

CONTRIBUTIONS TO THE THEORY OF MULTIVARIATE STATISTICAL ANALYSIS*

BY

WILLIAM G. MADOW†

The increasing application of statistical methods in the social sciences has resulted in the consideration of hypotheses more complicated than those for which the theory of univariate statistical processes and the elementary theory of multivariate statistical processes suffice. One cause of the complications may be found in the fact that in the social sciences the results of an experiment are frequently vectors of several components rather than of one component. As examples we may note that the prices and quantities of a given commodity in p localities are vectors of p components, and that mental traits are generally tested by batteries of tests and not by a single test. Although the replacement of these vectors by some function, say an average, of their components is adequate for some purposes, yet in certain situations, any such function may be shown to be unsatisfactory. The fact that vector variates must then be analyzed requires the construction of a statistical theory which will facilitate that analysis.

For the normal multivariate distribution, the only distribution for which results have been obtained, the beginnings of that theory may be found in the fairly intensive study, since Gauss, of least squares and multiple correlation, which culminated in the derivation by R. A. Fisher in 1928 of the distribution of the multiple correlation coefficient when the correlation parameter is not zero.‡ The general theory of vector variates has been developed since 1931. It has included the analysis of variance of a vector variate and the theory of relations between sets of variables. Hotelling derived the distribution of the generalization of Student's ratio§ in 1931. Distributions of statistics occurring in the generalized analysis of variance were derived by Wilks|| in 1932. In 1936 Hotelling stated and solved a general problem in the theory of relations between sets of variables.¶ We continue this analysis.

* Presented to the Society, March 26, 1937; received by the editors October 7, 1937. §2 was read before the Institute of Mathematical Statistics, December 27, 1936; §3 was read before the American Mathematical Society under the title, *The generalized analysis of variance*.

† This research was done under a grant-in-aid from the Carnegie Corporation of New York.

‡ See [13]. The number in the brackets refers to the bibliography at the end of the paper.

§ See [18].

|| See [34]. Other contributions to the theory are given in [2], [3], [24], [36], [37], [38].

¶ See [20].

In §1 there are given certain definitions and notation which are used throughout the paper.

In §2 we prove several theorems,* the purpose of which is the establishment of a routine method of obtaining the distributions needed in multivariate analysis.

In §3 the theorems of §2 are employed to obtain some of the basic distributions of the theory of vector variates. As illustrations of the uses of the methods, we examine the theory of generalized periodogram analysis,† derive the joint distribution of q^2 and z ,‡ and the joint distribution of the two canonical correlations which exist when one of the sets contains two variables and the other set contains at least two variables, under the assumption that not more than one canonical correlation parameter differs from zero. We also analyze the generalized covariance, the tetrad difference,|| and other statistics occurring in the theory of relation between sets of variables.

It will be noted that the variables are rarely expressed as deviations from a mean or regression function. This is permissible since Theorem 7 enables us to assume that those functions have been eliminated. Thus, if n' is the actual sample number, and if the mean value of the chance variable is a linear function of p parameters, then n , the sample number in terms of which our formulas are expressed, is equal to $n' - p$.

1. Definitions and notation. The following definitions and notation will be used throughout this paper unless explicit statement to the contrary is made.

If used as subscripts or superscripts, or as indices of summation or multiplication, the letters i, j, k will take on all integral values from 1 through p ; the letters g, h will take on all integral values from 1 through $i - 1$, the letter ν will take on all integral values from 1 through n , and the letter γ will take on all integral values from 1 through m . We shall denote by p_γ the sum $n_1 + \cdots + n_\gamma$, ($p_0 = 0$).

The notation $A \times B$, will be used for the combinatory product¶ of the two sets A and B . We shall write A^2 for $A \times A$. Thus, if we denote by R^1 the totality of all real numbers, then R^ν is the totality of all sets of ν real numbers. If V is a subset of $A \times B$ and b is an element of B , we shall denote by $V(b)$

* These are more fully described in the first few paragraphs of §2.

† For a discussion of the Schuster periodogram in one variable, see [14] and [17].

‡ See [20], p. 333. These statistics are measures of the degree of dependence of sets of chance variables.

|| M. A. Girshick has obtained some of these distributions under slightly less general conditions by other methods. They have not, as yet, been published.

¶ That is, $A \times B$ is the set consisting of all pairs (a, b) where a is an element of A and b is an element of B .

the totality of elements a of A , such that (a, b) is an element of V .

If x_1, x_2, \dots, x_n are chance variables* with the distribution function† $P(\xi_1, \xi_2, \dots, \xi_n)$, and if

$$(1.1) \quad P(\xi_1, \xi_2, \dots, \xi_n) = \int_{-\infty}^{\xi_n} \cdots \int_{-\infty}^{\xi_2} \int_{-\infty}^{\xi_1} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n,$$

where $f(x_1, \dots, x_n)$ is a real single-valued non-negative summable function of the n real variables x_1, \dots, x_n defined on R^n , then $f(x_1, \dots, x_n)$ is called the probability density of the chance variables x_1, x_2, \dots, x_n . If $n = p_m$ and

$$P(\xi_1, \dots, \xi_n) \equiv \prod_{\gamma} P(\xi_{p_{\gamma-1}+1}, \dots, \xi_{p_{\gamma}}),$$

then the sets of chance variables $x_{p_{\gamma-1}+1}, \dots, x_{p_{\gamma}}$ will be said to be independently distributed. The symbol $\mathcal{D}(x_1, \dots, x_n)$ will stand for "the probability density of the chance variables‡ x_1, \dots, x_n ."

It is easy to see that if a set of chance variables has a probability density, then any subset of them will have a probability density. Furthermore if x_1, x_2, \dots, x_n have a probability density and if $n = p_m$, then a necessary and sufficient condition that the sets of chance variables $x_{p_{\gamma-1}+1}, \dots, x_{p_{\gamma}}$ be independently distributed is

$$\mathcal{D}(x_1, \dots, x_n) \equiv \prod_{\gamma} \mathcal{D}(x_{p_{\gamma-1}+1}, \dots, x_{p_{\gamma}}).$$

If $\phi(x_1, \dots, x_n)$ is a measurable function§ of x_1, \dots, x_n on R^n , then the symbol $\mathcal{E}\{\phi(x_1, \dots, x_n)\}$ will stand for "the mathematical expectation of $\phi(x_1, \dots, x_n)$," and, by definition,

$$\mathcal{E}\{\phi(x_1, \dots, x_n)\} = \int \cdots \int_{R^n} \phi(x_1, \dots, x_n) dP(x_1, \dots, x_n),$$

* For definitions of chance variables, see [7], p. 8 and [21], p. 20. Unfortunately there are differences in the terminology used in papers in statistics and in probability. To avoid confusion, one must remember that variate, statistic, and chance variable are synonyms, as are probability density and distribution. Writers on statistical subjects sometimes use the term cumulated distribution for distribution function. It is particularly to be noted that the term distribution function in the literature of the theory of probability is the integral of the distribution referred to in papers on the theory of statistics.

† For all sets of real numbers $\xi_1, \xi_2, \dots, \xi_n$, the probability that $x_1 < \xi_1, x_2 < \xi_2, \dots, x_n < \xi_n$ is given by $P(\xi_1, \xi_2, \dots, \xi_n)$. For a discussion of probability density and distribution function, see [7] and [21].

‡ In using \mathcal{D} we, of course, assert that x_1, \dots, x_n are chance variables and have a probability density.

§ We will usually extend the domain of definition of functions defined on a subset of R^n to cover the set R^n by setting the functions equal to zero outside their original domain of definition.

where the integral on the right-hand side is a Lebesgue-Stieltjes integral.*

The positive definite quadratic forms q and q_v will be defined by the equations

$$q = \sum_{i,j} \sigma^{ij} x_i x_j$$

and

$$q_v = \sum_{i,j} \sigma^{ij} x_{iv} x_{jv},$$

respectively. The matrix of q and q_v will be denoted by (σ^{-1}) . The determinant of (σ^{-1}) will be denoted by σ^{-1} . The inverse of (σ^{-1}) will be written (σ) , and the elements of (σ) will be denoted by σ_{ij} , where σ_{ij} is the cofactor of σ^{ij} in σ^{-1} divided by σ^{-1} . The determinant of (σ) is, of course, σ .

The function $N(x; c)$ will be defined by the equations

$$\begin{aligned} N(x; c) &= (2\pi c)^{-1/2} \exp [-x^2/2c], & -\infty < x < \infty; 0 < c, \\ N(x; c) &= 0, & -\infty < x < \infty; c \leq 0. \end{aligned}$$

The functions $N((x); (\sigma))$ and $N((x_v); (\sigma))$ will be defined by the equations

$$\begin{aligned} N((x); (\sigma)) &= (2\pi\sigma)^{-1/2} \exp [-q/2], & -\infty < x_i < \infty, \\ N((x_v); (\sigma)) &= (2\pi\sigma)^{-1/2} \exp [-q_v/2], & -\infty < x_{iv} < \infty. \end{aligned}$$

Of course the elements σ_{ij} may assume any values as long as the requirement that q and q_v are positive definite is satisfied.

If $\mathcal{D}(\mathbf{x}) = N(\mathbf{x}; \sigma^2)$, then \mathbf{x} will be said to be normally distributed with variance σ^2 or to have a normal distribution with variance σ^2 . If $\mathcal{D}(\mathbf{x}_1, \dots, \mathbf{x}_p) = N((\mathbf{x}); (\sigma))$, then $\mathbf{x}_1, \dots, \mathbf{x}_p$ will be said to have a normal multivariate distribution.

The function $G(x; n, c)$ will be defined by the equations

$$\begin{aligned} G(0; 0, c) &= 1, & -\infty < c < \infty, \\ G(x; n, c) &= [\Gamma(n/2)]^{-1} (2c)^{-n/2} x^{n/2-1} \exp [-x/2c], \\ & & 0 < x < \infty; 0 < n < \infty; 0 < c < \infty, \end{aligned}$$

and $G(x; n, c) = 0$ for all other values of x, n, c . The function $\bar{G}(\lambda; n, c)$ will be defined by the equation,

$$\bar{G}(\lambda; n, c) = \int_0^\infty x^{\lambda/2} G(x; n, c) dx, \quad \lambda \geq -n + 1.$$

* If a measurable function $y = \phi(x_1, \dots, x_n)$ has been defined, we shall mean by y the function $\phi(\mathbf{x}_1, \dots, \mathbf{x}_n)$.

Then it follows that

$$\bar{G}(\lambda; n, c) = (2c)^{\lambda/2} \frac{\Gamma((n + \lambda)/2)}{\Gamma(n/2)}.$$

If $\mathcal{D}(\mathbf{x}) = G(\mathbf{x}; n, 1)$, then \mathbf{x} will be said to have a χ^2 distribution with n degrees of freedom. In that case the distribution of $\mathbf{x}/2$ will be said to be an incomplete Γ -function distribution of index n .

The function $D((\mathbf{x}); (n))$ will be defined by the equations

$$D((\mathbf{x}); (n)) = \frac{\Gamma(n/2)}{\prod_{i=1}^{p+1} \Gamma(n_i/2)} \left(\prod_{i=1}^p x_i^{n_i/2-1} \right) (1 - x_1 - \cdots - x_p)^{n_{p+1}/2-1},$$

$$0 < x_i; x_1 + \cdots + x_p < 1, 1 \leq n_i, n = n_1 + \cdots + n_{p+1},$$

and $D=0$ otherwise. The function $\bar{D}(\lambda; (n))$ is defined by the equation,

$$\bar{D}(\lambda; (n)) = \int_{0 < x_p} \cdots \int_{0 < x_1}^{x_1 + \cdots + x_p < 1} \left(\prod_{i=1}^p x_i^{\lambda_i/2} \right) D((\mathbf{x}); (n)) dx_1 \cdots dx_p.$$

Then letting $\lambda = \lambda_1 + \cdots + \lambda_p$, we find that

$$\bar{D}(\lambda; (n)) = \frac{\Gamma(n/2)}{\Gamma((n + \lambda)/2)} \cdot \prod_{i=1}^p \frac{\Gamma((n_i + \lambda_i)/2)}{\Gamma(n_i/2)}.$$

If $\mathcal{D}(\mathbf{x}_1, \cdots, \mathbf{x}_p) = D((\mathbf{x}); (n))$, then $\mathbf{x}_1, \cdots, \mathbf{x}_p$ will be said to have an incomplete Dirichlet distribution of indices n_1, \cdots, n_{p+1} . If $p=1$, then \mathbf{x}_1 and \mathbf{x}_2 will be said to have an incomplete β -function distribution of indices n_1 and n_2 .

2. General theorems. One of the most important theorems of univariate statistical analysis is that which states necessary and sufficient conditions that a set of quadratic forms in normally and independently distributed chance variables be themselves independently distributed with a joint probability density which is a product of χ^2 distributions.*

The fundamental part of this theorem may be stated as follows:

THEOREM 1. (Cochran's theorem.) *If x_1, x_2, \cdots, x_n are normally and independently distributed, with zero means and unit variances, if q_1, q_2, \cdots, q_m are quadratic forms in x_1, x_2, \cdots, x_n , with constant coefficients such that*

$$(2.1) \quad \sum_{\gamma} q_{\gamma} = \sum_{\gamma} x_{\gamma}^2,$$

and if the rank of q_{γ} is n_{γ} , then a necessary and sufficient condition that there exist n linear functions with constant coefficients,

* The theorem is stated by W. G. Cochran [6], p. 178, and essentially stated by R. A. Fisher [12], p. 97.

$$(2.2) \quad z_{\mu} = \sum_{\nu} c_{\mu\nu} x_{\nu}, \quad \mu = 1, \dots, n,$$

which are normally and independently distributed with zero means and unit variances, and are such that

$$(2.3) \quad \sum_{\nu=p_{\gamma-1}+1}^{p_{\gamma}} z_{\nu}^2 = q_{\gamma},$$

is*

$$(2.4) \quad p_m = n.$$

The necessity of the condition (2.4) is obvious. For a detailed proof of the sufficiency of the condition, the reader is referred to [6], p. 178. It may be remarked, however, that the proof depends on the fact that in reducing the q_{γ} 's to sums of squares, the n linear functions thus obtained must be linearly independent or contradict (2.1).

In this section we prove several theorems by means of which many problems in multivariate statistical analysis may be solved. Of these, Theorems 7, 8, and 9 are generalizations of the Fisher-Cochran theorem.

A problem of statistical analysis may be formulated in the following way: If the distribution of several functions of certain chance variables has been derived, then what is the distribution of the same functions of other chance variables?

As an example of this problem we may cite the extension of the distribution of the multiple correlation coefficient of variables having a normal multivariate distribution with zero multiple correlation parameter† to the distribution of the multiple correlation coefficient of variables having a normal multivariate distribution, the multiple correlation parameter of which need not vanish.‡

In Theorem 2 the problem is solved under fairly general conditions. In particular the problem is solved when the functions are sufficient statistics§ for both sets of chance variables. Various properties of sufficient statistics are then studied in Theorems 3 and 4 in order to determine to what extent

* The algebraic content of this theorem may be stated as follows: If the real quadratic forms q_1, q_2, \dots, q_m in x_1, x_2, \dots, x_n are such that $\sum_{\gamma} q_{\gamma} = \sum_{\nu} x_{\nu}^2$, and if the rank of q_{γ} is n_{γ} , then a necessary and sufficient condition that there exist an orthogonal transformation $z_{\mu} = \sum_{\nu} c_{\mu\nu} x_{\nu}$, ($\mu=1, \dots, n$), such that $\sum_{\nu=p_{\gamma-1}+1}^{p_{\gamma}} z_{\nu}^2 = q_{\gamma}$, is $p_m = n$.

† This distribution was derived by R. A. Fisher. See [10], p. 91 and [11], p. 811.

‡ This distribution was derived by Fisher four years after the appearance of the preceding distribution. See [13], p. 660.

§ See [28], §4, for a general definition of a system of sufficient statistics.

the possession of a set of sufficient statistics defines the underlying chance variables and to derive an expression for the probability density of any system of sufficient statistics.

In Theorem 5 we show that certain useful transformations may be treated as though they were the results of iterating linear transformations.

Frequently, in statistical investigations one is confronted with the necessity of finding distributions subject to the condition that certain variables are held constant or the condition that the effect of certain variables be previously eliminated. In such cases one must learn if the distributions obtained are functions of the values of the variables held constant and of the distribution of these variables. In Theorem 6 we shall provide a means of obtaining this information in situations which essentially depend on the normal distribution.

The following well known lemma will be used in the proofs of Theorems 2 and 4:

LEMMA 1. *Let the real single-valued Borel-measurable* functions*

$$(2.5) \quad y_i = g_i(x_1, \dots, x_n)$$

be defined on R^n , and let $\psi(y_1, \dots, y_p)$ be a single-valued measurable function of y_1, \dots, y_p defined on R^p . Upon substituting from (2.5), it is seen that $\psi(y_1, \dots, y_p)$ is a single-valued measurable function, $\phi(x_1, \dots, x_n)$, of x_1, \dots, x_n .

Then, if x_1, \dots, x_n are chance variables, it follows that y_1, \dots, y_p , where $y_i = g_i(x_1, \dots, x_n)$, are chance variables and

$$\mathcal{E}\{\psi(y_1, \dots, y_p)\} = \mathcal{E}\{\phi(x_1, \dots, x_n)\}.$$

A proof of this Lemma is given in [21], p. 35. A similar result is true when the functions (2.5) are measurable. It is noted that if

$$\psi(y_1, \dots, y_p) = \exp \left[i \sum_j t_j \mu_j(y_1, \dots, y_p) \right],$$

where μ_j is a real single-valued measurable function of y_1, \dots, y_p , then we have the following equation, which is often used in applying the Fourier transform theory in the derivation of probability densities,

$$\mathcal{E} \left\{ \exp \left[i \sum_j t_j \lambda_j(x_1, \dots, x_n) \right] \right\} = \mathcal{E} \left\{ \exp \left[i \sum_j t_j \mu_j(y_1, \dots, y_p) \right] \right\},$$

* For a definition and discussion of Borel-measurable functions, see, for example, [31], chaps. 1, 2, 3.

where $\lambda_j(x_1, \dots, x_n)$ is the real single-valued measurable function of x_1, \dots, x_n obtained by substituting from (2.5) in μ_j .

THEOREM 2. *Let*

$$\mathcal{D}(x_1, \dots, x_n) = f(x_1, \dots, x_n),$$

and let

$$\mathcal{D}(x'_1, \dots, x'_n) = F(x'_1, \dots, x'_n),$$

where F vanishes whenever f vanishes. Let the real single-valued Borel-measurable functions y_1, \dots, y_p be defined as in (2.5), and let y'_1, \dots, y'_p be the same functions of x'_1, \dots, x'_n that y_1, \dots, y_p are of x_1, \dots, x_n . Suppose that, when we substitute for y_i and y'_i ,

$$(2.6) \quad f(x_1, \dots, x_n) = k(y_1, \dots, y_p) \cdot l(x_1, \dots, x_n)$$

and

$$(2.7) \quad F(x'_1, \dots, x'_n) = K(y'_1, \dots, y'_p) \cdot l(x'_1, \dots, x'_n).$$

Then if

$$\mathcal{D}(y_1, \dots, y_p) = h(y_1, \dots, y_p),$$

it follows that

$$\mathcal{D}(y'_1, \dots, y'_p) = \frac{K(y'_1, \dots, y'_p) \cdot h(y'_1, \dots, y'_p)}{k(y'_1, \dots, y'_p)},$$

where $\mathcal{D}(y'_1, \dots, y'_p) = 0$, whenever $k = 0$.

Proof. For any Borel-measurable set A of R^p , let

$$\psi_A(y'_1, \dots, y'_p) = \frac{K(y'_1, \dots, y'_p)}{k(y'_1, \dots, y'_p)}$$

if y'_1, \dots, y'_p is an element of the set A and $k(y'_1, \dots, y'_p) \neq 0$, and let $\psi_A(y'_1, \dots, y'_p) = 0$ otherwise. Also let

$$\phi_{A^{-1}}(x'_1, \dots, x'_n) = \frac{F(x'_1, \dots, x'_n)}{f(x'_1, \dots, x'_n)}$$

if x'_1, \dots, x'_n is an element of A^{-1} and $f(x'_1, \dots, x'_n) \neq 0$, where A^{-1} is the set of values of x'_1, \dots, x'_n such that* y'_1, \dots, y'_p is an element of A , and let $\phi_{A^{-1}}(x'_1, \dots, x'_n) = 0$ otherwise.

Then, from Lemma 1, for all Borel-measurable sets A of R^p

* The fact that y'_1, \dots, y'_p are Borel-measurable functions of x'_1, \dots, x'_n causes A^{-1} to be a Borel set of R_n .

$$\int \cdots \int_A \frac{K(t_1, \cdots, t_p) \cdot h(t_1, \cdots, t_p)}{k(t_1, \cdots, t_p)} dt_1 \cdots dt_p \\ = \int \cdots \int_{A^{-1}} F(x_1, \cdots, x_n) dx_1 \cdots dx_n.$$

Since the integral on the right is equal to the probability that y'_1, \cdots, y'_p is an element of A , the theorem follows from (1.1).

As an example of the possible uses of Theorem 2, let us consider the derivation of the joint distribution of variances and covariances of chance variables which have a normal multivariate distribution.*

Hotelling† has, by a simple geometric argument, obtained the joint distribution of sample correlation coefficients when all the "true" correlation coefficients are zero. In this case, it is known that the variances are independent of the correlation coefficients.‡ We may therefore state Hotelling's result more generally as follows:

Let $\mathcal{D}(x_{11}, \cdots, x_{pn}) = \prod_i \prod_j N(x_{ij}; 1)$. If

$$a_i = \sum_j x_{ij}^2; \quad r_{ij} = \frac{\sum_j x_{ij} x_{ij}}{a_i^{1/2} a_j^{1/2}},$$

then

$$\mathcal{D}(a_1, \cdots, a_p, r_{12}, \cdots, r_{p-1,p}) = \pi^{-p(p-1)/4} \omega^{(n-p+1)/2-1} \prod_j G(a_j; n-j+1, 1),$$

where

$$\omega = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ r_{1p} & r_{2p} & \cdots & 1 \end{vmatrix}.$$

We are able, by means of Theorem 2, to extend Hotelling's result to dependent chance variables.

Let $\mathcal{D}(x'_{11}, \cdots, x'_{pn}) = \prod_j N((x_j); (\sigma))$. If a'_i, r'_{ij} , and ω' are the same functions of x'_{11}, \cdots, x'_{pn} that a_i, r_{ij} , and ω are of x_{11}, \cdots, x_{pn} , then

$$\mathcal{D}(a'_1, \cdots, a'_p, r'_{12}, \cdots, r'_{p-1,p}) = 2^{-pn/2} \pi^{-p(p-1)/4} \sigma^{-n/2} \omega'^{(n-p+1)/2-1} \\ (2.8) \quad \cdot \left\{ \prod_j [\Gamma((n-j+1)/2)]^{-1} a_j'^{n/2-1} \right\} \cdot \exp \left[-\frac{1}{2} \sum_{i,j} \sigma^2 r'_{ij} a_i'^{1/2} a_j'^{1/2} \right].$$

* This distribution was first found by J. Wishart [39], p. 38, and [4], p. 270.

† See [19], p. 519.

‡ An analytic proof of this statement and Hotelling's distribution is readily obtained by means of Theorem 8.

This conclusion is immediate when one considers that in the notation of Theorem 2, $l(x_{11}, \dots, x_{pn}) = l(x'_{11}, \dots, x'_{pn}) \equiv 1$,

$$k(a_1, \dots, a_p, r_{12}, \dots, r_{p-1,p}) = (2\pi)^{-pn/2} \exp \left[-\frac{1}{2}(a_1 + \dots + a_p) \right]$$

and

$$K(a_1, \dots, a_p, r_{12}, \dots, r_{p-1,p}) = (2\pi)^{-pn/2} \sigma^{-n/2} \exp \left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} r_{ij} a_i^{1/2} a_j^{1/2} \right].$$

The transformation into variances and covariances is obvious.

The chief reason for the introduction of the following theorems on sufficient statistics is the desire to learn to what extent the requirements that (2.6) and (2.7) are satisfied restrict the applicability of Theorem 2. We will show, in the corollary to Theorem 3, that at least for sufficient statistics, using the definition based on (2.9), the chief value of Theorem 2 is that it permits the extension of probability densities derived on the assumption that certain parameters have fixed values, to probability densities which are functions of the values of those parameters just as in the preceding example.

Let the chance variable \mathbf{x} , have a probability density $f(x_r, \theta_1, \dots, \theta_q)$, ($q < n$), which depends on several parameters, $\theta_1, \dots, \theta_q$, and let the chance variables x_1, \dots, x_n be independently distributed. Let the real single-valued Borel-measurable functions y_1, \dots, y_p be defined as in (2.5). If

$$(2.9) \quad \prod_r f(x_r, \theta_1, \dots, \theta_q) = k(y_1, \dots, y_p, \theta_1, \dots, \theta_q) \cdot l(x_1, \dots, x_n),$$

where $k(y_1, \dots, y_p, \theta_1, \dots, \theta_q)$ and $l(x_1, \dots, x_n)$ are real single-valued measurable functions of their indicated variables, then y_1, \dots, y_p are called a sufficient system of statistics* with respect to the estimation of $\theta_1, \dots, \theta_q$.

Now let the chance variable \mathbf{x}' have a probability density $F(x'_r, \lambda_1, \dots, \lambda_s)$ which depends on several parameters $\lambda_1, \dots, \lambda_s$, and let the chance variables x'_1, \dots, x'_n be independently distributed.

The question suggested by (2.6) and (2.7) is the following: If

$$\prod_r F(x'_r, \lambda_1, \dots, \lambda_s) = K(y'_1, \dots, y'_p, \lambda_1, \dots, \lambda_s) \cdot l'(x'_1, \dots, x'_n),$$

where the Borel-measurable function y'_i is the same function of x'_1, \dots, x'_n that y_i is of x_1, \dots, x_n , and where K and l' are real single-valued measurable functions of their indicated variables, then what is the relation between $\mathcal{D}(\mathbf{x}_r)$ and $\mathcal{D}(\mathbf{x}'_r)$?

The answer to that question is contained, for the absolutely continuous

* Sufficient statistics were first defined by R. A. Fisher; see [9], p. 319. Also see [26], where certain definitions of sufficient statistics are compared.

statistics and probability densities, in the corollary to the following derivation of the form which $\mathcal{D}(x_r)$ must have in order that a system of sufficient statistics exist.*

THEOREM 3. Let $f(x_r, \theta_1, \dots, \theta_q)$ be a non-vanishing continuous function of $x_r, \theta_1, \dots, \theta_q$ on an open subset† V of‡ $R^1 \times R^q$, and let f be an absolutely continuous function of x_r on $V(\theta_1^0, \dots, \theta_q^0)$ for every set of fixed values, $\theta_1^0, \dots, \theta_q^0$, of $\theta_1, \dots, \theta_q$ in Θ . Set $\mathcal{D}(x_r) = f(x_r, \theta_1, \dots, \theta_q)$, and assume x_1, \dots, x_n to be independently distributed. Let (2.9) be satisfied, it now being assumed that y_i is an absolutely continuous function of x_r on an open subset B of R^1 containing A , for fixed values of $x_1, \dots, x_{r-1}, x_{r+1}, \dots, x_n$. If we assume that the jacobian, $J(\delta_1, \dots, \delta_p)$, where

$$J(\delta_1, \dots, \delta_p) = \frac{\partial(y_1, \dots, y_p)}{\partial(x_{\delta_1}, \dots, x_{\delta_p})},$$

does not vanish on $(B-Z)^n$, where Z is a closed set of measure zero, for any of the C_p^n possible selections $\delta_1, \dots, \delta_p$ of p of the positive integers $1, \dots, n$, then there exists an open subset V' of V such that the measure of $V - V'$ is zero, and $J(\delta_1, \dots, \delta_p)$ does not vanish on any of the set $[V'(\theta_1, \dots, \theta_q)]^n$ for all possible sets of values of $\delta_1, \dots, \delta_p$. Furthermore, at each point $x_r^0, \theta_1^0, \dots, \theta_q^0$ of V' , there is neighborhood $N = \omega \leq r$, where the definitions of ω and r are

$$\omega: |x_r - x_r^0| < h, \quad r: |\theta_i - \theta_i^0| < h_i, \quad i = 1, \dots, q,$$

in which there exist real single-valued absolutely continuous non-constant functions $\alpha_j(x_r)$ and $\gamma(x_r)$ of x_r , and real single-valued continuous functions $\beta_j(\theta_1, \dots, \theta_q)$ and $\beta(\theta_1, \dots, \theta_q)$ of $\theta_1, \dots, \theta_q$, such that in N

$$(2.10) \quad f(x_r, \theta_1, \dots, \theta_q) = \exp \left[\sum_j \alpha_j(x_r) \beta_j(\theta_1, \dots, \theta_q) + \gamma(x_r) + \beta(\theta_1, \dots, \theta_q) \right].$$

Proof. It is easy to see that if we define V' to be the intersection of V and $(B-Z) \times \Theta$, then V' is open, and the measure of $V - V'$ is zero. Now let the

* The result (2.10) has been obtained by Koopman and Darmois. See [23], p. 402 and [8], p. 1265. I have not been able to obtain the corollary to Theorem 3 from Koopman's proof, which depends on a definition of sufficient statistics that includes the definition based on (2.9). It may be remarked that Koopman needs only the continuity of the functions y_i whereas absolute continuity is assumed in Theorem 3. The proofs of Theorems 3 and 4 are generalizations of proofs for $p=q=1$ in (2.9) given by Doob in an unpublished seminar lecture at Columbia University. R. A. Fisher has derived Theorems 3 and 4 in cases which are specializations of those here treated. See [15].

† The totality of elements in the sets $V(\theta_1, \dots, \theta_q)$ will be denoted by A .

‡ The values of x_r are elements of R^1 , and the values of $\theta_1, \dots, \theta_q$ are elements of an open subset Θ of R^q .

neighborhood N be defined as in the hypotheses. Then the difference,

$$\sum_v [\log f(x_v, \theta_1, \dots, \theta_q) - \log f(x_v, \theta'_1, \dots, \theta'_q)],$$

is a real single-valued continuous function of the variables $x_1, \dots, x_n, \theta_1, \dots, \theta_q, \theta'_1, \dots, \theta'_q$, on $\omega^n \times r^2$ and is absolutely continuous in x_v for fixed values of the other variables. Hence the equation

$$\begin{aligned} \frac{\partial}{\partial x_v} [\log f(x_v, \theta_1, \dots, \theta_q) - \log f(x_v, \theta'_1, \dots, \theta'_q)] \\ = \sum_i \frac{\partial y_i}{\partial x_v} \cdot \frac{\partial}{\partial y_i} [\log k(y_1, \dots, y_p, \theta_1, \dots, \theta_q) \\ - \log k(y_1, \dots, y_p, \theta'_1, \dots, \theta'_q)] \end{aligned}$$

is valid throughout $\omega^n \times r^2$. Letting $\delta_j = j+1$ in $J(\delta_1, \dots, \delta_p)$, which involves no loss of generality, we see that the equation

$$\begin{aligned} \frac{\partial}{\partial y_i} [\log k(y_1, \dots, y_p, \theta_1, \dots, \theta_q) - \log k(y_1, \dots, y_p, \theta'_1, \dots, \theta'_q)] \\ = \sum_{\mu=2}^{p+1} a_{\mu i} \frac{\partial}{\partial x_\mu} [\log f(x_\mu, \theta_1, \dots, \theta_q) - \log f(x_\mu, \theta'_1, \dots, \theta'_q)], \end{aligned}$$

where the $a_{\mu i}$ is the cofactor of $\partial y_i / \partial x_\mu$ in $J(2, \dots, p+1)$ divided by $J(2, \dots, p+1)$, is valid throughout $\omega^n \times r^2$. Upon assigning fixed values, $x_2^0, \dots, x_n^0, \theta_1^0, \dots, \theta_q^0$ from $\omega^{n-1} \times r$ to $x_2, \dots, x_n, \theta_1, \dots, \theta_q$, we find that the equation

$$(2.11) \quad \frac{\partial}{\partial x_1} \log f(x_1, \theta_1, \dots, \theta_q) = h(x_1) + \sum_i \bar{\alpha}_i(x_1) \beta_i(\theta_1, \dots, \theta_q),$$

where

$$\begin{aligned} h(x_1) &= f(x_1, \theta_1^0, \dots, \theta_q^0), \\ \bar{\alpha}_i(x_1) &= \sum_i \frac{\partial y_i}{\partial x_1} a_{i+1, i}, \\ (2.12) \quad \beta_i(\theta_1, \dots, \theta_q) &= \frac{\partial}{\partial x_{i+1}} [\log f(x_{i+1}, \theta_1, \dots, \theta_q) - \log f(x_{i+1}, \theta_1^0, \dots, \theta_q^0)], \\ x_{i+1} &= x_{i+1}^0, \end{aligned}$$

is valid throughout $\omega \times r$. Then (2.10) may be obtained by integration of (2.11).

The absolute continuity and continuity of the functions may be easily proved. Obviously, if (2.10) is valid for all possible values of x_v and $\theta_1, \dots, \theta_q$, then the statistics, $\sum_v \alpha_j(x_v)$, are a sufficient system of statistics with respect to the estimation of $\theta_1, \dots, \theta_q$.

It is clear that a theorem similar to Theorem 3 may be proved for multivariate distributions.

The answer to the question of the relation of densities determining the same sufficient statistics is contained in the following corollary:

COROLLARY. *Let*

$$\begin{aligned}\mathcal{D}(x_v) &= f(x_v, \theta_1, \dots, \theta_q), \\ \mathcal{D}(x'_v) &= F(x'_v, \lambda_1, \dots, \lambda_s),\end{aligned}$$

and let f and F not vanish except on a set of values of x of measure zero for all values of $\theta_1, \dots, \theta_q$ and $\lambda_1, \dots, \lambda_s$ considered. Assume that $\mathcal{D}(x_v)$ and $\mathcal{D}(x'_v)$ satisfy the conditions of Theorem 3, the statistics y'_1, \dots, y'_p being the same functions of x'_1, \dots, x'_n that y_1, \dots, y_p are of x_1, \dots, x_n . Then in N and N' , when N' is a neighborhood of values of $x', \lambda_1, \dots, \lambda_s$ satisfying the conditions of Theorem 3, the x' values of N' being those of ω , the equations (2.10) and

$$(2.13) \quad F(x'_v, \lambda_1, \dots, \lambda_s) = \exp \left[\sum_j \alpha_j(x'_v) \beta'_j(\lambda_1, \dots, \lambda_s) + \beta'(\lambda_1, \dots, \lambda_s) + \gamma'(x'_v) \right]$$

are valid; hence $\mathcal{D}(x_v)$ and $\mathcal{D}(x'_v)$ differ only in the functions γ, β, β_j and $\gamma', \beta', \beta'_j$. Furthermore if

$$l'(x_1, \dots, x_n) \equiv l(x_1, \dots, x_n),$$

then $\gamma(x) \equiv \gamma'(x)$, any constant difference being eliminated by means of β and β' .

The proof is omitted. It depends on (2.12).

We now assume (2.9) to be valid throughout $R^1 \times \Theta$ and find the joint probability density of the sufficient statistics ϕ_1, \dots, ϕ_p where

$$\phi_j = \frac{1}{n} \sum_v \alpha_j(x_v).$$

Integrating f with respect to x_v we note that $\beta(\theta_1, \dots, \theta_q)$ is a function $\delta(\beta_1, \dots, \beta_p)$ of the values of $\beta_1(\theta_1, \dots, \theta_q), \dots, \beta_p(\theta_1, \dots, \theta_q)$, defined by the equation

$$\exp [\delta(\beta_1, \dots, \beta_p)] = \int_{-\infty}^{\infty} \exp \left[\sum_j \alpha_j(x) \beta_j(\theta_1, \dots, \theta_q) + \gamma(x) \right] dx.$$

THEOREM 4. Let $\mathcal{D}(\mathbf{x}_r)$ be given by (2.10) for all \mathbf{x}_r except on a set where $\mathcal{D}(\mathbf{x}_r) = 0$. Then

$$\mathcal{D}(\phi_1, \dots, \phi_p) = (2\pi)^{-p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left[i \sum_j t_j \phi_j - n\delta(\beta_1, \dots, \beta_p) + n\delta \left(\beta_1 + \frac{it_1}{n}, \dots, \beta_p + \frac{it_p}{n} \right) \right] dt_1 \cdots dt_p.$$

Proof. Since, by Lemma 1,

$$\mathcal{E} \left\{ \exp \left[i \sum_j t_j \phi_j \right] \right\} = \mathcal{E} \left\{ \exp \left[i \sum_j t_j \frac{\sum_r \alpha_j(\mathbf{x}_r)}{n} \right] \right\},$$

it follows that

$$\begin{aligned} \mathcal{E} \left\{ \exp \left[i \sum_j t_j \phi_j \right] \right\} \\ = \exp \left[-n\delta(\beta_1, \dots, \beta_p) + n\delta \left(\beta_1 + \frac{it_1}{n}, \dots, \beta_p + \frac{it_p}{n} \right) \right]. \end{aligned}$$

Then the proof is completed by using the Fourier transform theorem.*

One of the uses of Theorem 4 is to find the distribution of statistics which are functions of sufficient statistics, and which may, in fact, be an equivalent set of sufficient statistics.† It is easy to see that

$$\mathcal{E} \{ \alpha_j(\mathbf{x}) \} = \frac{\partial}{\partial \beta_j} \delta(\beta_1, \dots, \beta_p),$$

and that

$$\sigma_{jk} = \mathcal{E} \{ [\alpha_j(\mathbf{x}) - \mathcal{E} \{ \alpha_j(\mathbf{x}) \}] [\alpha_k(\mathbf{x}) - \mathcal{E} \{ \alpha_k(\mathbf{x}) \}] \} = \frac{\partial^2}{\partial \beta_j \partial \beta_k} \delta(\beta_1, \dots, \beta_p).$$

The purpose of the following theorem is to make possible the simplification of our proofs of Theorems 6, 7, and 8.

THEOREM 5. Let the real single-valued Borel-measurable functions,

$$(2.14) \quad y_r = f_r(x_1, \dots, x_n), \quad n > 1,$$

be defined on R^n . Let $y_{p_{\gamma-1}+1}, \dots, y_{p_{\gamma}}$ depend only on $x_1, \dots, x_{p_{\gamma}}$, and let $y_{p_{\gamma-1}+1}, \dots, y_{p_{\gamma}}$ be a linear transformation, having unit determinant, of $x_{p_{\gamma-1}+1}, \dots, x_{p_{\gamma}}$ for almost all sets of fixed values of $x_1, \dots, x_{p_{\gamma-1}}$, where the

* See, for example, [5], p. 191.

† Two sets of sufficient statistics may be called equivalent if both satisfy equations of the form (2.9).

transformation may depend on the fixed values of x_1, \dots, x_{p-1} and $n = p_m$. Then if A is a Borel-measurable set of values of y_1, \dots, y_n , and A^{-1} is the set of values of x_1, \dots, x_n such that y_1, \dots, y_n is an element of A , it follows that A^{-1} is Borel-measurable and has the same measure as A .

Proof. The proof is given for $m=2$. We shall use p for p_1 . It is well known that if A is Borel-measurable, then A^{-1} is Borel-measurable. Let us denote by $c_A(y_1, \dots, y_n)$ the function which has value one if y_1, \dots, y_n is an element of A and which is zero otherwise. If we denote the measure of A by $\mu(A)$, it follows that $c_A(y_1, \dots, y_n)$ is Borel-measurable if A is Borel-measurable, and

$$\mu(A) = \int \cdots \int_{R^n} c_A(y, \dots, y_n) dy_1 \cdots dy_n.$$

Now it may readily be shown that if $c_A(y_1, \dots, y_n)$ is Borel-measurable in R^n , and we assign fixed values, y_1^0, \dots, y_p^0 to y_1, \dots, y_p , then $c_A(y_1^0, \dots, y_p^0, y_{p+1}, \dots, y_n)$ is Borel-measurable in R^{n-p} .^{*} From Fubini's theorem,[†] it follows that

$$\begin{aligned} \mu(A) = \int \cdots \int_{R^p} \left[\int \cdots \int_{R^{n-p}} c_A(y, \dots, y_p, \right. \\ \left. y_{p+1}, \dots, y_n) dy_{p+1} \cdots dy_n \right] dy_1 \cdots dy_p. \end{aligned} \quad (2.15)$$

In a precisely similar fashion, it may be shown that

$$\begin{aligned} \mu(A^{-1}) = \int \cdots \int_{R^p} \left[\int \cdots \int_{R^{n-p}} c_{A^{-1}}(x, \dots, x_p, \right. \\ \left. x_{p+1}, \dots, x_n) dx_{p+1} \cdots dx_n \right] dx_1 \cdots dx_p. \end{aligned}$$

Inasmuch as the linear transformation having unit determinant is measure preserving, it follows that if our fixed values of x_1, \dots, x_p are determined by means of the assigned fixed values of y_1, \dots, y_p and the linear transformation, then, except for a set of measure zero,

$$\begin{aligned} \int \cdots \int_{R^{n-p}} c_A(y_1, \dots, y_p, y_{p+1}, \dots, y_n) dy_{p+1} \cdots dy_n \\ = \int \cdots \int_{R^{n-p}} c_{A^{-1}}(x_1, \dots, x_p, x_{p+1}, \dots, x_n) dx_{p+1} \cdots dx_n. \end{aligned}$$

^{*} See, for example, [31], p. 82

[†] See [31], p. 77.

Since the transformation of x_1, \dots, x_p into y_1, \dots, y_p is measure preserving, the proof is completed by transforming (2.15).

COROLLARY. *Let $\psi(y_1, \dots, y_n)$ be a real single-valued measurable function of y_1, \dots, y_n on R^n . Then $\psi(y_1, \dots, y_n)$ is a real single-valued measurable function $\phi(x_1, \dots, x_n)$ of x_1, \dots, x_n on R^n when we substitute from (2.14), and*

$$\int \cdots \int_A \psi(y_1, \dots, y_n) dy_1 \cdots dy_n = \int \cdots \int_{A^{-1}} \phi(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Proof. If A has finite measure, we may use Lemma 1 to complete the proof, and if A is not of finite measure, we express A as a sum of non-overlapping sets of finite measure and then proceed as above.

More general results than Theorem 5 and its corollary may readily be obtained. It is only necessary to require that A and A^{-1} be in 1-1 correspondence, to replace the words "a linear transformation having unit determinant" by the words "absolutely continuous functions with non-vanishing jacobian," and to replace the words "and has the same measure as A " by the words "and has measure

$$\int \cdots \int_{A^{-1}} J_1 \cdots J_m dx_1 \cdots dx_n,$$

where

$$J_\gamma = \left| \frac{\partial y_i}{\partial x_j} \right|, \quad i, j = p_{\gamma-1} + 1, \dots, p_\gamma."$$

Similar theorems may be proved when R^n has a completely additive measure function defined on it,* and the functions (2.14) are measurable with respect to that measure function.

The theorems stated next are of constant service in the analysis of the normal multivariate distribution.

THEOREM 6. *Let† $p = p_1 + \cdots + p_m$, let*

$$\mathcal{D}(x_{11}, \dots, x_{pn}) = \prod_{\gamma} \mathcal{D}(x_{p'_{\gamma-1}+1,1}, \dots, x_{p'_{\gamma},n}),$$

$$p'_0 = 0; p'_\gamma = p_1 + \cdots + p_\gamma,$$

and assume that a positive integer, $\gamma_0 < m$, exists such that for $\gamma > \gamma_0$

$$\mathcal{D}(x_{p'_{\gamma-1}+1,1}, \dots, x_{p'_{\gamma},n}) = g_\gamma(a_{ij}),$$

* See, for example, [31], chap. 1.

† In this theorem we do not define p_γ to be $n_1 + \cdots + n_m$.

where it is understood that g_γ is a function of all the a_{ij} , ($i, j = p_{\gamma-1} + 1, \dots, p'_\gamma$) and $a_{ij} = \sum_\nu x_{i\nu} x_{j\nu}$. Let the real single-valued Borel-measurable functions

$$y_{i\nu} = f_{i\nu}(x_{11}, \dots, x_{pn})$$

be defined on R^{pn} , and let the equations,

$$y_{i\nu} = \sum_{\mu=1}^n \phi_{\mu\nu}^{(\gamma)}(x_{11}, \dots, x_{p'_{\gamma-1}, n}) x_{i\mu}, \quad i = p'_{\gamma-1} + 1, \dots, p'_\gamma,$$

define a non-singular linear transformation of x_{i1}, \dots, x_{in} into y_{i1}, \dots, y_{in} for fixed values of $x_{11}, \dots, x_{p'_{\gamma-1}, n}$ and each value of i , the transformation being orthogonal if $\gamma > \gamma_0$.

Then, if

$$b_{ij} = \sum_\nu y_{i\nu} y_{j\nu},$$

it follows that

$$\mathcal{D}(y_{11}, \dots, y_{pn}) = \mathcal{D}(y_{11}, \dots, y_{p'_{\gamma_0-1}, n}) \cdot \prod_{\gamma=\gamma_0+1}^m \mathcal{D}(y_{p'_{\gamma-1}+1, 1}, \dots, y_{p'_\gamma, n}),$$

where

$$\mathcal{D}(y_{p'_{\gamma-1}+1, 1}, \dots, y_{p'_\gamma, n}) = g_\gamma(b_{ij})$$

and it is understood that g_γ is a function of all the b_{ij} , ($i, j = p'_{\gamma-1} + 1, \dots, p'_\gamma$).

Proof. The functions $y_{i\nu}$ have been so defined that the corollary to Theorem 5 may be applied.

We come now to the generalizations of the Fisher-Cochran theorem.

THEOREM 7. Let

$$\mathcal{D}(x_{1\nu}, \dots, x_{p\nu}) = N((x_\nu); (\sigma)),$$

and let

$$\mathcal{D}(x_{11}, \dots, x_{pn}) = \prod_\nu \mathcal{D}(x_{1\nu}, \dots, x_{p\nu}).$$

Let, for each value of i, j , and γ the real single-valued function

$$(2.16) \quad b_{ij\gamma} = \sum_{\mu, \nu=1}^n d_{\mu\nu}^\gamma x_{i\mu} x_{j\nu},$$

be a bilinear form in $x_{i\nu}$ and $x_{j\nu}$ with constant coefficients, and let the matrix of $b_{ij\gamma}$ be denoted by (d^γ) . Suppose that

$$(2.17) \quad \sum_\gamma b_{ij\gamma} = \sum_\nu x_{i\nu} x_{j\nu}.$$

Then, if n_γ be the rank of (d^γ) , a necessary and sufficient condition that there exist linear functions $z_{i\gamma}$ of x_{i1}, \dots, x_{in} such that $\mathcal{D}(z_{1\gamma}, \dots, z_{p\gamma}) = N((z_\gamma); (\sigma))$,

$$\mathcal{D}(z_{11}, \dots, z_{p\gamma}) = \prod_\gamma \mathcal{D}(z_{1\gamma}, \dots, z_{p\gamma})$$

and

$$\sum_{\nu=p_{\gamma-1}+1}^{p_\gamma} z_{i\nu} z_{j\nu} = b_{ij\gamma}$$

is

$$(2.18) \quad p_m = n.$$

Proof. The necessity of the condition (2.18) is obvious. In order to prove that (2.18) is sufficient, let $i=j=1$. Then by Theorem 1, there exist n linear functions (2.2) which have the desired properties. Consider the system of linear functions

$$(2.19) \quad z_{i\mu} = \sum_\nu c_{\mu\nu} x_{i\nu}, \quad \mu = 1, \dots, n,$$

the coefficients of which are those of the linear functions (2.2). The functions (2.19) are an orthogonal transformation of the chance variables $x_{i\nu}$ for fixed i and have the desired properties.

COROLLARY 1. *If*

$$(2.20) \quad n_\gamma \geq p, \quad \gamma = 1, \dots, k; k \leq m,$$

and if $n \geq p_k$, then the chance variables

$$(2.21) \quad b_{ij\gamma}, z_{i\mu}, \quad \gamma = 1, \dots, k; \mu = p_\gamma + 1, \dots, n,$$

have density

$$(2\pi)^{-p(p-1)k/4 - (n-p_k)/2} \sigma^{-n/2} \left\{ \prod_{\gamma=1}^k \prod_{i=1}^{p_\gamma} [\Gamma((n_\gamma - i + 1)/2)]^{-1} \right\} \\ \cdot \left\{ \prod_{\gamma=1}^k |b_{ij\gamma}|^{(n_\gamma - p + 1)/2 - 1} \right\} \exp \left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} \left\{ \sum_{\gamma=1}^k b_{ij\gamma} + \sum_{\nu=p_k+1}^n z_{i\nu} z_{j\nu} \right\} \right],$$

where $|b_{ij\gamma}|$ is a p rowed determinant for each value of γ .

Proof. The condition (2.20) permits the use of Wishart's distribution.*

COROLLARY 2. *Let $\mathcal{D}(x_{11}, \dots, x_{pn}) = \prod_\nu N((x_\nu); (\sigma))$, and let the bilinear forms (2.16) have the property (2.17). Then if the linear functions $z_{i\gamma}$ exist for $i=1, \dots, p; \gamma=1, \dots, p_{m-1}$, and if*

$$\mathcal{D}(z_{11}, \dots, z_{p, p_{m-1}}) = \prod_{\nu=1}^{p_{m-1}} N((z_\nu); (\sigma)),$$

* See equation (2.8) of this paper, or Wishart [39], p. 38, or Wishart and Bartlett [4], p. 270.

it follows that

$$n_m = n - p_{m-1},$$

that the complete set of variables (2.19) exists, and that

$$\mathcal{D}(z_{11}, \dots, z_{pn}) = \prod_{\nu} N((z_{\nu}); (\sigma)).$$

Proof. See [6], p. 181.*

In Theorem 7 a direct generalization of the Fisher-Cochran theorem was considered. In Theorem 8, the generalization is one which permits the use of more general quadratic forms than could be examined by means of Theorem 7. However, transformations of the chance variable (2.21) such as that of Wilks [34], p. 484, equation (30), will provide another derivation, by the moment method, of many of the distributions which we shall obtain, in §3, by means of Theorems 8 and 9.

THEOREM 8. Let the $n_1 + n_2 + \dots + n_p$ chance variables $x_{i\nu}$, ($i = 1, \dots, p$; $\nu = 1, \dots, n_i$), have probability density $\prod_{i=1}^p \prod_{\nu=1}^{n_i} N(x_{i\nu}; 1)$. Let the real single-valued Borel-measurable function

$$(2.22) \quad q_{ij} = \sum_{\mu, \nu=1}^{n_i} \omega_{\mu\nu}^{ij}(x_{11}, \dots, x_{i-1, n_{i-1}}) x_{i\mu} x_{i\nu}$$

of x_{11}, \dots, x_{in} , defined on $R^{n_1 + \dots + n_p}$, be, for each pair of values of i and j , ($i = 1, \dots, p$; $j = 1, \dots, m_i$), and for almost all sets of fixed values of $x_{11}, \dots, x_{i-1, n_{i-1}}$, a quadratic form in x_{i1}, \dots, x_{in_i} of rank n_{ij} , and let

$$(2.23) \quad \sum_{j=1}^{m_i} q_{ij} = \sum_{\nu=1}^{n_i} x_{i\nu}^2, \quad i = 1, \dots, p.$$

Then a necessary and sufficient condition that

$$(2.24) \quad \mathcal{D}(q_{11}, \dots, q_{p k_p}) = \prod_{i=1}^p \prod_{j=1}^{m_i} G(q_{ij}; n_{ij}, 1)$$

is

$$(2.25) \quad \sum_{j=1}^{m_i} n_{ij} = n_i.$$

* It is not difficult to generalize Theorems II and III of Cochran ([16], pp. 179, 181) by following the same procedure that has sufficed for the generalization of Cochran's Theorem II. It may be remarked that a necessary and sufficient condition for the satisfaction of equation (3.1) of Wilks ([37], p. 326) is provided by the generalization of Cochran's Theorem III. The generalizations of Cochran's Theorems I and II are also very useful in the moment approach to the theory of multivariate statistical analysis.

Proof. Necessity. The additive property of the χ^2 distribution suffices; for, $\sum_{\nu=1}^{n_i} x_{i\nu}^2$ has the χ^2 distribution with n_i degrees of freedom, and from (2.22) and (2.23), $\sum_{\nu=1}^{n_i} x_{i\nu}^2$ has the χ^2 distribution with $n_{i1} + n_{i2} + \cdots + n_{im_i}$ degrees of freedom.

Sufficiency. We obtain the analogue of Theorem 1 for the hypotheses of Theorem 8. For each value of i , assigning fixed values to $x_{i1}, \cdots, x_{i-1, n_{i-1}}$, we obtain from Theorem 1 n_i real linear functions

$$(2.26) \quad z_{i\nu} = \sum_{\mu=1}^{n_i} c_{\mu\nu}^i x_{i\mu},$$

which are such that the corresponding chance variables have density

$$\prod_{\nu=1}^{n_i} N(z_{i\nu}; 1);$$

and for the fixed values of $x_{i1}, \cdots, x_{i-1, n_{i-1}}$

$$\sum_{\nu} z_{i\nu}^2 = q_{ij}, \quad j = 1, \cdots, m_i,$$

where ν runs from $n_{i1} + \cdots + n_{i, j-1} + 1$ through $n_{i1} + \cdots + n_{ij}$.

The coefficients $c_{\mu\nu}^i$, of the linear forms (2.26) are real single-valued Borel-measurable functions of the coefficients of q_{ij} for fixed values of $x_{i1}, \cdots, x_{i-1, n_{i-1}}$. Let $c_{\mu\nu}^i(\omega)$ be the same function of the functions

$$\omega_{\mu\nu}^i(x_{i1}, \cdots, x_{i-1, n_{i-1}}), \quad \mu, \nu = 1, \cdots, n_i,$$

that $c_{\mu\nu}^i$ is of the coefficients of the quadratic form having constant coefficients. Since Borel-measurable functions of Borel-measurable functions are Borel-measurable, it follows that the functions

$$\begin{aligned} y_{i\nu} &= z_{i\nu}(x_{i1}, \cdots, x_{in_i}) \\ &= \sum_{\mu=1}^{n_i} c_{\mu\nu}^i(\omega) x_{i\mu} \end{aligned}$$

are Borel-measurable functions of x_{i1}, \cdots, x_{in_i} . For almost all sets of fixed values of $x_{i1}, \cdots, x_{i-1, n_{i-1}}$, it is clear that y_{i1}, \cdots, y_{in_i} is a linear orthogonal transformation of x_{i1}, \cdots, x_{in_i} . Hence the hypotheses of Theorem 5 are satisfied. Furthermore, the functions $y_{i\nu}$ have been so defined that

$$\sum_{\nu} y_{i\nu}^2 = q_{ij}, \quad j = 1, \cdots, m_i,$$

where ν runs from $n_{i1} + \cdots + n_{i, j-1} + 1$ through $n_{i1} + \cdots + n_{ij}$. Then by the corollary to Theorem 5, it follows that

$$\mathcal{D}(y_{11}, \dots, y_{pn_p}) = \prod_{i=1}^p \prod_{r=1}^{n_i} N(y_{ir}; 1),$$

and the proof is completed by using the additive property of the χ^2 distribution.

THEOREM 9. Let $\mathcal{D}(x_{11}, \dots, x_{pn_p}) = \prod_{i=1}^p N(x_i; (\sigma))$. Let $n_1 = \dots = n_p = n$ in Theorem 8, and let the functions (2.22) satisfy (2.23). If, for some i and j , $\sigma^{ij} \neq 0$, let $\psi_{ij}(q_{11}, \dots, q_{pm_p})$ be a real-valued Borel-measurable function of its indicated variables such that

$$\psi_{ij}(q_{11}, \dots, q_{pm_p}) = \sum_{\nu} x_{i\nu} x_{j\nu}.$$

Then a necessary and sufficient condition that

$$(2.27) \quad \mathcal{D}(q_{11}, \dots, q_{pm_p}) = \left\{ \prod_i \prod_{r=1}^{m_i} [\Gamma(n_{ir}/2)]^{-1} q_{ir}^{n_{ir}/2-1} \right\} \\ \cdot 2^{-pn/2} \sigma^{-n/2} \exp \left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} \psi_{ij}(q_{11}, \dots, q_{pm_p}) \right]$$

is

$$(2.28) \quad \sum_{r=1}^{m_i} n_{ir} = n.$$

Proof. Necessity. The necessity of the condition (2.28) may be demonstrated as in the proof of Theorem 8.

Sufficiency. If, in Theorem 2, we let $l = l' \equiv 1$,

$$K(q_{11}, \dots, q_{pm_p}) = (2\pi)^{-np/2} \sigma^{-n/2} \exp \left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} \psi_{ij}(q_{11}, \dots, q_{pm_p}) \right],$$

and

$$k(q_{11}, \dots, q_{pm_p}) = (2\pi)^{-np/2} \exp \left[-\frac{1}{2} \sum_i \psi_{ii}(q_{11}, \dots, q_{pm_p}) \right],$$

then the functions (2.22) are such that (2.6) and (2.7) are true. Hence, the desired result, (2.27), follows from the straightforward application of Theorems 2 and 8.

It is apparent that many distributions of functions of chance variables having a normal multivariate distribution may be derived by using Theorem 9. Some of these will be considered in §3.

3. Vector variates. In this section we apply the methods developed in §2

to various problems in the generalized analysis of variance, relations between sets of variables, and generalized periodogram analysis.

In order to do this with a minimum of duplication, two theorems in the theory of quadratic forms are first stated. These theorems and their corollaries are useful in the theory of univariate statistical analysis. In Theorems 12 and 13, the previous theorems are used to obtain joint distributions of functions which are quadratic forms in certain variables for fixed values of other variables. These latter distributions are then transformed in order to derive distributions which are useful in the theory of vector variates. The notation and terminology have already been given in the first section of this paper.

Let

$$\sigma_i = \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1i} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2i} \\ . & . & \cdots & . \\ \sigma_{i1} & \sigma_{i2} & \cdots & \sigma_{ii} \end{vmatrix}, \quad \sigma_p = \sigma,$$

let $\sigma^i = \sigma_i / \sigma_{i-1}$, and let σ^{ijk} be the cofactor of σ_{jk} in σ_i divided by σ_i , ($j, k = 1, \cdots, i$), $\sigma^{jkp} = \sigma^{jk}$.

Set

$$a_{ij} = \sum_p x_{ip} x_{jp},$$

and assume that a_i , a^i , and a^{ijk} are the same functions of the a_{ij} that σ_i , σ^i , and σ^{ijk} are of the σ_{ij} . We shall refer to any determinant a_i as a generalized sum of squares since

$$a_i = \sum_{\mu} (\Delta_{\mu})^2,$$

where

$$(3.1) \quad \Delta_{\mu} = \begin{vmatrix} x_{1\mu_1} & x_{1\mu_2} & \cdots & x_{1\mu_i} \\ x_{2\mu_1} & x_{2\mu_2} & \cdots & x_{2\mu_i} \\ . & . & \cdots & . \\ x_{i\mu_1} & x_{i\mu_2} & \cdots & x_{i\mu_i} \end{vmatrix},$$

and the summation is for all the C_i^n distinct selections of i integers $\mu_1, \mu_2, \cdots, \mu_i$ from $1, \cdots, n$.

We remark that if β_{ig} is the regression coefficient of x_g in the "true" regression function of x_i on x_1, \cdots, x_{i-1} , where x_1, \cdots, x_p , ($p \geq i$), are assumed to have a normal multivariate distribution, then

$$\beta_{ig} = - \frac{\sigma^{ig i}}{\sigma^{i i i}}.$$

Similarly, if b_{ig} is the regression coefficient of x_g in the sample regression function of x_i on x_1, \dots, x_{i-1} , then

$$b_{ig} = - \frac{a^{ig i}}{a^{iii}}.$$

In the following theorem it is shown that the coefficients of a transformation by which a definite quadratic form may be reduced to a sum of squares are the regression coefficients. It is necessary to emphasize that although much of this material is known, yet no complete treatment seems to exist.

THEOREM 10. *There exist p linear forms*

$$(3.2) \quad y_i = x_i - \sum_g \beta_{ig} x_g$$

such that

$$q = \sum_i y_i^2 / \sigma^i.$$

The proof is omitted. It is noted that the transformation (3.2) has unit determinant.

COROLLARY 1. *Let*

$$\mathcal{D}(x_{1\nu}, \dots, x_{p\nu}) = N((x_\nu); (\sigma)),$$

and let

$$\mathcal{D}(x_{11}, \dots, x_{pn}) = \prod_\nu \mathcal{D}(x_{1\nu}, \dots, x_{p\nu}).$$

Then if $y_{i\nu} = x_{i\nu} - \sum_g \beta_{ig} x_{g\nu}$, it follows that

$$\mathcal{D}(y_{i\nu}) = N(y_{i\nu}; \sigma^i)$$

and

$$\mathcal{D}(y_{11}, \dots, y_{pn}) = \prod_\nu \prod_i \mathcal{D}(y_{i\nu}).$$

The proof is omitted.

In Theorem 10 and Corollary 1 it was shown that variables which have a joint normal distribution may be expressed in terms of variables which are distributed independently in normal distributions. In the following corollaries it will be shown that certain functions of the dependent variables are invariant under the transformation into independent variables and hence have the distributions of the same functions of independent variables.

COROLLARY 2. Let a'_i be the generalized sum of squares with the elements

$$a'_{jk} = \sum_{\nu} y_{j\nu} y_{k\nu}, \quad j, k = 1, \dots, i.$$

Then

$$(3.3) \quad a'_i = a_i.$$

Proof. Equation (3.3) is obvious when a_i and a'_i are written as sums of squares of determinants as in (3.1).

COROLLARY 3. If $a'^i = a'_i / a'_{i-1}$,* then, except for a set of measure zero,

$$(3.4) \quad a'^i = a'_{ii} - \sum_{\theta} w_{i\theta}^2,$$

where

$$(3.5) \quad w_{i\theta} = \sum_{\nu} p_{\theta\nu} y_{i\nu}$$

and the coefficients

$$(3.6) \quad p_{\theta\nu} = (a'^{\theta})^{-1/2} \left(y_{\theta\nu} - \sum_{j=1}^{\theta-1} b'_{\theta j} y_{j\nu} \right)$$

are such that

$$(3.7) \quad \sum_{\nu} p_{\theta\nu} p_{k\nu} = \delta_{\theta k}.$$

Proof. It follows from the definition of a'^i that†

$$(3.8) \quad a'^i = a'_{ii} - \sum_{\theta, h} a'_{i\theta} a'_{ih} a'^{\theta h i-1}.$$

From Theorem 1 and (3.8), it follows that

$$a'_{ii} - a'^i = \sum_{\theta} w_{i\theta}^2,$$

for almost all sets of fixed values of $y_{11}, \dots, y_{i-1, n}$, where‡

$$(3.9) \quad \begin{aligned} w_{i\theta} &= (a'^{\theta})^{-1/2} \left(a'_{i\theta} - \sum_{k=1}^{\theta-1} b'_{\theta k} a'_{ik} \right) \\ &= (a'^{\theta})^{1/2} b'_{i \cdot \theta}. \end{aligned}$$

* The functions a'^{ijk} and $b'_{i\theta}$ are to be defined as a^{ijk} and $b_{i\theta}$ but with a'_{jk} replacing a_{jk} , ($j, k = 1, \dots, i$). From Corollary 2 it follows that $a'^i = a^i$.

† It is, of course, well known that two alternative expressions for a'^i are $a'^i = a_{ii} - \sum_{\theta} a_{i\theta} b'_{i\theta}$ and $a'^i = a_{ii} - \sum_{\theta, h} b'_{i\theta} b'_{ih} a'_{\theta h}$.

‡ The function $b_{i \cdot \theta}$ is the coefficient of x_{θ} in the sample regression equation of x_i on x_1, \dots, x_{θ} , and $b'_{i \cdot \theta}$ is the same function of the $y_{i\nu}$.

Hence, equations (3.4), (3.5), and (3.6) are true. Equation (3.7) may be verified by direct computation, or by employing the known properties of deviations from sample regression functions to avoid that computation.

COROLLARY 4. *The function w_{ig} may be written as*

$$(3.10) \quad w_{ig} = (a^i)^{1/2} \left(b_{i \cdot g} - \sum_{k=g+1}^{i-1} b_{k \cdot g} \beta_{ik} - \beta_{ig} \right).$$

Proof. From (3.9) it follows that

$$w_{ig} = (a'^g)^{-1/2} (a'^{g-1})^{-1} \begin{vmatrix} a'_{11} & \cdots & a'_{1g} \\ \cdot & \cdots & \cdot \\ a'_{g-1,1} & \cdots & a'_{g-1,g} \\ a'_{i1} & \cdots & a'_{ig} \end{vmatrix}.$$

Now it may be shown that if

$$K'_{qr} = \begin{vmatrix} a'_{11} & \cdots & a'_{1,g-1} & a'_{1q} \\ \cdot & \cdots & \cdot & \cdot \\ a'_{g-1,1} & \cdots & a'_{g-1,g-1} & a'_{g-1,q} \\ a'_{r1} & \cdots & a'_{r,g-1} & a'_{r,q} \end{vmatrix},$$

where q and r are positive integers such that $g \leq q \leq i$; $g \leq r \leq i$, then

$$K'_{qr} = \sum_{\mu=g}^q \sum_{\nu=g}^r \beta_{q\mu} \beta_{r\nu} K_{\mu\nu},$$

where K_{qr} is the same function of the a_{ij} that K'_{qr} is of the a'_{ij} . Then (3.10) is a consequence of the fact that

$$b_{kg} = \frac{K_{kg}}{a_g}, \quad k = g, \cdots, i.$$

Although Theorem 10 and Corollaries 1-4 are more than sufficient for the consideration of the distribution of the generalized sum of squares, they do not yield the distribution of the generalized correlation ratio and related statistics. It is for this reason that the following definitions, theorem, and corollaries are introduced.

Let

$$a_{ij\gamma} = \sum_{\nu} x_{i\nu} x_{j\nu}, \quad n_0 = 0; n_{\gamma} \geq p; n_1 + \cdots + n_m \leq n,$$

where ν runs from $n_1 + \cdots + n_{\gamma-1} + 1$ through $n_1 + \cdots + n_{\gamma}$, and let $a_{i(\gamma)}$ and $a_{(\gamma)}^i$ be defined as are a_i and a^i but with $a_{kij\gamma}$ replacing a_{kij} . Let $a'_{ij\gamma}$, $a'_{i(\gamma)}$,

and $a'_{(\gamma)}$ be similar functions of the y_{iv} . Then there exist linear functions $w_{iv}(\gamma)$ such that Corollary 3 is true for each value of γ .

Let s be a positive integer not greater than p , and let

$$\tau_i = \begin{vmatrix} \sigma_{ii} & \sigma_{i,s+1} & \cdots & \sigma_{ip} \\ \sigma_{s+1,i} & \sigma_{s+1,s+1} & \cdots & \sigma_{s+1,p} \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{pi} & \sigma_{p,s+1} & \cdots & \sigma_{pp} \end{vmatrix}, \quad i = 1, \cdots, s.$$

Let τ^{ijk} be the cofactor of σ_{jk} in τ_i divided by τ_i , ($j, k = i, s+1, \cdots, p$; $i = 1, \cdots, s$). Let τ_0 be the cofactor of σ_{ii} in τ_i , and let τ_0^{ik} be the cofactor of σ_{jk} in τ_0 divided by τ_0 . Let d be the same function of the a_{ij} that τ_0 is of the σ_{ij} .

THEOREM 11. *There exist p linear forms, for a preassigned value of v ,*

$$(3.11) \quad \begin{aligned} z_{iv} &= x_{iv} + \sum_{j=s+1}^p \frac{\tau^{iji}}{\tau^{iii}} x_{jv}, & i &= 1, \cdots, s, \\ z_{iv} &= x_{iv}, & i &= s+1, \cdots, p, \end{aligned}$$

such that

$$q_v = \sum_{i,j=1}^s \sigma^{ij} z_{iv} z_{jv} + \sum_{i,j=s+1}^p \tau_0^{ij} z_{iv} z_{jv}.$$

The proof is omitted. It is noted that the transformation (2.36) has unit determinant.

R. A. Fisher* has evaluated the difference between the sum of squares of the deviations of y_i from the sample regression function in y_1, \cdots, y_{i-1} , and the sum of squares of the deviations of y_i from the sample regression function in y_1, \cdots, y_k , ($k < i-1$). In order to state his result in terms of the functions which are generally computed in least squares solutions, we define the $i-k-1$ rowed determinant

$$\alpha' = |a'^{it, i-1}|, \quad j, t = k+1, \cdots, i-1,$$

and let α'_{jt} be the cofactor of $a'^{it, i-1}$ in α' divided by α' . Then we may state Fisher's result as the following corollary:

COROLLARY 1. *If $a'^{i(k)}$ is the sum of squares of the deviations of y_i from its sample regression function in y_1, \cdots, y_k , then the difference, $a'^{i(k)} - a'^i$, is, for almost all sets of fixed values of $y_{11}, \cdots, y_{i-1,n}$, a positive definite quadratic form in y_{i1}, \cdots, y_{in} of rank $i-k-1$:*

$$\sum_{j,t=k+1}^{i-1} b'_{ij} b'_{it} \alpha'_{jt}.$$

* See [16], §29.1.

The proof is omitted. A simple proof may be given by means of Theorem 11 and the Jacobi ratio theorem* in determinants.

We now use Corollary 1 to determine the change in the rank of a^t induced by the omission of several observations.

Let us define, for each value of t , ($t = 1, \dots, n-m$), a variable which assumes only the values zero and one,

$$y_{i+t, \nu} = \delta_{\nu, n_1+t}.$$

Then $a'_{(1)}$ is the sum of squares of the deviations of y_i from its sample regression function in $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+n-n_1}$.

COROLLARY 2. *The difference*

$$a'^i - a'_{(1)}^i$$

is, for fixed i and almost all sets of fixed values of $y_{11}, \dots, y_{i-1, n}$, a quadratic form in y_{i1}, \dots, y_{in} of rank $n - n_1$.†

The proof is omitted.

The following corollary is useful in determining whether or not a set of regression coefficients differ significantly from one another.

Let a_i'' be the generalized sum of squares, the elements of which are $\sum_{\gamma} a_{ijk\gamma}$, ($j, k = 1, \dots, i$), and let $v_i = a''^i - a_{(1)}'' - \dots - a_{(m)}''$.

COROLLARY 3. *For any fixed value of i , and for almost all sets of fixed values of $y_{11}, \dots, y_{i-1, n}$, the function v_i is a quadratic form in y_{i1}, \dots, y_{in} of rank $(m-1)(i-1)$ and is equal to*

$$(3.12) \quad \sum_{\gamma} \sum_{g, h} a_{gh\gamma} (b_{ig} - b_{ig}^{(\gamma)}) (b_{ih} - b_{ih}^{(\gamma)}). \ddagger$$

The proof is omitted.§

Let $n' = n_1 + \dots + n_m$, and let $u_i = a'^i - a''^i$, where $a''^i = a_i'' / a_{i-1}''$. Then we can state the following theorem:

THEOREM 12. *Let*

$$\mathcal{D}(x_1, \dots, x_{p'}) = N((x_r); (\sigma)),$$

and let

* Turnbull [32], p. 77.

† The formula obtained by substituting in (3.11) does not provide a simple evaluation of the difference.

‡ The functions $b_{ig}^{(\gamma)}$ are the regression coefficients in the γ th set of observations; in other words, $b_{ig}^{(\gamma)}$ is defined as is b_{ig} but with $a_{kij\gamma}$ replacing a_{kij} .

§ Certain special cases of (3.12) have been used by Welch [33] and Kolodziejczyk [22].

$$\mathcal{D}(\mathbf{x}_{11}, \dots, \mathbf{x}_{pn}) = \prod_{\gamma} \mathcal{D}(\mathbf{x}_{1\gamma}, \dots, \mathbf{x}_{p\gamma}).$$

Then if $n > n'$, the chance variables

$$(3.13) \quad \mathbf{a}'_{(\gamma)}, \quad \mathbf{w}_{ig}^2, \quad v_i, \quad u_i$$

are independently distributed, and have joint probability density

$$(3.14) \quad \prod_i \left\{ G(u_i; n - n', \sigma^i) \cdot G(v_i; (m-1)(i-1), \sigma^i) \right. \\ \left. \left[\prod_g G(w_{ig}^2; 1, \sigma^i) \right] \left[\prod_{\gamma} G(a'_{(\gamma)}; n_{\gamma} - i + 1, \sigma^i) \right] \right\}.$$

If $n = n'$, the chance variables $\mathbf{a}'_{(\gamma)}$, \mathbf{w}_{ig}^2 , and v_i are independently distributed and have a joint probability density which may be obtained from (3.14) by suppressing the terms involving u_i and v_1 .

Proof. The chance variables (3.13) have been shown, in the corollaries to Theorems 10 and 11, to satisfy the conditions of Theorem 8. The proof is then completed by means of Theorem 9.*

The following corollaries are devoted to a consideration of the densities, moments and possible ranges of certain chance variables. Applications of the distributions are indicated.

COROLLARY 1. *Under the hypotheses of Theorem 12, it follows that*

$$\mathcal{D}(\mathbf{a}^i) = G(\mathbf{a}^i; n - i + 1, \sigma^i),$$

and

$$\mathcal{D}(\mathbf{a}^1, \dots, \mathbf{a}^p) = \prod_i \mathcal{D}(\mathbf{a}^i).$$

Furthermore $\mathcal{E}\{\prod_i (\mathbf{a}^i)^{\lambda_i/2}\} = \prod_i \bar{G}(\lambda_i, n - i + 1, \sigma^i)$ and $0 < a^i < \infty$.

Proof. See Theorem 12 and Theorem 10, Corollary 2.

Since $a_j = a^1 \cdots a^i$, there may be deduced from Corollary 1 the distribution and moments of the generalized variance,† the ratio of the generalized variance to any of its principal minors,‡ and the quotient of any two of its principal minors. There may also be deduced the joint distribution and moments of the generalized variance and some of its principal minors.

* This distribution is a generalization of that obtained by Bartlett [1], p. 268, starting from Wishart's distribution. It may be remarked that the use of Theorem 6 in conjunction with Theorem 12 yields a similar generalization of the results of Bartlett's §2.

† See Wilks [34], pp. 477, 481.

‡ In the general case one has to transform $\mathcal{D}(\mathbf{a}^1, \dots, \mathbf{a}^p)$ to find the probability that a ratio of products of the variables \mathbf{a}^i exceeds a given quantity.

In the theory of relations between two sets of variables,* there occurs a statistic which measures the portion of the generalized variance of the first set of variables which is due to the variables of the second set.

Let $\zeta_{i\nu} = x_{i\nu} - z_{i\nu}$, and assume $\xi_{i\nu}$ to be the same function of the a_{ij} and $x_{i\nu}$ that $\zeta_{i\nu}$ is of the σ_{ij} and $x_{i\nu}$. Furthermore, let

$$\begin{aligned} g_{ij} &= \sum_{\nu} (x_{i\nu} - \zeta_{i\nu})(x_{j\nu} - \zeta_{j\nu}), \\ (3.15) \quad g_{ij1} &= \sum_{\nu} (x_{i\nu} - \xi_{i\nu})(x_{j\nu} - \xi_{j\nu}), \\ g_{ij2} &= \sum_{\nu} (\xi_{i\nu} - \zeta_{i\nu})(\xi_{j\nu} - \zeta_{j\nu}), \end{aligned}$$

and let

$$\begin{aligned} (3.16) \quad a_{ij1} &= g_{ij1}, \\ a_{ij2} &= \sum_{\nu} \xi_{i\nu} \xi_{j\nu}. \end{aligned}$$

It is well known that $g_{ij} = g_{ij1} + g_{ij2}$ and $a_{ij} = a_{ij1} + a_{ij2}$. Obviously

$$g_{ij} = a_{ij}, \quad i, j = s+1, \dots, p.$$

We shall define the residual variance in terms of what may be called the generalized covariances. Set $s \leq p-s$, and let†

$$W'_{js} = (-1)^{s-j+1} \begin{vmatrix} 0 & \cdots & 0 & g_{j,s+1} & \cdots & g_{jp} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & 0 & g_{s,s+1} & \cdots & g_{sp} \\ g_{s+1,j} & \cdots & g_{s+1,s} & g_{s+1,s+1} & \cdots & g_{s+1,p} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ g_{pj} & \cdots & g_{ps} & g_{p,s+1} & \cdots & g_{pp} \end{vmatrix}, \quad j = 1, \dots, s.$$

Then we note that from Theorem 11 and Theorem 8, the statistic

$$B'_{js} = \frac{W'_{js}}{d} = \frac{\begin{vmatrix} g_{jj2} & \cdots & g_{js2} \\ \cdot & \cdots & \cdot \\ g_{sj2} & \cdots & g_{ss2} \end{vmatrix}}{d}$$

is the generalized variance of $p-s$ observations of a vector variate which has $s-j+1$ components and hence has the distribution and moments of such a

* See Hotelling [20]. A knowledge of §§1-4 of that paper is assumed.

† By W_{js} we shall mean the same function of the a_{ij} that W'_{js} is of the g_{ij} . It is noted that W_{1s} is the numerator of the statistic q^2 defined by Hotelling [20], p. 333.

variance.* From Theorems 11 and 12 it follows that the parametric function estimated by B'_{1s} is σ/τ .

Assume $k > j$. Since

$$\frac{W'_{js}}{W'_{ks}} = \frac{B'_{js}}{B'_{ks}},$$

its distribution may readily be found in accordance with the preceding discussion of the distribution of the ratio of a generalized variance to one of its principal minors.

Since B'_{js} and d are independent chance variables, the chance variable W'_{js} is a product of independent variances and has the distribution and moments of such a product.†

COROLLARY 2. (a) If $n > n'$, the chance variables $a'_{(\gamma)}$, a^i have density

$$(3.17) \quad \prod_i G(a^i - a'_{(1)} - \cdots - a'_{(m)}; n - n' + (j-1)(m-1), \sigma^i) \cdot \left[\prod_{\gamma} G(a'_{(\gamma)}; n_{\gamma} - j + 1, \sigma^j) \right].$$

Furthermore

$$(3.18) \quad \mathcal{E} \left\{ \prod_i (a^i)^{\lambda_i/2} \left[\prod_{\gamma} (a'_{(\gamma)})^{\lambda_{i\gamma}/2} \right] \right\} \\ = \prod_i \left\{ \bar{G} \left(\lambda_i; n - j + 1 + \sum_{\gamma} \lambda_{i\gamma}, \sigma^i \right) \cdot \left[\prod_{\gamma} \bar{G}(\lambda_{i\gamma}; n_{\gamma} - j + 1, \sigma^j) \right] \right\},$$

and

$$(3.19) \quad 0 < a'_{(\gamma)} < a^i - a'_{(m)} - \cdots - a'_{(\gamma+1)}, \quad a'_{(m+1)} = 0.$$

(b) If $n = n'$, the chance variables

$$\begin{aligned} a^1_{(\gamma)}, & \quad \gamma = 1, \cdots, m-1, \\ a^j_{(\gamma)}, & \quad j = 2, \cdots, p; \gamma = 1, \cdots, m, \\ a^j, & \quad j = 1, \cdots, p, \end{aligned}$$

* The residual variance is defined to be $B_{1s} = W_{1s}/d$, and has the distribution of B_{1s} if and only if $g_{i2} = a_{i2}$, ($i, j = 1, \cdots, s$).

† If $s = p - s = 2$, the statistic W_{12} is the square of the covariance tetrad difference. Hence if the two sets are independent, the distribution and moments of the covariance tetrad difference have been derived. It is noted that the variables of either set may be intercorrelated. The correlation tetrad difference will be discussed after Corollary 3. This analysis generalizes the discussions of Wilks [35], and Wishart [40].

have a joint probability density, joint moments, and possible ranges which are of the same form as (3.17), (3.18), and (3.19) with the obvious differences caused by the omission of $a'_{(m)}$.

Proof. The corollary is an immediate result of the theorem on the transformation of multiple integrals.

If $n > n'$, the joint density and moments of the chance variables $a_{j(\gamma)}$, a_j , ($j = 1, \dots, i$), may readily be obtained by transforming (3.17). The jacobian of the transformation

$$a_{j(\gamma)} = a_{(\gamma)}^1 \cdots a_{(\gamma)}^j \\ a_j = a^1 \cdots a^i$$

is

$$\prod_{\theta} a_{\theta}^{-1} \left[\prod_{\gamma} a_{\theta(\gamma)}^{-1} \right],$$

and the possible ranges of the chance variables are

$$(3.20) \quad 0 < a_{k(\gamma)} < \prod_{j=1}^k \left(\frac{a_j}{a_{j-1}} - \frac{a_{j(m)}}{a_{j-1(m)}} - \cdots - \frac{a_{j(\gamma+1)}}{a_{j-1(\gamma+1)}} \right), \\ 0 < a_{k(\gamma)} < (a_k^{1/k} - a_{k(m)}^{1/k} - \cdots - a_{k(\gamma+1)}^{1/k})^k.$$

It is noted that if $n = n'$, one may find the joint density of the chance variables

$$a_{1(\gamma)}, \quad a_1, \quad \gamma = 1, \dots, m-1, \\ a_{j(\gamma)}, \quad a_j, \quad j = 2, \dots, i; \gamma = 1, \dots, m,$$

in a similar fashion. The moments of the chance variables are most easily obtained by means of Lemma 1 and (3.18).

The inequalities (3.20) are consequences of the fact that

$$0 < a_{k(\gamma)} < \prod_{j=1}^k (a^j - a_{(m)}^j - \cdots - a_{(\gamma+1)}^j)$$

and of the application of the procedure of finding the extrema of a function subject to several conditions.

From these distributions may be obtained the joint distribution of

$$a_{i(1)}, \dots, a_{i(m)}, a_i$$

by integrating the other variables over their possible ranges as stated in (3.20).

Let $c_{(\gamma)}^j = a_{(\gamma)}^j / a^j$. Then we may state the following corollary:

COROLLARY 3. (a). If $n > n'$, then

$$(3.21) \quad \begin{aligned} \mathcal{D}(\mathbf{c}_{(1)}^i, \dots, \mathbf{c}_{(m)}^i) \\ = D(\mathbf{c}_{(1)}^i, \dots, \mathbf{c}_{(m)}^i; n_1 - i + 1, \dots, n_m - i + 1, n - n' + (i - 1)(m - 1)), \\ \mathcal{D}(\mathbf{c}_{(1)}^i, \dots, \mathbf{c}_{(m)}^i, \mathbf{a}^i) = \mathcal{D}(\mathbf{a}^i) \cdot \mathcal{D}(\mathbf{c}_{(1)}^i, \dots, \mathbf{c}_{(m)}^i), \end{aligned}$$

and

$$\mathcal{D}(\mathbf{c}_{(1)}^1, \dots, \mathbf{c}_{(m)}^p, \mathbf{a}^1, \dots, \mathbf{a}^p) = \prod_i \mathcal{D}(\mathbf{c}_{(1)}^i, \dots, \mathbf{c}_{(m)}^i, \mathbf{a}^i).$$

Furthermore

$$\begin{aligned} \mathcal{E} \left\{ \prod_i \prod_{\gamma} (\mathbf{c}_{(\gamma)}^i)^{\lambda_{i\gamma}/2} \right\} \\ = \prod_i \bar{D}(\lambda_{i1}, \dots, \lambda_{im}; n_1 - i + 1, \dots, n_m - i + 1, n - n' + (m - 1)(i - 1)), \end{aligned}$$

and

$$0 < c_{(\gamma)}^j < 1 - c_{(m)}^j - \dots - c_{(\gamma+1)}^j.$$

(b) If $n = n'$, the chance variables

$$\begin{aligned} \mathbf{c}_{(\gamma)}^1, & \quad \gamma = 1, \dots, m - 1, \\ \mathbf{c}_{(\gamma)}^j, & \quad j = 2, \dots, p; \gamma = 1, \dots, m, \end{aligned}$$

have a joint probability density, joint moments, and ranges which may be determined as above.

We shall call any statistic of the form

$$c_{i(\gamma)} = \frac{a_{i(\gamma)}}{a_i}$$

a generalized correlation ratio. Then, since

$$c_{j(\gamma)} = c_{(\gamma)}^1 \cdots c_{(\gamma)}^j,$$

we may obtain the distribution and moments of the generalized correlation ratio* and of the generalized Neyman-Pearson criteria.†

Applications of Corollary 3 occur in the theories of generalized period-

* See Wilks [34], p. 482.

† See Pearson and Wilks [29], and Wilks [34], p. 488. These criteria turn out to be functions of products of generalized correlation ratios.

gram analysis and relations between sets of variables. Of these, we first discuss generalized periodogram analysis.*

Let

$$z_{iv} = x_{iv} - \sum_{j=0}^k \alpha_{ij} \phi_j(t_v),$$

and let the functions $\phi_0(t), \dots, \phi_k(t)$ be such that

$$\sum_v \phi_i(t_v) \phi_j(t_v) = \delta_{ij}, \quad i, j = 0, \dots, k.$$

It is well known that if

$$\mathcal{D}(z_{1v}, \dots, z_{pv}) = N((z_v); (\sigma)), \quad \mathcal{D}(z_{11}, \dots, z_{pn}) = \prod_v \mathcal{D}(z_{1v}, \dots, z_{pv}),$$

and if

$$y_{i\mu} = \sum_v \phi_\mu(t_v) x_{iv}, \quad i = 1, \dots, p; \mu = 0, \dots, k,$$

then

$$\mathcal{D}(y_{1\mu} - \alpha_{1\mu}, \dots, y_{p\mu} - \alpha_{p\mu}) = N((y_\mu - \alpha_\mu); (\sigma)), \quad \mu = 0, \dots, k,$$

and

$$\mathcal{D}(y_{10} - \alpha_{10}, \dots, y_{pk} - \alpha_{pk}) = \prod_{\mu=0}^k \mathcal{D}(y_{1\mu} - \alpha_{1\mu}, \dots, y_{p\mu} - \alpha_{p\mu}).$$

Furthermore there exist linear functions

$$y_{i\mu} = \sum_v \phi_\mu(t_v) x_{iv}, \quad \mu = k+1, \dots, n-1,$$

such that

$$\begin{aligned} \mathcal{D}(y_{1\mu}, \dots, y_{p\mu}) &= N((y_\mu); (\sigma)), \\ \mathcal{D}(y_{1,k+1}, \dots, y_{p,n-1}) &= \sum_{\mu=k+1}^{n-1} \mathcal{D}(y_{1\mu}, \dots, y_{p\mu}), \end{aligned}$$

and†

$$\mathcal{D}(y_{10} - \alpha_{10}, \dots, y_{pn}) = \mathcal{D}(y_{10} - \alpha_{10}, \dots, y_{pk} - \alpha_{pk}) \cdot \mathcal{D}(y_{1,k+1}, \dots, y_{p,n-1}).$$

We consider our sample to contain an odd number‡ n , ($n = 2n' + 1$), of observations, and let

* See Fisher [14], and Greenstein [17], for discussions of the Schuster periodogram in one variable.

† This follows from Corollary 2 of Theorem 7.

‡ The adjustment to be made if n' is even will be obvious. See Fisher [14], p. 60.

$$\begin{aligned}\phi_0(t_\nu) &= n^{-1/2}, \\ \phi_{2j}(t_\nu) &= \left(\frac{2}{n}\right)^{1/2} \sin \frac{2\pi\nu j}{n}, \\ \phi_{2j-1}(t_\nu) &= \left(\frac{2}{n}\right)^{1/2} \cos \frac{2\pi\nu j}{n}, \quad j = 1, \dots, n'; \nu = 1, \dots, n.\end{aligned}$$

If $k < n-1$, it may be shown that a Neyman-Pearson λ criterion* for the hypothesis

$$\alpha_{ij} = 0, \quad i = 1, \dots, p; j = 1, \dots, k,$$

is given by a generalized correlation ratio

$$\lambda = \frac{\left| \sum_{\nu=k+1}^{n-1} y_{i\nu} y_{j\nu} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|}, \quad i, j = 1, \dots, p.$$

However, if $k = n-1$, or if it is desired that the hypothesis be rejected if and only if at least one of the statistics

$$(3.22) \quad \bar{c}_{p(2\gamma)} = \frac{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} - y_{i,2\gamma-1} y_{j,2\gamma-1} - y_{i,2\gamma} y_{j,2\gamma} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|},$$

$$\gamma = 1, \dots, n'; i, j = 1, \dots, p,$$

differs significantly from one, some other approach to the problem of testing the hypothesis that no period exists must be given.

Inasmuch as the joint distributions of statistics such as $\bar{c}_{p(2\gamma)}$ is quite complicated,† we shall extend Fisher's distribution‡ only to the case $p=2$, and shall use, in doing this, not the statistics (3.22) but the statistics

$$(3.23) \quad c_{2(2\gamma)} = \frac{\left| y_{i,2\gamma-1} y_{j,2\gamma-1} + y_{i,2\gamma} y_{j,2\gamma} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|}, \quad \gamma = 1, \dots, n'; i, j = 1, 2.$$

* See Neyman and Pearson [27], and Wilks [34], p. 485.

† They may be found by means of the distributions of Corollary 3 and variables having a normal multivariate distribution. Some notion of the complications may be gained by examining the distribution for $p=1$.

‡ R. A. Fisher [14], p. 57. The solution of a similar problem, if the variance and covariance parameters are known, is carried through in the same way.

The joint probability density of a subset $c_{2(2\gamma_1)}, c_{2(2\gamma_2)}, \dots, c_{2(2\gamma_k)}$ of the set (3.23), is, from (3.21),

$$\frac{\Gamma(n')\Gamma(n' - 1/2)}{\pi^{k/2}\Gamma(n' - k)\Gamma(n' - (k+1)/2)} \left\{ \prod_{\delta=1}^k c_{2(2\gamma_\delta)}^{-1/2} \right\} \\ \cdot \left\{ \int_{c_{2(2\gamma_k)}}^1 \cdots \int_{c_{2(2\gamma_1)}}^{1-c_{1(2\gamma_k)}-\cdots-c_{1(2\gamma_1)}} \left[\prod_{\delta=1}^k c_{1(2\gamma_\delta)}^{-1/2} \right] \right. \\ \cdot (1 - c_{1(2\gamma_1)} - \cdots - c_{1(2\gamma_k)})^{n'-k+1} \\ \cdot \left(1 - \frac{c_{2(2\gamma_1)}}{c_{1(2\gamma_1)}} - \cdots - \frac{c_{2(2\gamma_k)}}{c_{1(2\gamma_k)}} \right)^{n'-(k+1)/2-1} d c_{1(2\gamma_1)} \cdots d c_{1(2\gamma_k)} \left. \right\}$$

while the joint density of the entire set is

$$\frac{\Gamma(n')\Gamma(n' - 1/2)}{\pi^{k/2}\Gamma((k - \frac{1}{2})/2)} \left\{ \prod_{\gamma=1}^{n'} c_{2(2\gamma)}^{-1/2} \right\} \left\{ \int_{c_{2(2n'-2)}}^1 \cdots \int_{c_{2(2)}}^{1-c_{1(2n'-2)}-\cdots-c_{1(4)}} \left(\prod_{\gamma=1}^{n'-1} c_{1(2\gamma)}^{-1/2} \right) \right. \\ \cdot (1 - c_{1(1)} - \cdots - c_{1(2n'-2)})^{-1/2} \left(1 - \frac{c_{2(2n')}}{1 - c_{1(2)} - \cdots - c_{1(2n'-2)}} \right. \\ \left. \left. - \frac{c_{2(2n'-2)}}{c_{1(2n'-2)}} - \cdots - \frac{c_{2(2)}}{c_{1(2)}} \right)^{(n'-1)/2-1} d c_{1(2)} \cdots d c_{1(2n'-2)} \right\}.$$

If we denote the probability that

$$c_{2(2\gamma_i)} > \theta_{\gamma_i}, \quad i = 1, \dots, k,$$

for each selection of k integers from $1, 2, \dots, n'$, ($k=1, 2, \dots, n'$) by $P(\theta_{\gamma_1}, \dots, \theta_{\gamma_k})$, then the probability that at least j of the events

$$c_{2(2\gamma)} > \theta_{\gamma}, \quad \gamma = 1, \dots, n',$$

occur is*

$$(3.24) \quad P^{(j)} = \sum_{\nu=0}^{n'-j} (-1)^\nu c_\nu^{j+\nu-1} S_{j+\nu}$$

where

$$S_{j+\nu} = \sum P(\theta_{\gamma_1}, \dots, \theta_{\gamma_{j+\nu}})$$

and the summation is for all selections of $j+\nu$ integers from $1, \dots, n'$.†

From (3.24), we can find the probability that the maximum of the statistics $c_{2(2\gamma)}$ exceeds a given quantity g , by letting $j=1$ and $\theta_1 = \dots = \theta_{n'} = g$. It is, however, a question whether only the maximum need be considered or

* See, for example, [25], p. 162.

† If $\theta_1 = \theta_2 = \dots = \theta_{n'} = g$, then $S_{j+\nu} = c_{j+\nu}^{n'} P(\theta_1, \dots, \theta_{j+\nu})$.

whether each statistic should first be tested and then, if h are significant, the probability that the h greatest of the statistics should exceed a given quantity g should be found.

We note that by using the corollaries to Theorem 13, other distributions which may be used to test hypotheses in generalized periodogram analysis may be obtained for $p > 2$. However we shall not examine these here.

We now derive the joint distribution of q^2 and z^* and the joint distribution of the two sample canonical correlations,† which exist if $s=2$, when the two sets are not completely independent.

The degree of independence which we admit is expressed by the assumption that there exist s linear functions $y_{1\nu}, \dots, y_{s\nu}$ of $x_{1\nu}, \dots, x_{s\nu}$, and $p-s$ linear functions $y_{s+1,\nu}, \dots, y_{p\nu}$ of $x_{s+1,\nu}, \dots, x_{p\nu}$, such that the probability density of the chance variables

$$(3.25) \quad y_{i\nu}, \quad i = 1, \dots, p; \nu = 1, \dots, n,$$

is

$$(2\pi)^{pn/2} (1 - \rho^2)^{n/2} \exp \left[- \frac{1}{2(1 - \rho^2)} \sum_{\nu} \left\{ \sum_i y_{i\nu}^2 - 2\rho y_{1\nu} y_{s+1,\nu} \right\} \right].$$

We first note that if $\rho=0$, the joint probability density of

$$q^2 = \frac{W_{1s}}{a_s d} = \frac{a_s(2)}{a_s},$$

and

$$z = \frac{a_p}{a_s d} = \frac{a_s(1)}{a_s},$$

where‡

$$\begin{aligned} a_s(1) &= |a_{ij1}|, \\ a_s(2) &= |a_{ij2}|, \end{aligned} \quad i, j = 1, \dots, s,$$

may be obtained from Theorem 12, Corollary 3, since both q^2 and z are invariant under the transformation, which we have assumed, carrying the chance variables $x_{i\nu}$ into the chance variables (3.25), ($\rho=0$).

It follows from Theorem 9 and Corollary 3, that if we denote a_{112}/a_{11} by R^2 , then the chance variables

* Hotelling ([20], p. 333) defined q^2 and obtained on page 354 its distribution when one non-vanishing canonical correlation parameter exists. The distribution of z in that case may readily be obtained by the same procedure as that by which we shall obtain the joint distribution of q^2 and z .

† Hotelling ([20], p. 375) derived this distribution assuming $s=p-s=2$ and complete independence of sets.

‡ a_{ij1} and a_{ij2} are defined in (3.16).

$$R^2, c_{(1)}^i, c_{(2)}^i, \quad i = 1, \dots, s,$$

have density

$$(3.26) \quad \left\{ \prod_{j=1}^s \frac{\Gamma((n-j+1)/2)}{\Gamma((p-s-j+1)/2) \cdot \Gamma((n-p+s-j+1)/2) \cdot \Gamma((j-1)/2)} \right\} \\ \cdot (1 - \rho^2)^{n/2} (R^2)^{(p-s)/2-1} (1 - R^2)^{(n-p+s)/2-1} F(n/2, n/2, (p-s)/2, \rho^2 R^2) \\ \cdot \left\{ \prod_{i=2}^s (c_{(1)}^i)^{(n-p+s-i+1)/2-1} (c_{(2)}^i)^{(p-s-i+1)/2-1} (1 - c_{(1)}^i - c_{(2)}^i)^{(i-1)/2-1} \right\},$$

where $F(n/2, n/2, (p-s)/2, \rho^2 R^2)$ is the hypergeometric function.* If $s=2$, then $q^2 = R^2 \cdot c_{(2)}^2$, $z = (1 - R^2) c_{(1)}^2$, and

$$(3.27) \quad \mathcal{D}(q^2, z) = \frac{\Gamma(n)(1 - \rho^2)^{n/2}}{32\pi\Gamma(p-2)\Gamma(n-p-2)} z^{(n-p+1)/2-1} (q^2)^{(p-3)/2-1} \\ \cdot \int_0^1 F(n/2, n/2, (p-2)/2, \rho^2 [x(1-z-q^2) + q^2]) \\ \cdot \left[x(1-x) + \frac{q^2 z}{1-z-q^2} \right]^{-1} dx.$$

Then the general distribution may be similarly obtained as a multiple integral. The joint moments of q^2 and z are obtainable directly from (3.26).

Another statistic which may occasionally be of interest is

$$D = \frac{q^2}{z}$$

the distribution of which may be obtained as above. The factors of D are ratios of independent variances rather than correlation ratios if the sets are independently distributed.

The joint distribution of the two sample canonical correlation coefficients may be evaluated by means of the relations†

$$q^2 = r_1^2 r_2^2 \\ z = (1 - r_1^2)(1 - r_2^2).$$

Since $\partial(q^2, z)/\partial(r_1^2, r_2^2) = r_1^2 - r_2^2$, if $r_1^2 > r_2^2$, it follows that the chance variables r_1^2 and r_2^2 have a density which may be obtained by transforming (3.27).

* The details consist of letting $c = [a_{11}a_{s+1,s+1} - (a_{1,s+1})^2]/a_{s+1,s+1}$, noting that $a_{1,s+1} = a_{11}^{1/2} a_{s+1,s+1}^{1/2} \cdot (1 - c/a_{11})^{1/2}$, using Corollary 3 and Theorem 9, and then employing the same procedure as that adopted by R. A. Fisher [13].

† Hotelling [19], p. 335.

The first measure of relations between two sets of variables, both of which contained more than one variable, was the tetrad difference. However, as will be shown, the probability density of the tetrad difference depends on the "internal" correlation parameters even when the two sets are assumed to be independent. It is this fact which makes it necessary that q^2 be used instead of the tetrad difference unless the correlation parameters are known.

To discuss the statistic

$$(3.28) \quad T_{1s} = \frac{W_{1s}}{a_{11} a_{22} \cdots a_{pp}},$$

which is a generalization of the tetrad difference, we first note that if all the variables are independent, or if we consider the statistic

$$T'_{1s} = \frac{W'_{1s}}{g_{11} g_{22} \cdots g_{pp}}$$

in either of these two cases, the statistic is a product of independent determinants of correlation coefficients and has the distribution and moments of such a product.*

We now assume that $s = p - s = 2$ and derive the distribution of the square of the correlation tetrad difference under the assumption that the two variables of each set are correlated but that the two sets are independent.

Let the chance variables $x_{i\nu}$, ($i = 1, 2, 3, 4$; $\nu = 1, \cdots, n$), have probability density†

$$(2\pi)^{-2n} (1 - \rho_{12}^2)^{-n/2} (1 - \rho_{34}^2)^{-n/2} \cdot \exp \left[-\frac{1}{2(1 - \rho_{12}^2)} \left\{ \sum_{\nu} (x_{1\nu}^2 - 2\rho_{12}x_{1\nu}x_{2\nu} + x_{2\nu}^2) \right\} - \frac{1}{2(1 - \rho_{34}^2)} \left\{ \sum_{\nu} (x_{3\nu}^2 - 2\rho_{34}x_{3\nu}x_{4\nu} + x_{4\nu}^2) \right\} \right].$$

Now by (3.28),

$$T_{12} = \frac{a_{112}}{a_{11}} \cdot \frac{a_{(2)}}{a_{22}} \cdot (1 - r_{34}^2).$$

* If all the variables are independent, then the p -rowed determinant of correlation coefficients is the product of $p-1$ independent multiple correlation coefficients and has the corresponding distribution and moments. Wilks, [34], p. 492, first derived this distribution by the moment method. Certain more general distributions, which do not depend on the vanishing of all the correlation parameters may easily be derived.

† The assumption $\sigma_{ii} = 1$, ($i = 1, \cdots, 4$), involves no loss of generality since T_{1s} is invariant under the transformation $y_{i\nu} \sigma_{ii} = x_{i\nu}$.

Let

$$R^2 = \frac{a_{112}}{a_{11}}, \quad V_1 = \frac{a_{(2)}^2}{a_{22}}, \quad V_2 = (1 - r_{34}^2).$$

Then R^2 , V_1 , and V_2 are independent, and

$$\begin{aligned} \mathcal{D}(R^2, V_1, V_2) &= \frac{2[\Gamma(n/2)]^3(1 - \rho_{12}^2)^{n/2}(1 - \rho_{34}^2)^{n/2}}{\pi\Gamma((n-3)/2)\Gamma((n-2)/2)\Gamma((n-1)/2)} \cdot (1 - V_1)^{(n-4)/2} \\ &\cdot (R^2)^{1/2}(1 - R^2)^{(n-5)/2} F(n/2, n/2, 3/2, \rho_{12}^2 R^2) \\ &\cdot V_2^{(n-3)/2}(1 - V_2)^{1/2} F(n/2, n/2, 1/2, \rho_{34}^2 V_2). \end{aligned}$$

From this may be found at once the moments and distribution of T_{12} . The distribution and moments of T_{1s} may also be found.

Let $b_{ij\gamma} = \sum_{t=1}^{\gamma} a_{ijt}$, and let $b_{i(\gamma)}$, $b_{(\gamma)}^i$ be defined as are $a_{i(\gamma)}$ and $a_{(\gamma)}^i$. Let $b'_{ij\gamma}$, and so on, be similar functions of the variables y_{kv} . Assume that $b'_0 = 0$, and let

$$u_{i(\gamma)} = b_{(\gamma)}^i - b_{(\gamma-1)}^i.$$

We shall assume that $n = n_1 + \dots + n_m$. This involves no loss of generality, for if $n > n_1 + \dots + n_m$, we may extend the system $b'_{ij\gamma}$ to include b'_{ijm+1} .

THEOREM 13. Let $\mathcal{D}(x_{1v}, \dots, x_{pv}) = N((x_v); (\sigma))$, and let $\mathcal{D}(x_{11}, \dots, x_{pn}) = \prod \mathcal{D}(x_{1v}, \dots, x_{pv})$. Then

$$\mathcal{D}(u_{i(1)}) = G(u_{i(1)}; n_1 - i + 1, \sigma^i),$$

$$\mathcal{D}(u_{i(\gamma)}) = G(u_{i(\gamma)}; n_{\gamma}, \sigma^i), \quad \gamma = 2, \dots, m,$$

$$\mathcal{D}(w_{ig}^2) = G(w_{ig}^2; 1, \sigma),$$

and

$$\mathcal{D}(u, w^2) = \prod_i \left[\prod_{\gamma} \mathcal{D}(u_{i(\gamma)}) \right] \left[\prod_g \mathcal{D}(w_{ig}^2) \right].$$

Proof. See Theorem 11.

COROLLARY 1.

$$(3.29) \quad \mathcal{D}(b_{(1)}^i, \dots, b_{(m)}^i) = G(b_{(1)}^i; n_1 - j + 1, \sigma^i) \prod_{\gamma=2}^m G(b_{(\gamma)}^i - b_{(\gamma-1)}^i; n_{\gamma}, \sigma^i)$$

and

$$\mathcal{D}(b_{(1)}^1, \dots, b_{(m)}^p) = \prod_j \mathcal{D}(b_{(1)}^1, \dots, b_{(m)}^j).$$

Furthermore

$$(3.30) \quad \mathcal{E} \left\{ \prod_i \prod_\gamma (b_{(\gamma)}^i)^{\lambda_{j\gamma}/2} \right\} \\ = \prod_i \prod_\gamma \bar{G} \left(\lambda_{i\gamma}; \sum_{\delta=1}^{\gamma} (n_\delta + \lambda_{j\delta}) - j + 1 - \lambda_{i\gamma}, \sigma^j \right)$$

and

$$0 < b_{(1)}^j < b_{(2)}^j < \cdots < b_{(m)}^j < \infty.$$

The proof is omitted.

Let $d_{(\gamma)}^j = b_{(\gamma)}^j / b_{(m)}^j$. The probability density and joint moments of the chance variables $d_{(\gamma)}^j$ may be obtained by transforming (3.29) and (3.30). Furthermore

$$0 < d_{(1)}^j < \cdots < d_{(m-1)}^j < 1.$$

Let $\omega_{(\gamma)}^j = b_{(\gamma)}^j / b_{(\gamma+1)}^j$. We can then state the following corollary:

COROLLARY 2.

$$\mathcal{D}(\omega_{(\gamma)}^j) = D(\omega_{(\gamma)}^j; n_1 + \cdots + n_\gamma - j + 1, n_{\gamma+1}),$$

$$\mathcal{D}(\omega_{(1)}^1, \cdots, \omega_{(m-1)}^p) = \prod_i \prod_{\gamma=1}^{m-1} \mathcal{D}(\omega_{(\gamma)}^i),$$

and

$$\mathcal{D}(\omega_{(1)}^1, \cdots, \omega_{(m-1)}^p, \mathbf{a}^1, \cdots, \mathbf{a}^p) = \mathcal{D}(\omega_{(1)}^1, \cdots, \omega_{(m-1)}^p) \cdot \mathcal{D}(\mathbf{a}^1, \cdots, \mathbf{a}^p).$$

The proof is omitted. The moments and possible ranges of the chance variables $\omega_{(\gamma)}^j$ are too well known to require statement here.

We are now able to derive several additional distributions of the generalized analysis of variance.

To do so, we need only note that

$$b_{i(\gamma)} = \prod_{i=1}^j b_{(\gamma)}^i,$$

and that the probability density and moments of the statistics $b_{i(\gamma)}$ may therefore be obtained from Theorem 13, Corollary 1.

The probability density and moments of the statistics $d_{i(\gamma)}$, where

$$d_{i(\gamma)} = \frac{b_{i(\gamma)}}{b_{j(m)}},$$

may be similarly derived.

It may be remarked that the independence of the chance variables $\omega_{(\gamma)}^j$

permits the distributions depending on them to be expressed in terms of previously obtained distributions.

Distributions derived from Theorem 13 may be used in testing various hypotheses in generalized analysis of variance, comparative analysis,* and relations between sets of variables. The method to be used should be sufficiently clear to make any further transformations here unnecessary.

BIBLIOGRAPHY

1. M. S. Bartlett, *On the theory of statistical regression*, Proceedings of the Royal Society of Edinburgh, vol. 53 (1933), pp. 260-283.
2. M. S. Bartlett, *The problem in statistics of testing several variances*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 164-169.
3. M. S. Bartlett, *The vector representation of a sample*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 327-340.
4. M. S. Bartlett and J. Wishart, *The generalized product moment distribution in a normal system*, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 260-270.
5. S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig, 1932.
6. W. G. Cochran, *The distribution of quadratic forms in a normal system, with applications to the analysis of covariance*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 178-191.
7. H. Cramer, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 36, London, 1937.
8. G. Darmon, *Sur les lois de probabilité à estimation exhaustive*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 200 (1935), pp. 1265-1266.
9. R. A. Fisher, *On the mathematical foundations of theoretical statistics*, Philosophical Transactions of the Royal Society of London, (A), vol. 222 (1921), pp. 309-368.
10. R. A. Fisher, *The influence of rainfall on the yield of wheat at Rothamstead*, Philosophical Transactions of the Royal Society of London, (B), vol. 213 (1924), pp. 89-142.
11. R. A. Fisher, *On a distribution yielding the error functions of several well known statistics*, Proceedings of the International Mathematical Congress, Toronto, 1924, pp. 805-812.
12. R. A. Fisher, *Applications of "Student's" distribution*, Metron, vol. 5 (1926), pp. 90-104.
13. R. A. Fisher, *The general sampling distribution of the multiple correlation coefficient*, Proceedings of the Royal Society of London, (A), vol. 121 (1928), pp. 654-673.
14. R. A. Fisher, *Tests of significance in harmonic analysis*, Proceedings of the Royal Society of London, (A), vol. 125 (1929), pp. 54-60.
15. R. A. Fisher, *Two new properties of mathematical likelihood*, Proceedings of the Royal Society of London, (A), vol. 144 (1934), pp. 285-307.
16. R. A. Fisher, *Statistical Methods for Research Workers*, 6th edition. Oliver and Boyd, London, 1936.
17. B. Greenstein, *Periodogram analysis with special application to business failures in the United States, 1867-1932*, Econometrica, vol. 3 (1935), pp. 170-198.
18. H. Hotelling, *The generalization of Student's ratio*, Annals of Mathematical Statistics, vol. 2 (1931), pp. 170-198.
19. H. Hotelling, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology, vol. 24 (1933), pp. 417-441, 498-520.

* We use comparative analysis to denote the type of statistical analysis of which periodogram analysis is an example, that is, statistical analysis in which the effects of certain types of selection must be excluded.

20. H. Hotelling, *Relations between two sets of variables*, Biometrika, vol. 28 (1936), pp. 321-377.
21. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.
22. S. Kolodziejczyk, *On an important class of statistical hypotheses*, Biometrika, vol. 27 (1935), pp. 161-190.
23. B. O. Koopman, *On distributions admitting a sufficient statistic*, these Transactions, vol. 39 (1936), pp. 399-410.
24. S. Kullback, *An application of characteristic functions to the distribution problem of statistics*, Annals of Mathematical Statistics, vol. 5 (1934), pp. 264-307.
25. W. G. Madow, *Contributions to the theory of comparative statistical analysis. I Fundamental theorems of comparative analysis*, Annals of Mathematical Statistics, vol. 8 (1937), pp. 159-176.
26. J. Neyman, *Su un teorema concernente le cosidette statistiche sufficienti*, Giornale dell'Istituto Italiano degli Attuari, vol. 6 (1935), pp. 320-334.
27. J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika, vol. 20^A (1928), pp. 175-240, 264-294.
28. J. Neyman and E. S. Pearson, *Sufficient Statistics and Uniformly Most Powerful Tests*, Statistical Research Memoirs, Cambridge University Press, London, vol. 1, 1936.
29. E. S. Pearson and S. S. Wilks, *Methods of statistical analysis appropriate for k samples of two variables*, Biometrika, vol. 25 (1933), pp. 363-377.
30. E. S. Pearson and J. Neyman, see [27] and [28].
31. S. Saks, *Theory of the Integral*, 2d edition, Stechert, New York, 1937.
32. H. W. Turnbull, *The Theory of Determinants, Matrices, and Invariants*, Blackie and Son, London, 1926.
33. B. L. Welch, *Some problems in the analysis of regression among k samples of two variables*, Biometrika, vol. 27 (1935), pp. 145-160.
34. S. S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika, vol. 24 (1932), pp. 472-494.
35. S. S. Wilks, *The standard error of a tetrad in samples from a normal population of independent variables*, Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 562-565.
36. S. S. Wilks and E. S. Pearson, see [29].
37. S. S. Wilks, *Moment-generating operators for determinants of product moments in samples from a normal system*, Annals of Mathematics, (2), vol. 35 (1933), pp. 312-340.
38. S. S. Wilks, *On the independence of k sets of normally distributed statistical variables*, Econometrika, vol. 3 (1935), pp. 309-326.
39. J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika, vol. 20^A (1928), pp. 32-52.
40. J. Wishart, *Sampling errors in the theory of two factors*, British Journal of Psychology, vol. 19 (1928), pp. 32-52.
41. J. Wishart and M. S. Bartlett, see [4].

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.