RANDOM ORTHOGONAL TRANSFORMATIONS AND THEIR USE IN SOME CLASSICAL DISTRIBUTION PROBLEMS IN MULTIVARIATE ANALYSIS[1]

By Robert A. Wijsman

Department of Statistics, University of California, Berkeley

0. Summary. Orthogonal matrices having elements depending on certain random vectors provide a useful tool in various distribution problems in multivariate analysis. The method is applied to the derivation of the distributions of Hotelling's T^2 and Wilks' generalized variance, the Bartlett decomposition, and the Wishart distribution.

1. Introduction. The purpose of this paper is to demonstrate a method for treating some distribution problems in multivariate normal analysis, and to apply this method to the derivation of the Wishart distribution, the Bartlett decomposition, and the distributions of Hotelling's T^2 and Wilks' generalized variance. A large number of different derivations of these statistics exist in the literature ([1], [2], [4], [6] to [23]), and which one is preferable is a matter of taste. The motivation for presenting yet another derivation of well-known results is that it is believed that the method presented here leads to the results faster than existing derivations, without the necessity of extensive preparation, and almost without computations. A further advantage of the method is that it leads immediately to a representation of the statistics mentioned in terms of combinations of independent normal variables. More specifically, apart from constant factors, Hotelling's T^2 is obtained as an F variable, Wilks' generalized variance as a product of independent χ^2 variables, while the Wishart distribution is simply related to the joint distribution of independent normal and χ^2 variables (Bartlett decomposition [3]). These are known facts, stated explicitly by some authors ([2], [6], [16], [18]), but clearly demonstrated by only few derivations in the literature: Elfving [6] and Ogawa [16] obtain the Bartlett decomposition; Anderson [2] and Elfving [6] obtain the generalized variance essentially as a product of χ^2 variables; while Anderson [2] obtains T^2 essentially as an F variable in a rather indirect way, by relating it to a multiple correlation coefficient. The method presented in this paper will lead to the results in a simple direct, and unified way.

It can be expected that orthogonal transformations provide at least as powerful a tool in the multivariate case as in the univariate case. Indeed, an example can be found in the work of James [12]. The method followed in the present paper will also lean very heavily on orthogonal transformations of random variables. In order to utilize this tool to the utmost, most of the useful transforma-

Received June 18, 1956.

[1] This investigation was supported (in part) by a research grant (RG-3666) from the Institutes of Health, U. S. Public Health Service.

tions will be performed with orthogonal matrices, the elements of which depend on a random vector. This idea is not new, but is often couched in geometrical language.[2] In this respect the treatment in this paper will have something in common with that of Elfving [6] and Ogawa [16]. However, the method followed here does not seem to have appeared in the literature in the same form.

2. Methods and main results. *Notation*: Boldface symbols denote matrices and column vectors, prime denotes transposition, $\mathbf{0}$ is a zero vector, $\mathbf{I}_n$ an $n \times n$ identity matrix, $\mathbf{\Omega}$ an orthogonal matrix. If $\mathbf{A}$ is a square matrix, then $|\mathbf{A}|$ denotes the absolute value of its determinant, and tr $\mathbf{A}$ its trace. $\mathfrak{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the distribution of a normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, $\mathfrak{N}(0, 1)$ the distribution of a normal variable with zero mean and unit variance. A χ^2 variable with n degrees of freedom is denoted by χ_n^2, an F variable with n_1 and n_2 degrees of freedom by F_{n_1, n_2}. Of a $k \times n$ matrix of variables u_{ir}, the i-th row is denoted by $\mathbf{U}_i'$, the r-th column by $\mathbf{U}_{(r)}$.

The dominant method used throughout this paper is transformation by an orthogonal matrix, the elements of which depend on a random vector. The usefulness of this method depends on the following lemma.

LEMMA 1. *Let* $\mathbf{X}$ *be a random vector with components* $x_1 \cdots x_n$, *and let* $\mathbf{\Omega}(\mathbf{Z})$ *be a random* $n \times n$ *orthogonal matrix whose elements depend in a measurable way on a random vector* $\mathbf{Z}$ *which is independent of* $\mathbf{X}$. *Let* $\mathbf{Y} = \mathbf{\Omega X}$; *then if* $\mathbf{X}$ *is* $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{Y}$ *is also* $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$ *and independent of* $\mathbf{Z}$.

The proof of Lemma 1 follows immediately from the fact that the conditional distribution of $\mathbf{Y}$, given $\mathbf{Z}$, is $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$ and is therefore independent of $\mathbf{Z}$.

The lemma will usually be applied in cases where $\mathbf{X}$ and $\mathbf{Z}$ have the same number of components, and $\mathbf{\Omega}(\mathbf{Z})$ is defined in such a way that $\mathbf{\Omega Z}$ has all but its last component equal to zero. In Appendix 1 it will be shown that $\mathbf{\Omega}$ can be uniquely defined in a measurable way.

Throughout this paper we have to consider random matrices, the elements of which are independent $\mathfrak{N}(0, 1)$ variables. A $k \times n$ matrix of independent $\mathfrak{N}(0, 1)$ variables x_{ir} $(i = 1 \cdots k, r = 1 \cdots n)$, will be denoted by $\mathbf{M}_{kn}^x$. We shall assume $k \leqq n$. The i-th row of $\mathbf{M}_{kn}^x$ will be denoted by $\mathbf{X}_i'$ $(i = 1 \cdots k)$. With $\mathbf{M}_{kn}^x$ we form the symmetric matrix $\mathbf{A}_{kn}^x$

$$\mathbf{A}_{kn}^x = \mathbf{M}_{kn}^x (\mathbf{M}_{kn}^x)', \tag{1}$$

whose ij-th element is $\mathbf{X}_i'\mathbf{X}_j$ $(i, j = 1 \cdots k)$.

Consider the transformation

$$\mathbf{X}_i'\mathbf{\Omega} = \mathbf{Z}_i' \qquad (i = 1 \cdots k),$$

in which the orthogonal matrix $\mathbf{\Omega}$ depends on $\mathbf{X}_1$ in such a way as to reduce the first $n - 1$ components of $\mathbf{Z}_1$ to 0. The last component of any $\mathbf{Z}_i$—that is, z_{in}—

[2] In a course at Stanford University, Dr. Charles M. Stein uses this idea in the derivation of Hotelling's T^2 distribution (private communication).

will, in the following, be denoted for short by z_i $(i = 1 \cdots k)$. For z_1 we have

$$z_1^2 = \mathbf{X}_1'\mathbf{X}_1, \tag{2}$$

and z_1^2 is clearly a χ_n^2 variable. Inserting the identity matrix $\mathbf{I}_n = \mathbf{\Omega\Omega}'$ between the two factors on the right-hand side of (1) we obtain

$$\mathbf{A}_{kn}^x = \mathbf{M}_{kn}^x \mathbf{\Omega\Omega}'(\mathbf{M}_{kn}^x)' = \begin{Vmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_k \\ z_1 z_2 & \mathbf{Z}_2'\mathbf{Z}_2 & \cdots & \mathbf{Z}_2'\mathbf{Z}_k \\ \vdots & \vdots & & \vdots \\ z_1 z_k & \mathbf{Z}_k'\mathbf{Z}_2 & \cdots & \mathbf{Z}_k'\mathbf{Z}_k \end{Vmatrix}. \tag{3}$$

Let $\mathbf{Y}_i$ $(i = 1 \cdots k - 1)$ be the $(n - 1)$-component vector obtained from $\mathbf{Z}_{i+1}$ by deleting its last component z_{i+1}:

$$y_{ir} = z_{i+1,r} \qquad (i = 1 \cdots k - 1, r = 1 \cdots n - 1),$$

so that we have

$$\mathbf{X}_{i+1}'\mathbf{X}_{j+1} = \mathbf{Z}_{i+1}'\mathbf{Z}_{j+1} = \mathbf{Y}_i'\mathbf{Y}_j + z_{i+1}z_{j+1} \qquad (i, j = 1 \cdots k - 1). \tag{4}$$

It is now possible to write (3) in the following way:

$$\mathbf{A}_{kn}^x = \begin{Vmatrix} z_1 & & & & \\ z_2 & 1 & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ z_k & & & & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 & \cdots & 0 \\ 0 & \mathbf{Y}_1'\mathbf{Y}_1 & \cdots & \mathbf{Y}_1'\mathbf{Y}_{k-1} \\ \vdots & \vdots & & \vdots \\ 0 & \mathbf{Y}_{k-1}'\mathbf{Y}_1 & \cdots & \mathbf{Y}_{k-1}'\mathbf{Y}_{k-1} \end{Vmatrix} \begin{Vmatrix} z_1 & z_2 & \cdots & z_k \\ & 1 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \cdot \\ & & & & 1 \end{Vmatrix}, \tag{5}$$

in which z_1 is χ_n, $z_2 \cdots z_k$ and the y_{ir} $(i = 1 \cdots k - 1, r = 1 \cdots n - 1)$ are $\mathfrak{N}(0, 1)$, and all variables are independent. Equation (5) can be written more concisely if we denote by $\mathbf{Z}$ the vector with components $z_2 \cdots z_k$:

$$\mathbf{A}_{kn}^x = \begin{Vmatrix} z_1 & \mathbf{0}' \\ \mathbf{Z} & \mathbf{I}_{k-1} \end{Vmatrix} \begin{Vmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{A}_{k-1,n-1}^y \end{Vmatrix} \begin{Vmatrix} z_1 & \mathbf{Z}' \\ \mathbf{0} & \mathbf{I}_{k-1} \end{Vmatrix}. \tag{6}$$

Almost everything will follow from (5) or (6), which is essentially the first step in the Bartlett decomposition. Taking determinants in (6) we get

$$| \mathbf{A}_{kn}^x | = z_1^2 \, | \mathbf{A}_{k-1,n-1}^y |, \tag{7}$$

in which the two factors on the right-hand side are independent, and z_1^2 has a χ_n^2 distribution. Upon repeated application of (7) we get the following result:

LEMMA 2. *The distribution of* $| \mathbf{A}_{kn}^x |$ *is the distribution of the product of* k *independent* χ^2 *variables with* $n, n - 1, \cdots, n - k + 1$ *degrees of freedom, respectively.*

Equation (7) holds for any k. Writing (7) with k replaced by $k - h$ $(h \leqq k - 1)$ and forming ratios, we have

$$(8) \qquad \frac{|\mathbf{A}^x_{kn}|}{|\mathbf{A}^x_{k-h,n}|} = \frac{|\mathbf{A}^y_{k-1,n-1}|}{|\mathbf{A}^y_{k-h-1,n-1}|},$$

in which we set by convention $\mathbf{A}^x_{on} = 1$. Upon iteration of (8) and use of Lemma 2 we get the following result.

LEMMA 3. *The ratio* $|\mathbf{A}^x_{kn}|\,|\mathbf{A}^x_{k-h,n}|^{-1}$, *for* $h \leqq k - 1$, *is distributed like the product of* h *independent* χ^2 *variables with* $n - k + h$, $n - k + h - 1, \cdots,$ $n - k + 1$ *degrees of freedom, respectively.*

COROLLARY. *The ratio* $|\mathbf{A}^x_{kn}|\,|\mathbf{A}^x_{k-1,n}|^{-1}$ *is a* χ^2_{n-k+1} *variable.*

The results so far obtained are sufficient to derive the distributions of T^2 and the generalized variance. For the Wishart distribution, however, it is necessary to consider the joint distribution of the $\frac{1}{2}k(k + 1)$ distinct elements $a_{ij} = \mathbf{X}'_i\mathbf{X}_j$ $(i = 1 \cdots k, j = i \cdots k)$ of $\mathbf{A}^x_{kn}$. The decomposition (5) expresses the a_{ij} as functions of the χ^2_n variable z_1^2, the $k - 1$ $\mathfrak{N}(0, 1)$ variables $z_2 \cdots z_k$, and the $\frac{1}{2}k(k - 1)$ variables $b_{ij} = \mathbf{Y}'_i\mathbf{Y}_j$ $(i = 1 \cdots k - 1, j = i \cdots k - 1)$. If the decomposition (5) is continued, then the a_{ij} are expressed as functions of $\frac{1}{2}k(k - 1)$ $\mathfrak{N}(0, 1)$ variables and k χ^2 variables with $n, n - 1, \cdots, n - k + 1$ degrees of freedom, respectively, all variables being independent.[3] The joint distribution of these variables, together with the Jacobian of the transformation, will produce the joint distribution of the a_{ij}.

According to (3), the a_{ij} are first expressed as functions of new variables z_1^2, z_i, and $\mathbf{Z}'_i\mathbf{Z}_j$ $(i = 2 \cdots k, j = i \cdots k)$. Subsequently a new set of variables $b_{ij} = \mathbf{Y}'_i\mathbf{Y}_j$ $(i = 1 \cdots k - 1, j = i \cdots k - 1)$ is introduced, connected with the $\mathbf{Z}'_i\mathbf{Z}_j$ through (4). The first transformation yields a Jacobian z_1^{k-1}, the second yields unity. Hence, the Jacobian of the transformation from the $\frac{1}{2}k(k + 1)$ variables a_{ij} $(i = 1 \cdots k, j = i \cdots k)$ to the $\frac{1}{2}k(k + 1)$ variables z_1^2, $z_2 \cdots z_k$, and b_{ij} $(i = 1 \cdots k - 1, j = i \cdots k - 1)$ is z_1^{k-1}. Let the density of the a_{ij} $(i = 1 \cdots k, j = i \cdots k)$ be denoted by $p(\mathbf{A}^x_{kn})$, and the density of the b_{ij} $(i = 1 \cdots k - 1, j = i \cdots k - 1)$ by $p(\mathbf{A}^y_{k-1,n-1})$. The joint density of the $\mathfrak{N}(0, 1)$ variables $z_2 \cdots z_k$, the χ^2_n variable z_1^2, and the variables b_{ij} is given by

$$(2\pi)^{-(1/2)(k-1)} \exp\left[-\tfrac{1}{2}\sum_{i=2}^{k} z_i^2\right] 2^{-(1/2)-n/2}\Gamma^{-1}\left(\frac{n}{2}\right) z_1^{n-2} \exp\left[-\tfrac{1}{2}z_1^2\right] p(\mathbf{A}^y_{k-1,n-1}).$$

Taking the Jacobian z_1^{k-1} into account, we have:

$$(9) \qquad p(\mathbf{A}^x_{kn}) = c_{kn} z_1^{n-k-1} \exp\left[-\tfrac{1}{2}\left(z_1^2 + \sum_{i=2}^{k} z_i^2\right)\right] p(\mathbf{A}^y_{k-1,n-1}),$$

[3] If the decomposition (5) is continued, the right hand side can be written as the product of a triangular matrix and its transposed, the elements in the triangular matrix being independent normal and χ^2 variables. The decomposition in that form was also obtained by Mauldon [15].

with

$$c_{kn}^{-1} = (2\pi)^{(k-1)/2} 2^{n/2} \Gamma\left(\frac{n}{2}\right). \tag{10}$$

In order to write (9) so that it can be iterated immediately, we observe first that $z_1^2 = | \mathbf{A}_{kn}^x | \, | \mathbf{A}_{k-1,n-1}^y |^{-1}$ by (7). Furthermore,

$$z_1^2 + \sum_{i=2}^{k} z_i^2 = \sum_{i=1}^{k} \mathbf{X}_i' \mathbf{X}_i - \sum_{i=1}^{k-1} \mathbf{Y}_i' \mathbf{Y}_i = \operatorname{tr} \mathbf{A}_{kn}^x - \operatorname{tr} \mathbf{A}_{k-1,n-1}^y,$$

using (2) and (4). Thus we can write (9) in the following way:

$$p(\mathbf{A}_{kn}^x) \, | \mathbf{A}_{kn}^x |^{-(n-k-1)/2} \exp [\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x] = c_{kn} p(\mathbf{A}_{k-1,n-1}^y) \, | \mathbf{A}_{k-1,n-1}^y |^{-(n-k-1)/2} \exp [\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{k-1,n-1}^y],$$

from which follows immediately by iteration

$$p(\mathbf{A}_{kn}^x) \, | \mathbf{A}_{kn}^x |^{-(n-k-1)/2} \exp [\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x] = \prod_{i=0}^{k-1} c_{k-i,n-i} = C_{kn},$$

with C_{kn} given by

$$C_{kn}^{-1} = 2^{(1/2)(kn)} \pi^{(1/4)k(k-1)} \prod_{i=0}^{k-1} \Gamma\left(\frac{n-i}{2}\right). \tag{11}$$

We have then, finally,

$$p(\mathbf{A}_{kn}^x) = C_{kn} \, | \mathbf{A}_{kn}^x |^{(n-k-1)/2} \exp[-\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x], \tag{12}$$

with C_{kn} given by (11).

3. Applications. Let $\mathbf{U}_{(1)} \cdots \mathbf{U}_{(n)}$ be n independent observations on a k-component random vector $\mathbf{U}$, which is $\mathfrak{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $k \leq n - 1$. The components of $\mathbf{U}_{(r)}$ will be denoted by u_{ir} ($i = 1 \cdots k$, $r = 1 \cdots n$). The sample mean is $\bar{\mathbf{U}} = (1/n) \sum_{r=1}^{n} \mathbf{U}_{(r)}$, having components $\bar{u}_1 \cdots \bar{u}_k$. The sample covariance matrix $\mathbf{S}$ has components s_{ij} ($i, j = 1 \cdots k$) given by

$$s_{ij} = \frac{1}{n-1} \sum_{r=1}^{n} (u_{ir} - \bar{u}_i)(u_{jr} - \bar{u}_j). \tag{13}$$

(a) *Hotelling's* T^2. Hotelling's T^2 is defined as

$$T^2 = n(\bar{\mathbf{U}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{U}} - \boldsymbol{\mu}_0), \tag{14}$$

in which $\boldsymbol{\mu}_0$ is some specified vector. T^2 is not defined on the null set in the sample space on which $\mathbf{S}$ is singular.

First consider the case $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. By making the proper transformations it can be shown[4] that $[1/(n-1)]T^2$ has the same distribution as T_1^2 defined by

[4] See, for example [8]. In order to make this paper self-contained, a proof is given in Appendix 2.

$$T_1^2 = \mathbf{X}'_{(n)}(\mathbf{A}^x_{k,n-1})^{-1}\mathbf{X}_{(n)}\,, \tag{15}$$

in which $\mathbf{X}_{(1)} \cdots \mathbf{X}_{(n)}$ are independent, $\mathfrak{N}(\mathbf{0}, \mathbf{I}_k)$, and $\mathbf{A}^x_{k,n-1}$ is defined by (1). In (15) all the variables are independent, and $\mathfrak{N}(0, 1)$. By subjecting the $\mathbf{X}_{(r)}$ to an orthogonal transformation with matrix $\mathbf{\Omega}: \mathbf{Y}_{(r)} = \mathbf{\Omega X}_{(r)}$ $(r = 1 \cdots n)$, we can write (15) also as

$$T_1^2 = \mathbf{Y}'_{(n)}(\mathbf{A}^y_{k,n-1})^{-1}\mathbf{Y}_{(n)}\,, \tag{16}$$

with $\mathbf{A}^y_{k,n-1} = \mathbf{M}^y_{k,n-1}(\mathbf{M}^y_{k,n-1})'$ and $\mathbf{M}^y_{k,n-1} = \mathbf{\Omega M}^x_{k,n-1}$. The columns of $\mathbf{M}^x_{k,n-1}$ are $\mathbf{X}_{(1)} \cdots \mathbf{X}_{(n-1)}$. Hence, if $\mathbf{\Omega}$ depends only on $\mathbf{X}_{(n)}$, then, by Lemma 1, the elements of $\mathbf{M}^y_{k,n-1}$ are still independent, $\mathfrak{N}(0, 1)$, and independent of $\mathbf{Y}_{(n)}$. We now choose $\mathbf{\Omega}$ such that the first $k - 1$ components of $\mathbf{Y}_{(n)}$ are 0. The k-th component of $\mathbf{Y}_{(n)}$ will be denoted by y. From (16) it follows that T_1^2 equals the product of y^2 and the kk-th element of $(\mathbf{A}^y_{k,n-1})^{-1}$, where it has to be remembered that these factors are independent. Now $y^2 = \mathbf{X}'_{(n)}\mathbf{X}_{(n)}$ is a χ^2_k variable, and the kk-th element of $(\mathbf{A}^y_{k,n-1})^{-1}$ equals $|\mathbf{A}^y_{k-1,n-1}|\,|\mathbf{A}^y_{k,n-1}|^{-1}$, the reciprocal of which is a χ^2_{n-k} variable by the corollary to Lemma 3. Hence T_1^2 is the ratio of two independent χ^2 variables, with k and $(n - k)$ degrees of freedom, respectively. It follows that $(n - 1)^{-1}k^{-1}(n - k)T^2$ is an $F_{k,n-k}$ variable.

If $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, then, in (15), $\mathbf{X}_{(n)}$ no longer has zero mean, with the consequence that y^2 is a noncentral χ^2_k variable. On the other hand, the distribution of $\mathbf{A}^x_{k,n-1}$ is unchanged. It follows then that $(n - 1)^{-1}k^{-1}(n - k)T^2$ is a noncentral $F_{k,n-k}$ variable. Its distribution was first derived by Hsu [9].

(b) *Wilks' generalized variance.* Wilks' generalized variance is defined as $|\mathbf{S}|$, the determinant of the sample covariance matrix given by (13). By making the same transformation which led to (15) (see also Appendix 2), we find that

$$(n - 1)\,\mathbf{CSC}' = \mathbf{A}^x_{k,n-1}\,, \tag{17}$$

in which $\mathbf{C}$ is a nonsingular matrix transforming $\mathbf{\Sigma}$ to the identity matrix

$$\mathbf{C\Sigma C}' = \mathbf{I}_k\,. \tag{18}$$

Taking determinants in (17) and (18) and using Lemma 2, we have then immediately the result that $(n - 1)\,|\mathbf{\Sigma}|^{-1}\,|\mathbf{S}|$ is distributed like the product of k independent χ^2 variables with $n - 1, \cdots, n - k$ degrees of freedom, respectively. The density of this distribution can be obtained easily only for $k = 1$ and $k = 2$. For $k \geqq 3$, expressions in terms of infinite series have been given by Kullback [13].

(c) *The Wishart distribution and Bartlett decomposition.* The Wishart distribution is the joint distribution of the $\frac{1}{2}k(k + 1)$ distinct elements of the sample covariance matrix $\mathbf{S}$, given by (13). It is more convenient to study $\mathbf{S}_1 = (n - 1)\mathbf{S}$. By (17) we have

$$\mathbf{CS}_1\mathbf{C}' = \mathbf{A}^x_{k,n-1}\,. \tag{19}$$

The linear transformation (19) relates the sample covariance matrix to $\mathbf{A}^x_{k,n-1}$, defined by (1). The decomposition (5) or (6) of $\mathbf{A}^x_{k,n-1}$ is essentially the first step in the Bartlett decomposition, giving rise to a χ^2_{n-1} variable z_1^2 and $k - 1$ $\mathfrak{N}(0, 1)$ variables $z_2 \cdots z_k$. If the decomposition is continued, then $\mathbf{A}^z_{k,n-1}$, and therefore by (19) also $\mathbf{S}_1$, is related in a simple way to k χ^2 variables and $\frac{1}{2} k(k - 1)$ $\mathfrak{N}(0, 1)$ variables, all independent, which provides the complete Bartlett decomposition.

The density of the $\frac{1}{2} k(k + 1)$ distinct elements of $\mathbf{A}^z_{k,n-1}$ is given by (12), after replacing n by $n - 1$. The Jacobian of the transformation (19) is[5]

$$\frac{\partial(\mathbf{A}^z_{k,n-1})}{\partial(\mathbf{S}_1)} = |\mathbf{C}|^{k+1}, \tag{20}$$

which we find equals $|\boldsymbol{\Sigma}|^{-\frac{1}{2}(k+1)}$, using (18). Furthermore, by (18) and (19) we have $|\mathbf{A}^z_{k,n-1}| = |\boldsymbol{\Sigma}|^{-1} |\mathbf{S}_1|$ and $\operatorname{tr} \mathbf{A}^z_{k,n-1} = \operatorname{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S}_1$. Substitution of these expressions into (12) gives the Wishart distribution

$$p(\mathbf{S}_1) = C_{k,n-1} |\boldsymbol{\Sigma}|^{-(n-1)/2} |\mathbf{S}_1|^{(n-k-2)/2} \exp[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\mathbf{S}_1],$$

in which it has to be remembered that $\mathbf{S}_1 = (n - 1)\mathbf{S}$, and $C_{k,n-1}$ is given by (11).

4. Acknowledgments. The writer wishes to thank Dr. Henry Scheffé for bringing the problem to his attention and for helpful suggestions, and Dr. Charles M. Stein for providing some valuable references.

APPENDICES

Appendix 1. Let $\boldsymbol{\Omega}^{(m)} = \boldsymbol{\Omega}^{(m)}(\mathbf{Z})$ be an $n \times n$ orthogonal matrix $(1 \leqq m \leqq n - 1)$ depending on an n-component column vector $\mathbf{Z}$. If $\mathbf{Z}$ has components $z_1 \cdots z_n$, then the elements $\omega_{ij}^{(m)}$ of $\boldsymbol{\Omega}^{(m)}$ will be defined as follows: $\omega_{ij}^{(m)} = \delta_{ij}$, for $i, j = 1 \cdots m - 1, m + 2 \cdots n$; $\omega_{mm}^{(m)} = \omega_{m+1,m+1}^{(m)} = z_{m+1}(z_m^2 + z_{m+1}^2)^{-\frac{1}{2}}$; $\omega_{m+1,m}^{(m)} = -\omega_{m,m+1}^{(m)} = z_m(z_m^2 + z_{m+1}^2)^{-\frac{1}{2}}$; and all other off-diagonal elements vanish. If both z_m and z_{m+1} are equal to 0, then we define $\boldsymbol{\Omega}^{(m)}$ to be $\mathbf{I}_n$. The effect of $\boldsymbol{\Omega}^{(m)}(\mathbf{Z})$ applied to $\mathbf{Z}$ is that all components of $\mathbf{Z}$ remain unchanged, except for the m-th and $(m + 1)$-st components, of which orthogonal linear combinations are taken such as to make the m-th component equal to 0 and the $(m + 1)$-st equal to $(z_m^2 + z_{m+1}^2)^{\frac{1}{2}}$. If we put $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{(n-1)} \cdots \boldsymbol{\Omega}^{(1)}$, where $\boldsymbol{\Omega}^{(1)} = \boldsymbol{\Omega}^{(1)}(\mathbf{Z})$, $\boldsymbol{\Omega}^{(2)} = \boldsymbol{\Omega}^{(2)}(\boldsymbol{\Omega}^{(1)}\mathbf{Z})$, $\cdots$, then the first $n - 1$ components of $\boldsymbol{\Omega}\mathbf{Z}$ are zero and the n-th is $(\sum_{i=1}^n z_i^2)^{\frac{1}{2}}$. $\boldsymbol{\Omega}$ is clearly measurable.

Appendix 2. Put $\mathbf{V}_{(r)} = \mathbf{C}(\mathbf{U}_{(r)} - \boldsymbol{\mu}_0)$, $r = 1 \cdots n$, where $\mathbf{C}$ is any nonsingular $k \times k$ matrix. Substitution in (13) and (14) yields the result that T^2 retains the expression given by (14), with $(u_{ir} - \mu_{oi})$ everywhere replaced by v_{ir} (T^2 is invariant under nonsingular linear transformations of $\mathbf{U}$). If $\mathbf{C}$ is chosen such as to transform $\boldsymbol{\Sigma}$ to $\mathbf{I}_k$ (see (18)), then the v_{ir} are independent and $\mathfrak{N}(0, 1)$. It is possible now to write $T_1^2 = [1/(n - 1)]T^2$ as follows:

$$T_1^2 = n\bar{\mathbf{V}}'\{\mathbf{M}^v_{kn}(\mathbf{I}_n - \boldsymbol{\Delta}_n)(\mathbf{M}^v_{kn})'\}^{-1}\bar{\mathbf{V}}, \tag{21}$$

in which $\boldsymbol{\Delta}_n$ is an $n \times n$ matrix of which every element equals $1/n$. Let $\boldsymbol{\Omega}$ be an $n \times n$ orthogonal matrix whose last column has all elements equal to $1/(n)^{1/2}$.

[5] See [1], [5]. For completeness, a proof is also indicated in Appendix 3.

We find that $\Omega'(\mathbf{I}_n - \Delta_n)\Omega = \mathbf{J}_n$, where $\mathbf{J}_n$ is obtained from $\mathbf{I}_n$ by replacing the nn-th element by 0. If the variables x_{ir} are related to the v_{ir} by $\mathbf{M}_{kn}^x = \mathbf{M}_{kn}^v\Omega$, then $\mathbf{X}_{(n)} = (n)^{1/2}\bar{\mathbf{V}}$, and the matrix in braces in (21) is

$$\mathbf{M}_{kn}^v\Omega\Omega'(\mathbf{I}_n - \Delta_n)\Omega\Omega'(\mathbf{M}_{kn}^v)' = \mathbf{M}_{kn}^x\mathbf{J}_n(\mathbf{M}_{kn}^x)' = \mathbf{M}_{k,n-1}^x(\mathbf{M}_{k,n-1}^x)' = \mathbf{A}_{k,n-1}^x,$$

which proves (15).

Appendix 3. Since any nonsingular matrix $\mathbf{C}$ can be written as the product of elementary matrices, equation (20) need only be verified for the latter ones. Multiplication of a square matrix $\mathbf{A}$ by an elementary matrix results in either of the following elementary operations on $\mathbf{A}$:

(i) interchange of two rows (columns),

(ii) multiplication of a row (column) by a constant $c \neq 0$,

(iii) subtraction of a row (column) from another row (column). The absolute values of the determinants of the corresponding elementary matrices are 1, $|c|$, 1, respectively. The corresponding Jacobians can easily be checked to be 1, $|c|^{k+1}$, 1, respectively. This completes the proof.

REFERENCES

[1] A. C. Aitken, "On the Wishart distribution in statistics," *Biometrika*, Vol. 36 (1949), pp. 59–62.

[2] T. W. Anderson, "Multivariate statistical analysis," mimeographed lecture notes.

[3] M. S. Bartlett, "On the theory of statistical regression, *Proc. Roy. Soc.* Edinburgh, Vol. 53 (1933), pp. 260–283.

[4] H. Cramér, "*Mathematical Methods of Statistics*," Princeton University Press, 1946.

[5] W. L. Deemer and I. Olkin, "The Jacobians of certain matrix transformations useful in multivariate analysis," based on lectures by P. L. Hsu, *Biometrika*, Vol. 38 (1951), pp. 345–367.

[6] G. Elfving, "A simple method of deducing certain distributions connected with multivariate sampling," *Skand. Aktuarietids.*, Vol. 30 (1947), pp. 56–74.

[7] D. Fog, "The geometrical method in the theory of sampling," *Biometrika*, Vol. 35 (1948), pp. 46–54.

[8] H. Hotelling, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[9] P. L. Hsu, "Notes on Hotelling's generalized T," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 231–243.

[10] P. L. Hsu, "A new proof of the joint product moment distribution," *Proc. Cambridge Philos. Soc.*, Vol. 25 (1939), pp. 336–338.

[11] A. E. Ingham, "An integral which occurs in statistics," *Proc. Cambridge Philos. Soc.*, Vol. 29 (1933), pp. 271–276.

[12] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 40–75.

[13] S. Kullback, "An Application of characteristic functions to the distribution problem of statistics," *Ann. Math. Stat.*, Vol. 5 (1934), pp. 264–307.

[14] P. C. Mahalanobis, R. C. Bose, and S. N. Roy, "Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions," *Sankhyā*, Vol. 3 (1937), pp. 1–40.

[15] J. G. Mauldon, "Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory," *J. Roy. Stat. Soc.*, Ser. B., Vol. 17 (1955), pp. 79–85.

[16] J. Ogawa, "On the sampling distributions of classical statistics in multivariate analysis," *Osaka Math. J.*, Vol. 5 (1953), pp. 13–52.

[17] I. Olkin, "On distribution problems in multivariate analysis," *Inst. of Stat. Mimeo. Series*, Vol. 43 (1951), pp. 1–126.

[18] C. R. Rao, *Advanced Statistical Methods in Biometric Research*," John Wiley and Sons, 1952.

[19] G. Rasch, "A functional equation for Wishart's distribution," *Ann. Math. Stat.*, Vol. 29 (1948), pp. 262–266.

[20] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

[21] J. Wishart, "The generalized product moment distribution in sampling from a normal multivariate population," *Biometrika* Vol. 20A (1928), pp. 32–52.

[22] J. Wishart, "Proofs of the distribution law of the second order moment statistics, *Biometrika*, Vol. 35 (1948), pp. 55–57.

[23] J. Wishart and M. S. Bartlett, "The generalized product moment distribution in a normal system," *Proc. Cambridge Philos. Soc.*, Vol. 29 (1933), pp. 260–270.