AB},$$

has been found by Wilks under this same hypothesis of complete independence and normality‡. This distribution, which was obtained by means of its moments, reduces for the case $s = 2$ which we are now studying to

$$\frac{(n-2)!}{2\,(t-1)!\,(n-t-2)!} z^{\frac{n-t-3}{2}} (1 - \sqrt{z})^{t-1}\, dz \quad\text{...........(11·3)}.$$

* "The Distribution of the Partial Correlation Coefficient" in *Metron*, Vol. III. (1924), pp. 329—332.

† *Biometrika* Office, 1934.

‡ Wilks, *op. cit.*

The range of possible values of z is from 0 to 1, the latter corresponding to complete independence between the two sets of variates, just as does $q = 0$. q and z are not functionally related; for a continuum of values of either can be found which is consistent with any value of the other. The field of the joint distribution of the two is easily delimited by reference to the canonical correlations r_1 and r_2. Indeed, if we always take $q \geqslant 0$, we have from (4·10) that the field of variation of a point of coordinates q, z is in the quadrant in which both are positive, and is bounded by the parabola

$$z = (1 - q)^2 \quad \text{.................................(11·4)},$$

shown in Figure 1. The best agreement with the hypothesis of complete independence is shown by a sample for which $z = 1$ and $q = 0$, and which therefore corresponds to the point in the upper corner of the figure.

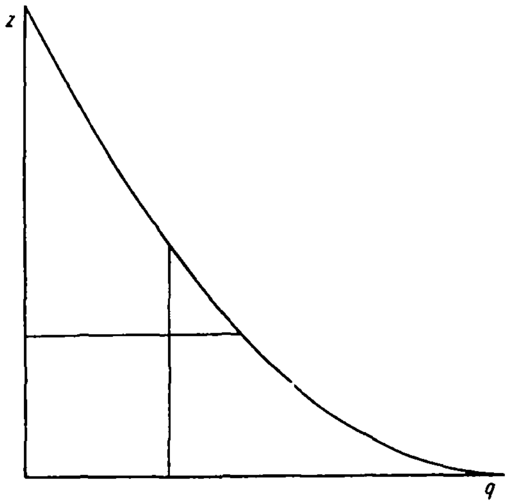

Fig. 1.

If we represent a sample by a point in a plane in which r_1 and r_2 are rectangular coordinates, and take r_1 as the greater, then the field of variation is the right triangle for which $0 \leqslant r_2 \leqslant r_1 \leqslant 1$. The point corresponding to best agreement with the hypothesis of complete independence is in this case the origin. The curves $q =$ constant and $z =$ constant, shown in Figure 2, are respectively hyperbolic arcs, and quartic curves which in the neighbourhood of the origin approximate circles. Their equations are

$$r_1 r_2 = q, \quad (1 - r_1^2)(1 - r_2^2) = z \quad \text{.....................(11·5)}.$$

To test complete independence by means of z, we need the integral of (11·3) from zero to the observed value. For $t = 2$ this is

$$P' = z^{\frac{n-3}{2}} \{(n-2)(1 - \sqrt{z}) + \sqrt{z}\} \text{.....................(11·6)}.$$

Like the integral of the distribution of q, that of (11·3) is easily found numerically from the Tables of the Incomplete Beta Function.

The existence of two different, though exact, tests of the same hypothesis makes us ask in what circumstances each should be used. No general answer appears to be possible to this question; but if we make sufficiently special assumptions about the nature of the deviations from our hypothesis that are likely to occur, or test for deviations of a sufficiently special character, a unique solution will exist.

In order that z differ from unity, it is seen from (11·5) to be sufficient for *either* of the canonical correlations to differ from zero. But in order that q differ from zero, it is necessary that *both* correlations differ from zero. This suggests that the z test will be the more sensitive to deviations from complete independence resulting from the existence of only a single component common to the two sets of variates; but if the correlations of one set with the other result from two independent common components operating to an approximately equal extent, the deviation from in-

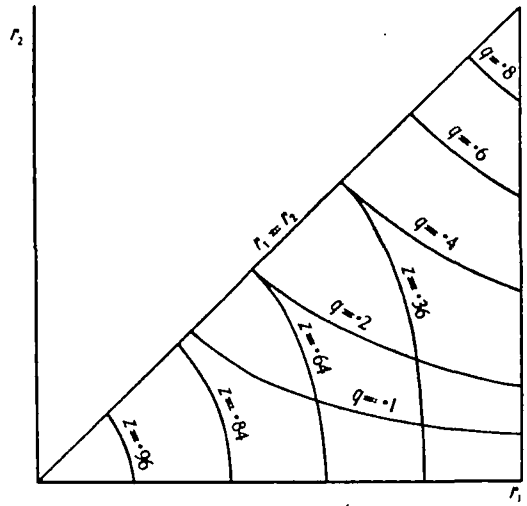

Fig. 2.

dependence will be revealed by q more clearly than by z. This conclusion is confirmed by a comparison of (11·2) with (11·6), putting for q and z their values from (11·5). If $r_2 = r_1$, then

$$P = (1 - r_1^2)^{n-2}, \quad P' = (1 - r_1^2)^{n-3}\{(n-2)\, r_1^2 + 1 - r_1^2\},$$

so that $P < P'$, and q provides the more sensitive test. If on the other hand $r_2 = 0$, $P = 1$, so that q provides no evidence whatever of deviation from independence, though for a large enough sample P' becomes arbitrarily small, supplying evidence to any desired extent, if r_1 has any constant value other than zero.

Let us apply both tests for complete independence to Kelley's correlations cited in Section 6. For the correlations of arithmetical with reading abilities the values (6·1) of q and z were obtained, with $s = t = 2$. From (11·2) the test for complete independence based on q gives $P = \cdot 023$, a probability so small that we may conclude that the two kinds of ability really have something in common.

The same conclusion is given even greater definiteness by the z test (11·6), from which we have $P' = \cdot 0001$.

The comparison of arithmetical with memory tests in Section 6 was for the values $s = 2$, $t = 3$. In this case we find from q that $P = 8{\cdot}6 \times 10^{-10}$, while the test for complete independence by means of z gives $P' = 1{\cdot}0 \times 10^{-24}$. Thus z gives a more sensitive test, and a more conclusive demonstration, of complete independence in both these cases than does q. The underlying reason for this is the considerable inequality between the two canonical correlations in each case.

One practical consideration in favour of the q test is that q is somewhat easier to calculate than z. The chief ground for distinction between them is however their sensitiveness to different types of deviations from complete independence.

Wilks in a later paper* derived the z test for complete independence from the likelihood criterion of Egon S. Pearson and J. Neyman. The considerations of this section therefore are relevant to an understanding of this criterion.

It is clear that a full understanding of the relations between two independent pairs of variates necessitates a knowledge of the bivariate distribution surfaces of r_1 and r_2, and of q and z. These we shall proceed to investigate.

12. *Alternants of a Plane and of a Sample.* The common method of specifying the orientation of a plane by the direction cosines of its normal is unsatisfactory in a space of more than three dimensions, since then a plane has an infinity of normals at each point. Instead we shall use determinants which may be regarded as of the form known as alternants. If in a space of n dimensions a flat space of k dimensions is determined by the origin and k other points, we shall call the k-rowed determinants in the matrix of the rectangular coordinates of these k points the *alternants* of the k-space. The alternants are C_k^n in number, and are connected by numerous quadratic relations. The number of degrees of freedom of the k-space through a fixed point is $k(n-k)$. The alternants depend only on the k-space, and not on the particular points used to determine it, except for multiplication of all the alternants by a constant; for to replace the k points by k others in the same k-space is to replace their matrix of coordinates by a new one whose rows are linear functions of the old; and this merely multiplies all the k-rowed determinants by a constant.

Taking the case $k = 2$, a plane containing the points

$$x_1, x_2, \ldots, x_n,$$

$$y_1, y_2, \ldots, y_n$$

is specified by the alternants

$$p_{ij} = x_i y_j - x_j y_i \qquad \text{.............................(12·1)},$$

which are analogous to the Plücker coordinates of a line. Indeed, the planes through a point in n-space are in one-to-one correspondence with the lines in which they meet an $(n-1)$-space not containing the point.

* "On the Independence of k Sets of Normally Distributed Statistical Variables" in *Econometrica*, Vol. III. (1935), pp. 309—326.

The relations connecting these alternants, apart from the obvious relations

$$p_{ij} = -p_{ji} \quad \text{..............................(12·2)},$$

are obtainable from the fact that, on account of identical rows,

$$\begin{vmatrix} x_i & x_j & x_k & x_m \\ y_i & y_j & y_k & y_m \\ x_i & x_j & x_k & x_m \\ y_i & y_j & y_k & y_m \end{vmatrix} = 0.$$

Applying a Laplace expansion to the first two rows of this determinant we obtain

$$p_{ij}p_{km} - p_{ik}p_{jm} + p_{im}p_{jk} = 0 \quad \text{.......................(12·3)}.$$

There is one of these relations for each combination of subscripts, but not all are independent. To obtain a set of independent relations which shall imply all the rest, we first observe that not all the p_{ij} can be zero if a definite plane is to be specified, for in that case the x- and y-points would be collinear with the origin. Let the notation be arranged so that $p_{12} \neq 0$. Then in terms of

$$\left.\begin{matrix} p_{12}, p_{13}, p_{14}, \ldots, p_{1n} \\ p_{23}, p_{24}, \ldots, p_{2n} \end{matrix}\right\} \quad \text{..........................(12·4)}$$

any other alternant p_{ij} is determined from (12·3), by putting $k = 1$, $m = 2$, so that

$$p_{ij} = \frac{p_{1i}p_{2j} - p_{1j}p_{2i}}{p_{12}} \quad \text{..........................(12·5)}.$$

The relations (12·5) constitute a complete set of independent relations of the form (12·3). For if in the left member of (12·3) we substitute for all the alternants the expressions obtained from (12·5) by putting the several combinations of subscripts in place of i and j, the resulting equation is satisfied identically. Furthermore, any set of quantities p_{ij} $(i, j = 1, \ldots, n)$ satisfying (12·2) and (12·5), and not all zero, determines uniquely a plane through the origin. For, supposing that $p_{12} \neq 0$, we may obtain the coordinates of two points not collinear with the origin in the following manner. Let $x_1 = y_2 = 0$, and let $x_2 = 1$. Putting $i = 1$, $j = 2$ in (12·1) we then have $y_1 = -p_{12}$. Then putting first $j = 1$ and then $j = 2$, we obtain $x_i = \frac{p_{i1}}{y_1} = \frac{-p_{i1}}{p_{12}}$; and $y_i = p_{2i}$. The points whose coordinates are thus determined cannot be collinear with the origin, for if they were the alternants would all be zero.

Since the alternants of a plane have so far been determined only to within a multiplicative constant, we may determine them uniquely if we add the condition

$$\Sigma p_{ij}^2 = 1 \quad \text{.................................(12·6)}.$$

Here we use the sign Σ to mean summation over the $\frac{n(n-1)}{2}$ alternants

$$p_{12}, p_{13}, \ldots, p_{n-1,n}$$

for which the first subscript is less than the second. This condition on the quantities (12·4) shows that the number of independent alternants is $2n - 4$, which is the number of degrees of freedom of the plane.

We define the alternants of a set of observations of two variates on n individuals (or rather with n degrees of freedom after elimination of the mean and possibly other variables and an appropriate orthogonal transformation of the observations) as the determinants p_{ij} in the matrix of observations, or of pseudo-observations, multiplied by a constant. It will frequently be convenient to choose this constant so that (12·6) is satisfied. This definition breaks down in the case where the determinants are all zero, but if the observations are a sample from a continuous distribution this is infinitely improbable, and we shall disregard this case.

It is clear that all relations between two pairs of variates that are invariant under internal linear transformations of the pairs, and are based on a sample, must be expressible in terms of the alternants of the two pairs; for such relations must correspond to relations between planes through the origin in n-space, independent of the particular points used to define the planes. The relations depending on correlations must also be invariant under rotations of the n-space about the origin, since the correlations are cosines of angles at the origin, which are invariant under rotations. We shall therefore suppose for simplicity that the axes have been rotated in such a way that the first two of them lie in the plane of the observations on one pair of variates, and contain the observation points for this pair. We shall suppose further that all the observation points are at unit distance from the origin, so that the sum of the squares of the observations on each variate is unity. None of these assumptions reduces the generality of our results. The matrix of observations now takes the form

$$\begin{Vmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & \dots & x_n \\ y_1 & y_2 & y_3 & y_4 & y_5 & \dots & y_n \\ 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \end{Vmatrix} \quad \text{.....................(12·7).}$$

The determinant D of the correlations among the four variates is the determinant of sums of squares and products, since each sum of squares is unity. Hence, by Section 2, D is the sum of the squares of the four-rowed determinants in the matrix (12·7). But all these determinants are zero except those containing the first and second columns. The determinant consisting of the first, second, ith and jth columns equals $x_i y_j - x_j y_i$. Defining this as p_{ij}, we obtain the result

$$D = \Sigma' p_{ij}^2 \quad \text{.................................(12·8),}$$

where Σ' denotes summation from 3 to n with respect to i, and from $i+1$ to n with respect to j. It is further evident from (12·7) that the determinants of correlations within the sets are

$$A = \begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix} = \Sigma p_{ij}^2, \; B = 1 \quad \text{.....................(12·9).}$$

The equation (3·6) for the canonical correlations of the sample is

$$\begin{vmatrix} \lambda & \lambda r_{12} & r_{13} & r_{14} \\ \lambda r_{12} & \lambda & r_{23} & r_{24} \\ r_{13} & r_{23} & \lambda & \lambda r_{34} \\ r_{14} & r_{24} & \lambda r_{34} & \lambda \end{vmatrix} = 0 \quad \text{..................(12·10).}$$

The coefficient of λ^4 in this equation is $AB = \Sigma p_{ij}^2$. The term independent of λ is the square of the sum of the products of the determinants in the first two rows of (12·7) by the corresponding determinants in the last two rows. The latter determinants, however, are all zero but the first; hence the constant term in the equation is p_{12}^2. The coefficient of λ^2 may be obtained by putting $\lambda = 1$ in the left member of (12·10) and subtracting the coefficient of λ^4 and the constant term. This coefficient is therefore equal to

$$D - AB - p_{12}^2 = -\Sigma p_{1i}^2 - \Sigma p_{2i},$$

where Σ denotes summation with respect to i from 1 to n. We recall in this connection that (12·2) shows that $p_{ii} = 0$. The equation may thus be written in terms of alternants

$$\lambda^4 \Sigma p_{ij}^2 - \lambda^2 \Sigma (p_{1i}^2 + p_{2i}^2) + p_{12}^2 = 0 \quad \text{..................(12·11)}.$$

If we regard this as a quadratic equation in λ^2, the roots are r_1^2 and r_2^2. Hence, from (11·5) and the expressions for the coefficients in terms of the roots,

$$q^2 = \frac{p_{12}^2}{\Sigma p_{ij}^2}, \quad z = 1 - \frac{\Sigma (p_{1i}^2 + p_{2i}^2) - p_{12}^2}{\Sigma p_{ij}^2} = \frac{\Sigma' p_{ij}^2}{\Sigma p_{ij}^2} \quad \text{.........(12·12)}.$$

If we adopt the further convention (12·6), which by (12·9) is seen to be equivalent to assuming that the first pair of variates has been reduced by an internal transformation so that the correlation is zero, (12·11) and (12·12) simplify to

$$\lambda^4 - \lambda^2 \Sigma (p_{1i}^2 + p_{2i}^2) + p_{12}^2 = 0 \quad \text{.....................(12·13)},$$

$$q = \pm p_{12}, \quad z = \Sigma' p_{ij}^2 \quad \text{...........................(12·14)}.$$

If we do not specialise one of our planes with reference to the coordinates, but take its alternants as q_{ij}, while those of the other are p_{ij}, it is easy to see in the foregoing manner, or from Section 4, that the vector correlation is

$$q = \frac{\Sigma p_{ij} q_{ij}}{\sqrt{(\Sigma p_{ij}^2)(\Sigma q_{ij}^2)}} \quad \text{...........................(12·15)}.$$

This has the form of the ordinary formula for a correlation coefficient, or of the cosine of an angle. From the latter fact comes the following conclusion, which is of the utmost importance for our purposes.

Let us take the alternants p_{ij} for which $i < j$ as Cartesian coordinates in a space of $\frac{n(n-1)}{2}$ dimensions, and denote by V the subspace in which the equations (12·5) and (12·6) hold. Then V is a curved space of $2n - 4$ dimensions, in which all the equations (12·3) hold, since they follow from (12·5). The points of V are in one-to-one correspondence with the planes through the origin in n-space. A property of this correspondence which we shall use is that it is *metrical*, in the sense that any rotation of the n-space about the origin engenders a transformation of V which is also a rotation. This fact follows from (12·15), which shows that q is the cosine of the angle at the origin between lines extending to points of V representing two planes. Under a rigid rotation of the n-space, the correlations defining q are all invariant, so that q is invariant. Hence the points of V representing the rotated pair of planes are exactly as far apart as the points representing the planes in their original position, since all these points are, by (12·6), equidistant from the origin. Thus the transformation of V satisfies the definition of a rotation.

13. *The Bivariate Distribution for Complete Independence* ($s = t = 2$, $n = 4$). If there is complete independence of one pair of variates from another, $\rho_1 = \rho_2 = 0$. We may without loss of generality regard the internal correlations also as zero. The planes corresponding to a sample from a normal distribution are then determined by lines drawn through the origin in n-space at random, in the sense that the probability of a line meeting any region on a surrounding sphere is proportional to the generalized area of the region. The chance selection of a plane in this way is equivalent to the selection of a point in V in such a way that the element of probability is proportional to the volume element. For, since any plane through the origin in n-space can be rotated into any other, any point in V can be rotated into any other, and will carry with it in this rotation its probability density, which must therefore be uniform over the whole of V. Thus all problems of finding distributions of statistics calculated from the pairs of variates in such a way as to be invariant under internal transformations reduce to purely geometrical problems of finding the $(2n-4)$-dimensional volumes of the corresponding regions in V.

The distribution of q and z, or of r_1 and r_2, will be deduced with the help of methods of parametric representation resembling those previously applied by the author to other statistical problems*. First we take the case $n = 4$. In the six-dimensional space in which the alternants are Cartesian coordinates, V is then a curved four-dimensional space having the equations

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0, \quad p_{12}^2 + p_{13}^2 + p_{14}^2 + p_{23}^2 + p_{24}^2 + p_{34}^2 = 1 \qquad \text{......(13·1)}.$$

It follows that V may be defined in terms of four parameters α, β, γ, δ by means of the equations

$$\left.\begin{aligned} p_{14} &= \tfrac{1}{2}(\sin\alpha\sin\beta + \sin\gamma\sin\delta) \\ p_{23} &= \tfrac{1}{2}(\sin\alpha\sin\beta - \sin\gamma\sin\delta) \\ p_{13} &= \tfrac{1}{2}(\cos\alpha\sin\beta + \cos\gamma\sin\delta) \\ p_{24} &= -\tfrac{1}{2}(\cos\alpha\sin\beta - \cos\gamma\sin\delta) \\ p_{12} &= \tfrac{1}{2}(\cos\beta + \cos\delta) \\ p_{34} &= \tfrac{1}{2}(\cos\beta - \cos\delta) \end{aligned}\right\} \qquad \text{..................(13·2)},$$

since these equations satisfy both the equations (13·1). All points of V are included when we allow α and γ to vary from 0 to 2π, and β and δ from 0 to π. The element of volume in V is of course

$$\sqrt{g}\, d\alpha\, d\beta\, d\gamma\, d\delta,$$

* "The Distribution of Correlation Ratios Calculated from Random Data" in *Proceedings of the National Academy of Sciences*, Vol. XI. (1925), pp. 657—662; "The Generalization of Student's Ratio" in *Annals of Mathematical Statistics*, Vol. II. (1931), pp. 360—378; "The Physical State of Protoplasm," *loc. cit.*

where g is the sum of the squares of the four-rowed determinants in the matrix of partial derivatives of the p_{ij} with respect to α, β, γ and δ. These derivatives are the halves of the elements of

$$\left\|\begin{array}{cccccc} \cos\alpha\sin\beta & \cos\alpha\sin\beta & -\sin\alpha\sin\beta & \sin\alpha\sin\beta & 0 & 0 \\ \sin\alpha\cos\beta & \sin\alpha\cos\beta & \cos\alpha\cos\beta & -\cos\alpha\cos\beta & -\sin\beta & -\sin\beta \\ \cos\gamma\sin\delta & -\cos\gamma\sin\delta & -\sin\gamma\sin\delta & -\sin\gamma\sin\delta & 0 & 0 \\ \sin\gamma\cos\delta & -\sin\gamma\cos\delta & \cos\gamma\cos\delta & \cos\gamma\cos\delta & -\sin\delta & +\sin\delta \end{array}\right\|.$$

The sum of products of corresponding elements in each pair of rows of this matrix is zero; the sums of squares are respectively $2\sin^2\beta$, 2, $2\sin^2\delta$, and 2, each of which sums must be divided by 4. Thus in accordance with Section 2 we have

$$g = \tfrac{1}{16}\sin^2\beta\sin^2\delta,$$

so that the volume element in V is

$$\tfrac{1}{4}\sin\beta\sin\delta\, d\alpha\, d\beta\, d\gamma\, d\delta, \quad \text{or} \quad \tfrac{1}{4}d(\cos\beta)\, d(\cos\delta)\, d\alpha\, d\gamma.$$

From this it follows, if we put

$$\xi = \cos\beta, \quad \xi' = \cos\delta \quad \text{..........................(13·3)},$$

that ξ and ξ' are independently distributed with uniform density between -1 and 1. If we take q to have the sign of p_{12}, and $\sqrt{z}$ to have that of p_{34}, we have, from (12·14), (13·2) and (13·3),

$$q = p_{12} = \tfrac{1}{2}(\xi + \xi'), \quad \sqrt{z} = p_{34} = \tfrac{1}{2}(\xi - \xi'),$$

whence $\xi = q + \sqrt{z}, \qquad \xi' = q - \sqrt{z}.$

These relations show, because they are linear, and since points of coordinates (ξ, ξ') are uniformly distributed in the square $\xi = \pm 1$, $\xi' = \pm 1$, that points of coordinates $(q, \sqrt{z})$ are uniformly distributed in the square bounded by the four lines

$$q \pm \sqrt{z} = \pm 1.$$

If we restrict q and $\sqrt{z}$ to positive values, their distribution will be

$$2\, dq\, d\sqrt{z},$$

within the triangle bounded by the coordinate axes and the line $q + \sqrt{z} = 1$. The distribution of q and z is therefore

$$z^{-\frac{1}{2}}dz\, dq \quad \text{..................................(13·4)},$$

subject to the limitations that both are positive, and that

$$z \leqslant (1-q)^2 \quad \text{................................(13·5)}.$$

If we integrate (13·4) with respect to q or z we obtain Wilks' distribution (11·3) of z, or the distribution (11·1) of q, respectively, for the case $s = t = 2$, $n = 4$.

We may regard (α, β) and (γ, δ) as the spherical coordinates of two points on a sphere in 3-space. The equations (13·2) thus establish a correspondence having metrical properties between planes through a point in 4-space and pairs of points on an ordinary sphere. In this representation q appears as the mean distance of the two points from a fixed plane, while $\sqrt{z}$ is half the difference of the distances from this plane. From the theorem of Archimedes that the area of a zone depends only on the distance between the bounding planes and the radius of the sphere (which in this case is unity), it is therefore evident that when z is fixed the distribution of q is of uniform density, confirming (13·4).

Let us call r the cosine of the angle between the two lines determining our variable plane in 4-space. The distribution of r, which is the sample correlation between two really uncorrelated variates, is readily seen geometrically, or by putting $n = 4$ in the general distribution of such sample correlations, to be

$$\frac{2}{\pi}(1-r^2)^{\frac{1}{2}}\,dr.$$

Moreover, this distribution is independent of that of q and z, since r depends only on the angle within the plane, and q and z on the plane itself. Consequently the joint distribution of the three is, for $n = 4$,

$$\frac{2}{\pi}(1-r^2)^{\frac{1}{2}} z^{-\frac{1}{2}}\,dr\,dz\,dq \qquad (13·6).$$

This result and the following theorem will be used in Section 15 in extending the distribution to a general value of n.

14. *Theorem on Circularly Distributed Variates.* The sum or difference of two variates distributed independently and with uniform density over a particular range is known to have a distribution represented by an isosceles triangle whose base has double the breadth of the original range. If however each value of the sum or difference is reduced with the original range as modulus—that is, is replaced by the remainder after dividing by the range—the resulting distribution is exactly the original one, with uniform density over the same range. This is a special case of the following rather remarkable

THEOREM: *If any number of variates are distributed independently and with uniform density from* 0 *to* a, *then any linear function of these variates with integral coefficients, when reduced modulo* a, *is likewise distributed with uniform density from* 0 *to* a. *Any number of such functions, if algebraically independent, are also independent in the probability sense.*

The truth of this theorem becomes evident when we regard each set of values of the variates as a point in a space having the metrical properties of a hypercube of as many dimensions as there are variates, but with a topological nature determined by making each pair of diametrically opposite faces of the hypercube correspond to a single region of the space. The space is thus a closed manifold generalizing a torus in its topology, but not contained in a euclidean space, because of its metrical nature. For two variates this representing space would be approximated by a torus obtained by revolving a very small circle about a very distant line in its plane. Another representation in this case would be by means of the squares of side a into which a plane is divided by two sets of parallel lines, all points occupying a particular position within their respective squares being regarded as identical. If we call the variates, or coordinates, $x_1, x_2, \ldots, x_m$, the linear functions

$$y_i = \sum_{j=1}^{m} a_{ij} x_j \; (i = 1, 2, \ldots, k) \quad \ldots\ldots\ldots\ldots\ldots\ldots(14\cdot1),$$

in which the coefficients a_{ij} are positive or negative integers or zero (but are not all zero for any value of i), are constants over loci which, on the representation on a plane or flat space of m dimensions, are parallel lines, or hyperplanes of $m-1$ dimensions. In the space itself, in which only one point corresponds to each set of values of the variates other than 0 and a, the loci (14·1) are closed curves, or closed hypersurfaces, because the coefficients are integers. It is obvious that the volume in this space contained between $y_i = b$ and $y_i = c$ must, on account of the homogeneity, be proportional to $b - c$.

If the k linear functions (14·1) are linearly independent, the loci obtained by giving each y_i a succession of constant values differing consecutively by $\frac{a}{p}$, where p is an integer, will divide the space into congruent parallelepipedons. If $k-1$ of the y_i are constrained to lie in certain of these intervals, the representing point is merely constrained to lie in a certain layer. Since all such layers must be congruent, the distribution of the kth of the y_i, reduced modulo a, is not affected by this constraint. Hence all the variates thus reduced are independent.

15. *Generalization of Section* 13 *for Samples of Any Size.* No direct extension to a larger number of dimensions of the method of using alternants in Section 13 appears to lead in any simple fashion to the generalization of the distribution there found. This generalization will however be obtained with the help of hyperspherical coordinates in the space of the observations.

On account of the spherical symmetry of the density distributions in n-space in the absence of true correlation, our distributions will not be affected if we assume the two points of coordinates $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ to be taken independently at random on a unit sphere about the origin in n dimensions, in such a way that the element of probability for each is proportional to the element of $(n-1)$-dimensional area on this hypersphere. If we define the hypersphere parametrically by the equations

$$\left.\begin{array}{ll}
x_1 &= \sin\theta_1 \sin\theta_2 \\
x_2 &= \cos\theta_1 \sin\theta_2 \\
x_3 &= \qquad \cos\theta_2 \cos\theta_3 \\
x_4 &= \qquad \cos\theta_2 \sin\theta_3 \cos\theta_4 \\
\multicolumn{2}{l}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
x_{n-1} &= \qquad \cos\theta_2 \sin\theta_3 \ldots \sin\theta_{n-2} \cos\theta_{n-1} \\
x_n &= \qquad \cos\theta_2 \sin\theta_3 \ldots \sin\theta_{n-2} \sin\theta_{n-1} \\
\multicolumn{2}{l}{(0 \leqslant \theta_1, \theta_{n-1} \leqslant 2\pi;\ 0 \leqslant \theta_2 \leqslant \pi/2;\ 0 \leqslant \theta_3, \theta_4, \ldots, \theta_{n-2} \leqslant \pi)}
\end{array}\right\} \ldots\ldots(15\cdot1),$$

which satisfy $\Sigma x^2 = 1$ identically, then the element of $(n-1)$-dimensional area for the x-point may be written

$$\sqrt{g}\, d\theta_1 \theta_2 \ldots d\theta_{n-1},$$

where g is a determinant of $n-1$ rows, in which the element in the ith row and jth column is

$$\sum_{\alpha=1}^{n} \frac{\partial x_\alpha}{\partial \theta_i} \frac{\partial x_\alpha}{\partial \theta_j} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15{\cdot}2).$$

All these quantities are readily seen from (15·1) to vanish except those for which $i = j$. The successive diagonal elements of g are

$$\sin^2 \theta_2, \quad 1, \quad \cos^2 \theta_2, \quad \cos^2 \theta_2 \sin^2 \theta_3, \ \ldots.$$

Hence the element of generalized area may be written

$$\sin \theta_2 \cos^{n-3} \theta_2 \sin^{n-4} \theta_3 \sin^{n-5} \theta_4 \ldots \sin \theta_{n-2}\, d\theta_1\, d\theta_2 \ldots d\theta_{n-1} \ \ldots\ldots (15{\cdot}3).$$

In the same way, if we put

$$\left.\begin{aligned} y_1 &= \sin \phi_1 \sin \phi_2 \\ y_2 &= \cos \phi_1 \sin \phi_2 \\ y_3 &= \qquad \cos \phi_2 \cos \phi_3 \\ y_4 &= \qquad \cos \phi_2 \sin \phi_3 \cos \phi_4 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= \qquad \cos \phi_2 \sin \phi_3 \sin \phi_4 \ldots \sin \phi_{n-1} \end{aligned}\right\} \ \ldots\ldots\ldots\ldots\ldots (15{\cdot}4),$$

the element of probability for the y-point is proportional to

$$\sin \phi_2 \cos^{n-3} \phi_2 \sin^{n-4} \phi_3 \ldots \sin \phi_{n-2}\, d\phi_1\, d\phi_2 \ldots d\phi_{n-1} \ \ldots\ldots\ldots (15{\cdot}5).$$

The distribution of the parameters defining the two points is obtained by multiplying (15·3) by (15·5). It is evident that all the quantities $\theta_1, \ldots, \theta_{n-1}, \phi_1, \ldots, \phi_{n-1}$ are independent in the probability sense, since the distribution function is a product of functions each involving only one of these parameters.

We now introduce quantities $u_3, u_4, \ldots, u_n, v_3, \ldots, v_n$ defined by the equations

$$x_i = u_i \cos \theta_2, \quad y_i = v_i \cos \phi_2 \quad (i = 3, 4, \ldots, n) \ldots\ldots\ldots\ldots\ldots (15{\cdot}6).$$

The u_i, by (15·1), are functions only of $\theta_3, \theta_4, \ldots, \theta_{n-1}$ and the v_i, by (15·4), are functions of $\phi_3, \ldots, \phi_{n-1}$. The u_i and v_i may be regarded as Cartesian coordinates of two points on a sphere in space of $n-2$ dimensions, these points being taken independently of each other and of the values of θ_1, θ_2, ϕ_1 and ϕ_2, with the element of probability proportional to the element of $(n-3)$-dimensional area. If we denote the angle between these points by $\lambda\ (0 \leqslant \lambda \leqslant \pi)$, it is evident that the distribution of λ is proportional to

$$\sin^{n-4} \lambda\, d\lambda \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15{\cdot}7),$$

and is independent of θ_1, θ_2, ϕ_1 and ϕ_2.

Let r be the cosine of the angle subtended at the origin in the n-dimensional space by the x- and y-points. Since

$$u_3 v_3 + \ldots + u_n v_n = \cos \lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15{\cdot}8),$$

it is evident from (15·1), (15·4), (15·6) and (15·8) that

$$r = \Sigma xy = \cos(\theta_1 - \phi_1)\sin\theta_2\sin\phi_2 + \cos\lambda\cos\theta_2\cos\phi_2 \qquad \text{.........(15·9)}.$$

The sample values of the variates of the second pair may in the absence of correlation in the population be represented by an arbitrary pair of fixed lines; we shall represent them by the first two coordinate axes. We then have from (12·12) and (12·1), with the help of Section 2,

$$q^2 = \frac{(x_1y_2 - x_2y_1)^2}{\Sigma x^2 \Sigma y^2 - (\Sigma xy)^2}, \quad z = \frac{\Sigma' x^2 \Sigma' y^2 - (\Sigma' xy)^2}{\Sigma x^2 \Sigma y^2 - (\Sigma xy)^2},$$

where Σ denotes summation from 1 to n, and Σ' from 3 to n. These expressions become, upon substitution from (15·1), (15·4), (15·6), (15·8) and (15·9),

$$q^2 = \frac{\sin^2(\theta_1 - \phi_1)\sin^2\theta_2\sin^2\phi_2}{1 - r^2}, \quad z = \frac{\cos^2\theta_2\cos^2\phi_2\sin^2\lambda}{1 - r^2} \qquad \text{......(15·10)}.$$

To simplify the notation we shall replace θ_2 and ϕ_2 simply by θ and ϕ respectively. We shall also put

$$\omega = \theta_1 - \phi_1.$$

Since only the sines and cosines of ω will enter into our discussion, we may regard ω as reduced modulo 2π. Now θ_1 and ϕ_1 vary independently and, as is seen from (15·3) and (15·5), with uniform density from 0 to 2π. Hence their difference ω must, by the theorem of the last section, have a distribution of uniform density from 0 to 2π. Moreover, since θ_1 and ϕ_1 have been seen to be independent of θ, ϕ and λ, it follows that ω is likewise independent of them; indeed, ω, θ, ϕ and λ constitute a completely independent set. The distributions of θ and ϕ are determined by integrating (15·3) and (15·5) between constant limits with respect to all the variates appearing in them except θ_2 and ϕ_2 respectively. Combining with (15·7) the result of this integration and the uniformity of the distribution of ω, we have that the element of probability is of the form

$$k_n \sin\theta\cos^{n-3}\theta\sin\phi\cos^{n-3}\phi\sin^{n-4}\lambda\, d\theta\, d\phi\, d\lambda\, d\omega \qquad \text{.........(15·11)},$$

where k_n depends only on n. The limits for θ and ϕ are 0 and $\frac{\pi}{2}$; for λ they are 0 and π; for ω they are 0 and 2π.

In the new notation, (15·9) and (15·10) become

$$r = \cos\omega\sin\theta\sin\phi + \cos\lambda\cos\theta\cos\phi \qquad \text{...............(15·12)},$$

$$q^2 = \frac{\sin^2\omega\sin^2\theta\sin^2\phi}{1 - r^2}, \quad z = \frac{\sin^2\lambda\cos^2\theta\cos^2\phi}{1 - r^2} \qquad \text{...........(15·13)}.$$

We next consider a transformation to the variates q, z, r and ω. Without troubling to compute the Jacobian J of this transformation, we need observe only that it is independent of n, since the functional relations (15·12) and (15·13) do not involve n. Substituting in (15·11) from the second of the equations (15·13) we find that the distribution is of the form

$$k_n \psi z^{\frac{n-3}{2}} (1 - r^2)^{\frac{n-3}{2}}\, dq\, dz\, dr\, d\omega,$$

where ψ does not involve n. Upon integrating this with respect to ω between certain limits depending on q, z and r, but not on n, we have the distribution

$$k_n \Psi z^{\frac{n-3}{2}} (1 - r^2)^{\frac{n-3}{2}} dq\, dz\, dr,$$

which for $n = 4$ must reduce to (13·6). Comparing with (13·6) we have $k_4 \Psi = \dfrac{2}{\pi z}$. Inasmuch as the distribution of r is known to be

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi}} (1-r^2)^{\frac{n-3}{2}} dr,$$

and to be independent of that of q and z, we have for the distribution of the latter two

$$h_n z^{\frac{n-5}{2}} dz\, dq,$$

where h_n depends only on n. Since the integral over the entire range of variation defined by the inequalities

$$0 \leqslant q \leqslant 1, \quad 0 \leqslant z \leqslant 1, \quad z \leqslant (1-q)^2$$

must be unity, the constant h_n is readily found. The distribution is

$$\tfrac{1}{2}(n-2)(n-3)\, z^{\frac{n-5}{2}} dz\, dq \quad \text{.....................(15·14)}.$$

Its form shows that, in the plane of Fig. 1, the loci of uniform density are horizontal lines.

The distribution of the canonical correlations is determined by (15·14), together with (11·5), which latter gives

$$\frac{\partial(q, z)}{\partial(r_1, r_2)} = 2(r_1^2 - r_2^2).$$

Thus the distribution of the correlations in case of complete independence is

$$(n-2)(n-3)(r_1^2 - r_2^2)(1 - r_1^2)^{\frac{n-5}{2}} (1 - r_2^2)^{\frac{n-5}{2}} dr_1\, dr_2 \text{......(15·15)}.$$

16. *Further Problems.* The foregoing treatment of sampling distributions is obviously incomplete. It would be desirable to have exact distributions, both of sample canonical correlations and of various functions of them, for cases in which the canonical correlations in the population have arbitrary values. The coefficients obtained for the canonical variates have sampling distributions which remain to be determined. Furthermore, various possible comparisons among different samples remain to be investigated; for example, there is the problem of testing the significance of the difference between vector correlations obtained from different samples.

A generalization of the problem of relations between two sets of variates invariant under internal linear transformations is that of invariants under such transformations of three or more sets of variates. A beginning of this theory has been made by

Wilks in the work previously alluded to; the z we have used is only a special case of a statistic of his, which in general is defined with reference to any number of sets of variates as a fraction, whose numerator is the determinant of the correlations among all the variates, and whose denominator is the product of the determinants of correlations within sets. It is obvious also that the invariants we have discussed, taken between every two of the sets, are invariants of such a system. An additional set of invariants will be the roots of the equation in λ resembling (3·6), obtained from the determinant of all the correlations or covariances by multiplying those between variates in the same set by $-\lambda$. It is easy to prove with the help of the theory of λ-matrices that the roots and coefficients of this equation are actually invariants.

A generalization of our work in a different direction would consider invariants, not under all linear internal transformations, but under a restricted class of these transformations. For example, a study of the relations of the prices to the quantities of several commodities might well consider transformations of *commodities*, such for example as the mixing of different grades of wheat, or the combination of raw materials and labour into finished products. If from quantities $q_1, q_2, \ldots$ of the old commodities there are formed quantities $q_1', q_2', \ldots$ of the new, we may, at least approximately, write

$$q_i' = \sum_j c_{ij} q_j \quad \text{.................................(16·1).}$$

If all the costs and profits of the mixing or manufacturing operation are regarded as prices of constituents, the *value* of one set of commodities will equal that of the other, so that

$$\Sigma p_i' q_i' = \Sigma p_j q_j \quad \text{.............................(16·2),}$$

where the p_j are the prices of the original commodities and the p_i' are those of the products. If we regard (16·1) as a linear transformation of the quantities, there will be a corresponding linear transformation of the prices, whose coefficients may be determined in terms of the c_{ij} by substituting (16·1) and

$$p_i' = \Sigma d_{ik} p_k$$

in (16·2) and then equating coefficients of like terms. This process shows that

$$\Sigma c_{ij} d_{ik} = \delta_{jk}, \quad = 1 \text{ if } j = k, \quad = 0 \text{ if } j \neq k.$$

These equations fully determine the d_{ik} as functions of the c_{ij}. The relation is such that the transformation of prices is *contragredient* to that of quantities.

An important class of relations between prices and quantities of a group of commodities would be the class of relations invariant under mixings of the kind described above. The canonical correlations and their functions, which are the main subject of this paper, are such invariants. But on account of the restriction that linear transformations of one set of variates shall be contragredient to those of the other, there will be additional invariants for this case, which remain to be investigated.

Other important problems are connected with the case of equal canonical correlations, which had to be excluded in deriving the approximate standard errors in Section 5. If two or more canonical correlations in the population are equal, it appears that the distribution of the corresponding sample values does not approach the multivariate normal form. This case is of much practical importance, owing to the practice of devising tests designed to measure the same character with equal accuracy. The psychologists' use of "reliability coefficients," and of "correlations corrected for attenuation" has been recognized as unsatisfactory. One symptom of trouble is that the formula for correlations corrected for attenuation sometimes gives values greater than unity. A satisfactory treatment of this difficulty should be possible with the help of the distribution function, when found, of sample canonical correlations, or of the vector correlation, when in the population the roots of the determinantal equation are equal