so that

$$h(i, u) = \binom{n}{i} u^{i-1}(1 - u)^{n-i-1}(1 - (n/i)u) \tag{15}$$

for $0 \leqq u \leqq i/n$ which integrates to give

$$H(i, u) = \frac{1}{i}\binom{n}{i} u^{i}(1 - u)^{n-i}. \tag{16}$$

The marginal distributions are given by

$$p_i = \Pr(i^* = i) = H\left(i, \frac{i}{n}\right) = \frac{1}{i}\binom{n}{i}\left(\frac{i}{n}\right)^{i}\left(1 - \frac{i}{n}\right)^{n-i} \tag{17}$$

and

$$\begin{aligned} K(u) = \Pr(U^* \leqq u) &= \sum_{i=1}^{n} H\left(i, \min\left[\frac{i}{n}, u\right]\right) \\ &= \sum_{i=1}^{n} \frac{1}{i}\binom{n}{i}\left(\min\left[\frac{i}{n}, u\right]\right)^{i-1}\left(1 - \min\left[\frac{i}{n}, u\right]\right)^{n-i} \end{aligned} \tag{18}$$

The algebraic identity implied by the relation $\sum_{i=1}^{n} p_i = 1$, like that in (8), has been indirectly derived. Both identities may be algebraically proved using the formula quoted as (5) in [2].

We repeat that the method used here to obtain $H(i, u)$ is applicable to the general class of barriers $(\delta(d), \epsilon(d))$.

REFERENCES

[1] Z. W. Birnbaum and F. H. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[2] Z. W. Birnbaum and R. Pyke, "On some distributions related to the statistic D_n^+," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 179–187.

[3] D. G. Chapman, "A Comparative study of several one-sided goodness-of-fit tests," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 655–674.

[4] R. Pyke, "The supremum and infimum of the Poisson process," *Technical Report No.* 39, Applied Mathematics and Statistics Laboratory, Stanford University, March 1958 (Abstract in *Ann. Math. Stat.*, Vol. 29 (1958), p. 327).

APPLICATIONS OF A CERTAIN REPRESENTATION OF THE WISHART MATRIX

By Robert A. Wijsman

University of Illinois

0. Summary. Apart from pre- and post-multiplication by a fixed matrix and its transpose, the Wishart matrix **A** can be written as the product of a triangular matrix and its transpose, whose elements are independent normal and chi variables. Various applications of this representation are indicated. Examples

Received July 17, 1958.

are given concerning the diagonal elements of $\mathbf{A}^{-1}$, the sample ordinary and multiple correlation coefficient, the characteristic roots of $\mathbf{A}$ and the sphericity criterion in the bivariate case.

1. Introduction. Let $\mathbf{X}$ be a $p \times n$ matrix, with $p \leqq n$, whose columns are independent and distributed according to a p-variate normal law with mean vector 0, and covariance matrix Σ. The matrix $\mathbf{A} = \mathbf{XX}'$ will be called the *Wishart matrix*. If $\mathbf{\Sigma} = \mathbf{I}_p$ (the $p \times p$ identity matrix), then $\mathbf{A}$ can be written in the form

$$\mathbf{A} = \mathbf{TT}' \tag{1}$$

in which $\mathbf{T}$ is a triangular $p \times p$ matrix whose elements are independent random variables, the off-diagonal elements being $N(0, 1)$, and the diagonal elements being χ variables with certain degrees of freedom.[1] More specifically, if we choose $\mathbf{T}$ to be lower diagonal (which we shall do from now on), then T_{ii} is a χ variable with $n - i + 1$ degrees of freedom. The representation (1) is known; in fact, it is implied by the Bartlett decomposition [2]. However, only few authors, like Mauldon [6], state the representation (1) and the nature of $\mathbf{T}$ explicitly.[2] Equation (1) is also implied in [8]. It is the purpose of this note to point out some of the applications of (1), or, as the case may be, of the more general equation (2) below.

If $\mathbf{\Sigma}$ does not necessarily equal $\mathbf{I}_p$, let $\mathbf{C}$ be a $p \times p$ matrix such that

$$\mathbf{CC}' = \mathbf{\Sigma}.$$

Then the Wishart matrix can be represented as

$$\mathbf{A} = \mathbf{CTT}'\mathbf{C}' \tag{2}$$

with $\mathbf{T}$ as before. It may be convenient in applications to choose $\mathbf{C}$ also lower triangular. If $\mathbf{\Sigma} = \mathbf{I}_p$, we may take $\mathbf{C} = \mathbf{I}_p$, so that (2) reduces to (1).

Equation (2) can be used in several ways. In the first place the Wishart distribution can be derived very easily starting from the distribution of $\mathbf{T}$, since the Jacobian of the transformation (2) from $\mathbf{T}$ to $\mathbf{A}$ is simple to compute. Secondly, if it is desired to generate values of $\mathbf{A}$, or of a function of $\mathbf{A}$, by a random process (for an application see [7]), this can be done conveniently by generating values of $\mathbf{T}$. In the third place, the distribution and certain properties of functions of $\mathbf{A}$ can sometimes be obtained quite easily by expressing them as functions of the elements of $\mathbf{T}$. It is of this third kind of application that we will give some examples. Concerning notation and nomenclature, the ratio of a normal and a central χ variable will be called a t' variable, and the ratio of a χ^2 variable and a central χ^2 variable will be called an F' variable (the primes are used here to distinguish these variables from the customary t and F variables, in which the χ^2 variables have been divided by their degrees of

[1] In [8], footnote 3, the diagonal elements were erroneously termed χ^2 variables.

[2] Very recently, A. M. Kshirsagar (*Ann. Math. Stat.*, Vol. 30 (1959), pp. 239–241) also gave the decomposition (1) and a simple derivation.

freedom). The degrees of freedom of a t' or F' variable will be indicated by subscripts on t' or F'.

2. Applications.

EXAMPLE 1: THE DIAGONAL ELEMENTS OF $\mathbf{A}^{-1}$ IF $\mathbf{\Sigma} = \mathbf{I}_p$. These diagonal elements are obviously identically distributed, so it suffices to consider $(A^{-1})_{pp}$. By (1) we have $\mathbf{A}^{-1} = \mathbf{T}'^{-1}\mathbf{T}^{-1}$ (with probability 1, $\mathbf{T}$ is non-singular). Now $\mathbf{T}^{-1}$ is also lower triangular, and its diagonal elements are the reciprocals of the corresponding diagonal elements of $\mathbf{T}$. Thus, we find

$$(A^{-1})_{pp} = (T'^{-1})_{pp}(T^{-1})_{pp} = 1/T^2_{pp}.$$

Hence, each diagonal element of $\mathbf{A}^{-1}$ is the reciprocal of a χ^2_{n-p+1} variable. This result can be applied to exhibit Hotelling's T^2 as a constant times an F variable [8].

EXAMPLE 2: THE SAMPLE CORRELATION COEFFICIENT. This is essentially a bivariate problem, so we may set $p = 2$. Let the population correlation coefficient be ρ, the sample correlation coefficient $r = A_{12}(A_{11}A_{22})^{-1/2}$. It suffices to assume $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = \rho$. This can most conveniently be effected by choosing $\mathbf{C}$ lower triangular, with $C_{11} = 1$, $C_{21} = \rho$, $C_{22} = (1 - \rho^2)^{1/2}$. From (2) we compute then

$$\frac{r}{\sqrt{1 - r^2}} = \frac{T_{21} + T_{11}\,\rho\,/\,\sqrt{1 - \rho^2}}{T_{22}}. \tag{3}$$

The same expression was also obtained by Elfving [3], following a different method. The right hand side of (3) can be described as a non-central t'_{n-1} variable, with a random non-centrality parameter $T_{11}\rho(1 - \rho^2)^{-1/2}$, which is a χ_n variable times $\rho(1 - \rho^2)^{-1/2}$. From this remark an expression for the density $p(\rho, \cdot)$ of $r(1 - r^2)^{-1/2}$ follows at once. Let $f(\rho, \cdot)$ be the density of $T_{11}\rho(1 - \rho^2)^{-1/2}$, and $g(\xi, \cdot)$ the density of a non-central t'_{n-1} variable with non-centrality parameter ξ. Then

$$p(\rho, x) = \int_0^\infty f(\rho, \xi)g(\xi, x)\,d\xi. \tag{4}$$

From (4) the monotonicity of the probability ratio follows then immediately by applying a theorem of Lehmann [5] (theorem 3), or a theorem of Karlin [4] (lemma 5).

EXAMPLE 3: THE SAMPLE MULTIPLE CORRELATION COEFFICIENT. Let $\bar{R}$ be the population multiple correlation coefficient between the pth variate and the first $p - 1$ variates and let R be the corresponding sample quantity. Then

$$1 - R^2 = |\mathbf{A}|\,/\,|\mathbf{A}^*|\,A_{pp} \tag{5}$$

where $\mathbf{A}^*$ is obtained from $\mathbf{A}$ by deleting the last row and column. It is sufficient to choose $\mathbf{C}$ to be lower triangular, with $C_{11} = \cdots = C_{p-1,\,p-1} = 1$, $C_{pp} =$

$(1 - \bar{R}^2)^{1/2}$, $C_{p1} = \bar{R}$, all other elements equal to 0. Substituting into (2) and using (5), we derive

$$\frac{R^2}{1 - R^2} = \frac{(T_{p1} + T_{11}\bar{R} / \sqrt{1 - \bar{R}^2})^2 + \sum_{i=2}^{p-1} T_{pi}^2}{T_{pp}^2} \tag{6}$$

This can be described as a non-central $F'_{p-1,n-p+1}$ variable, with a random non-centrality parameter $T_{11}^2\bar{R}^2/(1 - \bar{R}^2)$, which is $\bar{R}^2/(1 - \bar{R}^2)$ times a χ_n^2 variable. An expression for the density of $R^2/(1 - R^2)$ can then be written down at once, and the monotonicity of the probability ratio follows in a similar way as in Example 2.

Example 4: The characteristic roots of $\mathbf{A}$ and the sphericity criterion in the case $p = 2$ and $\mathbf{\Sigma} = \mathbf{I}_2$. Let the square roots of the characteristic roots of $\mathbf{A}$ be λ_1 and λ_2 ($\lambda_1 \geqq \lambda_2$). We have

$$\lambda_1^2 + \lambda_2^2 = \operatorname{tr} \mathbf{A} \tag{7}$$

$$\lambda_1^2\lambda_2^2 = |\mathbf{A}| \tag{8}$$

The joint distribution of λ_1 and λ_2 is determined by the joint distribution of $(\lambda_1 - \lambda_2)^2$ and $2\lambda_1\lambda_2$, which turns out to be very simple. Using (7), (8) and (1) we compute

$$(\lambda_1 - \lambda_2)^2 = (T_{11} - T_{22})^2 + T_{21}^2 \tag{9}$$

$$2\lambda_1\lambda_2 = 2T_{11}T_{22}\,. \tag{10}$$

By the lemma below, the right hand sides of (9) and (10) are independent, and distributed as χ_2^2 and χ_{2n-2}^2 respectively.

Lemma. *If X and Y are independent and distributed as χ_n, χ_{n-1}, respectively, then $(X - Y)^2$ and $2XY$ are independent and distributed as χ_1^2 and χ_{2n-2}^2 respectively.*

The proof is straightforward, and will be omitted.

The sphericity criterion (Anderson [1], section 10.7) in the bivariate case is the ratio $Z = 2\,|\mathbf{A}|^{1/2} / \operatorname{tr} \mathbf{A}$, which can also be written as

$$Z = 2\lambda_1\lambda_2/(\lambda_1^2 + \lambda_2^2).$$

Using (9) and (10) we find

$$\frac{Z}{1 - Z} = \frac{2T_{11}\,T_{22}}{(T_{11} - T_{22})^2 + T_{21}^2} \tag{11}$$

which is an $F'_{2n-2,2}$ variable, by the lemma.

The statistic Z, or, equivalently, $Z / (1 - Z)$, can be used to test the hypothesis H that $\mathbf{\Sigma} = \sigma^2\mathbf{I}_2$ (σ^2 unknown) against the alternative that this is not so. The likelihood ratio test rejects H if $Z <$ constant. It can be shown that this test is also uniformly most powerful invariant. In the first place, Z is maximal invariant. Secondly, the distribution of Z depends on a single parameter only, e.g. on $2 \mid \mathbf{\Sigma} \mid^{1/2} / \operatorname{tr} \mathbf{\Sigma}$. In the third place, it can be shown that the probability ratio is monotonic. This can be demonstrated either by starting from the Wishart distribution, or by using (2). However, in this example the latter way does not seem to be any simpler than the former. The moral seems to be that in some cases the utilization of the representation (1) or (2) leads to the results in a fast and elegant way, but in other cases the conventional approach may be more practical.

REFERENCES

[1] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.
[2] M. S. BARTLETT, "On the theory of statistical regression," *Proc. Roy. Soc. Edinburgh*, Vol. 53 (1933), pp. 260–283.
[3] G. ELFVING, "A simple method of deducing certain distributions connected with multivariate sampling," *Skand. Aktuarietids.*, Vol. 30 (1947), pp. 56–74.
[4] SAMUEL KARLIN, "Decision theory for Pólya type distributions. Case of two action, I." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 1956, pp. 115–128.
[5] E. L. LEHMANN, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.
[6] 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.
[7] DANIEL TEICHROEW AND ROSEDITH SITGREAVES, "Computation of an empirical sampling distribution for the classification statistic *W*," *Probability and Statistics in Item Analysis and Classification Problems*, School of Aviation Medicine, USAF, Texas, No. 57–98 (1957).
[8] ROBERT A. WIJSMAN, "Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 415–423.

ON DVORETZKY'S STOCHASTIC APPROXIMATION THEOREM

C. DERMAN[1] AND J. SACKS[2]

Columbia University

1. Introduction. A very general theorem was proved by Dvoretzky [2] on the convergence of transformations with superimposed random errors. This work followed that of Robbins-Monro [5] and others (see [6] for bibliography) and contains the most comprehensive results on convergence (with probability one and in mean square) of the stochastic approximation procedures of Robbins-

Received August 21, 1958.
[1] This research was sponsored by the Office of Naval Research under Contract Nonr-266(55), Project NR 042–099.
[2] This research was sponsored by the Office of Naval Research under Contract Nonr-266(51), Project NR 043–198.