ON THE EXACT DISTRIBUTIONS OF THE LIKELIHOOD RATIO CRITERIA FOR TESTING LINEAR HYPOTHESES ABOUT REGRESSION COEFFICIENTS

By P. C. Consul

University of Libya

1. Introduction. Wilks (1932) defined a number of likelihood ratio criteria for testing the equality of means, equality of variances and equality of covariances from several populations. These criteria are being very widely used in applied work for tests of significance in multivariate analysis. Wilks (1934) and Bartlett (1934) extended their use for testing some linear hypotheses about regression coefficients.

If $x_1, x_2, x_3, \cdots, x_N$ are a set of vector observations, with associated fixed vectors $z_1, z_2, z_3, \cdots, z_n$, where x_α is an observation from $N(\beta_{z\alpha}, \Sigma)$, and if the matrix β is partitioned such that $\beta = (\beta_1\beta_2)$ where β_1 has q_1 columns and β_2 has q_2 columns, then the likelihood ratio criterion λ, for testing the hypothesis that the matrix β_1 is equal to some given matrix, is given by

$$(1.1) \qquad \lambda = |\Sigma_1|^{N/2}/(|\Sigma_2|^{N/2})',$$

where Σ_1 and Σ_2 (sum of products matrices) are the maximum likelihood estimates of $p \times p$ matrix Σ over the full range and over the restricted range under the hypothesis.

Wilks (1932) has also obtained the hth moment of the criterion $U = \lambda^{2/N}$. When $N - q_1 - q_2 = n$ and $q_1 = m$, Anderson shows that the hth moment of $U_{p,m,n}$ can be put in the form

$$(1.2) \quad M_h(U_{p,m,n}) = \prod_{i=1}^{p} \Gamma[\tfrac{1}{2}(n+1-i)+h]\cdot\Gamma[\tfrac{1}{2}(n+m+1-i)]/ \Gamma[\tfrac{1}{2}(n+1-i)]\cdot\Gamma[\tfrac{1}{2}(n+m+1-i)+h].$$

Wilks (1935) obtained the distribution of $U_{p,m,n}$ in the form of a $(p-1)$ fold multiple integral, which he was able to evaluate for $p = 1, 2$; $p = 3$ with $m = 3, 4$ and for $p = 4$ with $m = 4$ only.

Bartlett (1938) suggested χ^2-significance points for $-m \log U_{p,m,n}$. Wald & Brookner (1941) obtained an asymptotic expansion for the distribution of $(-2 \log \lambda)$ and it was modified into a new form by Rao (1948). Box (1949) has given a general method of obtaining the asymptotic distributions of such criteria. Consul (1965) has given another similar general method. However, all these methods provide approximate values only.

Anderson (1958) has shown that the distribution of $U_{p,m,n}$ is that of a product of a number of independent beta variates and by integrating their joint densities he obtains explicit expressions for the distributions of $U_{p,m,n}$ for $p = 1$,

Received 7 July 1965; revised 23 May 1966.

2, while for $p = 3$, with $m = 3, 4$ and $p = 4$ with $m = 4$ he could obtain cumulative distributions only but his result for $p = 3$ with $m = 4$ disagrees with Wilks' result.

The exact distributions of $U_{p,m,n}$ for all p and m have been found by Schatzoff (1964) and he gives closed form expressions when either p or m is an even integer and provides exact representations in integral form for all cases where p and m are both odd.

In this paper we apply operational calculus to the expression (1.2) of hth moment of $U_{p,m,n}$ to obtain the exact probability distributions of $U_{p,m,n}$ for $p = 1, 2, 3, 4$, and show that the distributions are in the form of Gauss hypergeometric functions. Since detailed tabulated values of hypergeometric functions do not exist, we also obtain the algebraic forms of these distributions for $p = 1, 2$ and for $p = 3, 4$, with $m = 3, 4, 5, 6, 7, 8$, and the respective cumulative distribution functions by conversion and integration. Our result (3.3.8) confirms that for $p = 3$ with $m = 4$, Anderson's result is correct and Wilks' result is incorrect.

2. Some known results and integrals. For ready reference we quote here some known results and integrals as they are required at many places in the text.

(i) Titchmarsh gives the inverse Mellin transform

$$(2.1)\quad (1/2\pi i)\int_{c-i\infty}^{c+i\infty} x^{-s}\cdot[\Gamma(s+a)/\Gamma(s+a+m)]\,ds = (1/\Gamma(m))\cdot x^{a}(1-x)^{m-1},\quad 0 \leqq x \leqq 1.$$

(ii) Consul (1965) has obtained the inverse Mellin transform

$$(2.2)\quad (1/2\pi i)\int_{c-i\infty}^{c+i\infty} x^{-s}\cdot[\Gamma(s+a)\Gamma(s+b)/\Gamma(s+a+m)\Gamma(s+b+n)]\,ds = [x^{a}(1-x)^{m+n-1}/\Gamma(m+n)]\cdot F(n, a-b+m; m+n; 1-x)$$

which can be easily modified into

$$(2.3)\quad \begin{aligned} &(1/2\pi i)\int_{c-i\infty}^{c+i\infty} x^{-s} \\ &\qquad\cdot[\Gamma(2s+a)\Gamma(2s+b)/\Gamma(2s+a+m)\Gamma(2s+b+n)]\,ds \\ &= [x^{a/2}(1-x^{\frac{1}{2}})^{m+n-1}/2\Gamma(m+n)] \\ &\qquad\cdot F(n, a-b+m; m+n; 1-x^{\frac{1}{2}}). \end{aligned}$$

(iii) Consul (1965) has also proved that

$$(2.4)\quad \begin{aligned} &(1/2\pi i)\int_{c-i\infty}^{c+i\infty} x^{-s} \\ &\qquad\cdot[\Gamma(2s+a)\Gamma(s+b)/\Gamma(2s+a+m)\Gamma(s+b+n)]\,ds \\ &= [x^{a/2}(1-x)^{n}/2\Gamma(m)\Gamma(n+1)] \\ &\qquad\cdot\sum_{i=0}^{m-1}\binom{m-1}{i}\cdot(-x^{\frac{1}{2}})^{i}F(n, 1-b+\tfrac{1}{2}(a+i); n+1; 1-x). \end{aligned}$$

(iv) Gauss recurrence formula for hypergeometric functions viz.

(2.5) $$F(a, b; c + 1; x) = [cx^{-1}/(c - b)][F(a - 1, b; c; x) - (1 - x)F(a, b; c; x)].$$

(v) Consul has further proved the reduction formula

(2.6) $$F(a, b; c; x) = [(c - m)_m/(c - a - m)_m](x)^{-m} \cdot \sum_{i=0}^{m} \binom{m}{i}(-1)^i(1 - x)^i F(a, b - m + i; c - m; x)$$

where $c > a + m$.

(vi) Consul has also proved the results

(2.7) $$\begin{aligned} F(2, 1 + b; 3; x) &= [2(1 - x)^{-b}/b(b - 1)x^2][(1 - x)^b + bx - 1], \quad b \neq 0, 1, \\ &= -2x^{-2}[\log (1 - x) + x], \quad b = 0, \\ &= 2x^{-2} \cdot [x(1 - x)^{-1} + \log (1 - x)], \quad b = 1; \end{aligned}$$

(2.8) $$\begin{aligned} F(3, 1 + b; 4; x) &= [6x^{-3}(1 - x)^{-b}/b][\tfrac{1}{2}x^2 - x(1 - x)/(b - 1) \\ &\quad + [(1 - x)^2 - (1 - x)^b]/(b - 1)(b - 2)], \quad b \neq 0, 1, 2, \\ &= -3x^{-3}[x + x^2/2 + \log (1 - x)], \quad \text{if } b = 0, \\ &= 3x^{-3}[(1 - x)^{-1} - (1 - x) + 2 \log (1 - x)], \quad \text{if } b = 1, \\ &= 3x^{-3}(1 - x)^{-2}[\tfrac{3}{2}x^2 - x - (1 - x)^2 \log (1 - x)], \quad \text{if } b = 2. \end{aligned}$$

3. Exact probability distributions.

3.0 *Exact probability distribution of the likelihood criterion* $U_{p,m,n}$. By applying Mellin's inversion theorem on the hth moment, given by the expression (1.2), of the criterion $U_{p,m,n}$, we get the exact probability distribution of the criterion $U_{p,m,n}$ as

(3.0.1) $$\begin{aligned} f(U_{p,m,n}) &= \prod_{i=1}^{p} [\Gamma[\tfrac{1}{2}(n + m + 1 - i)]/\Gamma[\tfrac{1}{2}(n + 1 - i)]] \\ &\quad \cdot (1/2\pi i) \int_{c'-i\infty}^{c'+i\infty} U^{-h-1} \\ &\quad \cdot \prod_{i=1}^{p} [\Gamma[\tfrac{1}{2}(n + 1 - i) + h]/\Gamma[\tfrac{1}{2}(n + m + 1 - i) + h]]\, dh \end{aligned}$$

which, on putting $h + \frac{1}{2}(n + 1 - p) = t$ and on further simplification, takes the form

(3.0.2) $$\begin{aligned} f(U_{p,m,n}) &= U^{(n-1-p)/2} \\ &\quad \cdot \prod_{i=1}^{p} [\Gamma[\tfrac{1}{2}(n + m + 1 - i)]/\Gamma[\tfrac{1}{2}(n + 1 - i)]] \\ &\quad \cdot (1/2\pi i) \int_{c-i\infty}^{c+i\infty} U^{-t} \\ &\quad \cdot \prod_{i=1}^{p} [\Gamma[t + \tfrac{1}{2}(p - i)]/\Gamma[t + \tfrac{1}{2}(m + p - i)]]\, dt. \end{aligned}$$

The above representation of the exact distribution as an inverse Mellin transform is interesting because the distribution splits into a factor depending on n and an integral depending only upon p and U. However, the expression does not seem to get integrated for all values of p. We use it to determine the exact distributions $f(U)$ of the criterion $U_{p,m,n}$ for some values of p viz. $p = 1, 2, 3, 4$.

3.1 *Exact distributions of* $U_{1,m,n}$ *and* $U_{2,m,n}$. When $p = 1$ and 2, the expression (3.0.2) together with the results (2.1) and (2.2), gives the exact distributions of $U_{1,m,n}$ and $U_{2,m,n}$ in the form

$$(3.1.1) \quad f(U_{1,m,n}) = [\Gamma[\tfrac{1}{2}(n+m)]/\Gamma(n/2)\Gamma(m/2)]\cdot U^{\frac{1}{2}n-1}(1-U)^{\frac{1}{2}m-1}, \quad 0 \leqq U \leqq 1,$$

and

$$(3.1.2) \quad f(U_{2,m,n}) = [\Gamma[\tfrac{1}{2}(n+m-1)\Gamma[\tfrac{1}{2}(n+m)]/\Gamma[\tfrac{1}{2}(n-1)]\Gamma(\tfrac{1}{2}n)\Gamma(m)] \cdot U^{\frac{1}{2}(n-3)}(1-U)^{m-1}\cdot F(\tfrac{1}{2}m, \tfrac{1}{2}m-\tfrac{1}{2}; m; 1-U), \quad 0 \leqq U \leqq 1.$$

The result (3.1.2) can be easily put into another form given below:

$$(3.1.3) \quad f(U_{2,m,n}) = [\Gamma(n+m-1)/2\Gamma(n-1)\Gamma(m)]\cdot U^{\frac{1}{2}(n-3)}\cdot(1-U^{\frac{1}{2}})^{m-1}, \quad 0 \leqq U \leqq 1.$$

The results (3.1.1) and (3.1.3) are very well known and were obtained by Wilks (1935) and Anderson (1958) by other methods.

3.2 *Exact distribution of* $U_{3,m,n}$. When $p = 3$, we find from the expression (3.0.2) that the exact distribution of $U_{3,m,n}$ is given by

$$[\Gamma[\tfrac{1}{2}(n+m-2)]\Gamma[\tfrac{1}{2}(n+m-1)]\Gamma[\tfrac{1}{2}(n+m)]/\Gamma[\tfrac{1}{2}(n-2)]\Gamma[\tfrac{1}{2}(n-1)]\Gamma(\tfrac{1}{2}n)] \cdot U^{\frac{1}{2}(n-4)}\cdot(1/2\pi i)\int_{c-i\infty}^{c+i\infty} U^{-t} \cdot[\Gamma(t)\Gamma(t+\tfrac{1}{2})\Gamma(t+1)/\Gamma(t+\tfrac{1}{2}m)\Gamma(t+\tfrac{1}{2}m+\tfrac{1}{2})\Gamma(t+\tfrac{1}{2}m+1)]\,dt$$

which, on simplification with the help of Legendre's duplication formula, becomes

$$[\Gamma(n+m-1)\Gamma[\tfrac{1}{2}(n+m)-1]/\Gamma(n-1)\Gamma(\tfrac{1}{2}n-1)] \cdot U^{n/2-2}\cdot(1/2\pi i)\int_{c-i\infty}^{c+i\infty} U^{-t}\cdot[\Gamma(2t)\Gamma(t+1)/\Gamma(2t+m)\Gamma(t+\tfrac{1}{2}m+1)]\,dt.$$

Now, by evaluating the integral with the help of Consul's inverse Mellin transform (2.3), we obtain the exact distribution of $U_{3,m,n}$ as

$$(3.2.0) \quad f(U_{3,m,n}) = [\Gamma(n+m-1)\Gamma[\tfrac{1}{2}(n+m)-1]/ \Gamma(n-1)\Gamma(\tfrac{1}{2}n-1)\cdot 2\Gamma(m)\Gamma(\tfrac{1}{2}m+1)] \cdot U^{n/2-2}(1-U)^{m/2} \cdot \sum_{i=0}^{m-1}\binom{m-1}{i}(-U^{\frac{1}{2}})^i\cdot F(\tfrac{1}{2}m, \tfrac{1}{2}i; \tfrac{1}{2}m+1; 1-U)$$

where $0 \leqq U \leqq 1$.

The above expression shows that for all values of m, the exact distribution of $U_{3,m,n}$ consists of a number of terms each of which contains a hypergeometric function. Wilks (1935) and Anderson (1958) obtained the cumulative distributions of $U_{3,3,n}$ and $U_{3,4,n}$ in the form of algebraic expressions but their results for $U_{3,4,n}$ differ with each other.

For particular values of m, the exact distribution of $U_{3,m,n}$, given by our expression (3.2.0), can be expressed in terms of known algebraic functions by first reducing the values of the parameters in the hypergeometric functions with the help of the reduction formula (2.6) and the repeated use of the recurrence relation (2.5) and then by replacing the resulting hypergeometric functions with their known values by the formulae (2.7) or (2.8).

To obtain the probabilities that $U_{3,m,n} \leqq u$ ($\leqq 1$), for different values of m, we can integrate the resulting algebraic expressions with respect to U between the limits 0 to u and the respective integrals, on simplification, will denote the cumulative distribution function $\Pr(U_{3,m,n} \leqq u)$.

For convenience, we list below the exact distribution functions and the cumulative distribution functions of $U_{3,m,n}$ in algebraic form for the particular cases $m = 3, 4, 5, 6, 7$ and 8.

CASE I. For $m = 3$, the exact distribution of $U_{3,3,n}$ is given by

$$(3.2.1)\quad f(U_{3,3,n}) = CU^{n/2-2}[(1 - U)^{\frac{3}{2}} - 3U^{\frac{1}{2}} \sin^{-1}(1 - U)^{\frac{1}{2}} + 3U \log\{U^{-\frac{1}{2}} + U^{-\frac{1}{2}}(1 - U)^{\frac{1}{2}}\}]$$

where $0 \leqq U \leqq 1$ and

$$C = [\Gamma(n + 2)\Gamma[\tfrac{1}{2}(n + 1)]/\Gamma(n - 1)\Gamma(\tfrac{1}{2}n - 1)\cdot 3\pi^{\frac{1}{2}}]$$

and the cumulative distribution function of $U_{3,3,n}$ is given by

$$\begin{aligned}(3.2.2)\quad &\Pr(U_{3,3,n} \leqq u)\\ &= \tfrac{1}{2}[n(n - 1)I_u(\tfrac{1}{2}n - 1, \tfrac{5}{2}) - (n + 1)(n - 2)I_u(\tfrac{1}{2}n, \tfrac{1}{2})]\\ &\quad - 6CU^{(n-1)/2}[(n - 1)^{-1}\sin^{-1}(1 - u)^{\frac{1}{2}}\\ &\quad - n^{-1}u^{\frac{1}{2}}\log\{u^{-\frac{1}{2}} + u^{-\frac{1}{2}}(1 - u)^{\frac{1}{2}}\}]\end{aligned}$$

where the incomplete beta function, $I_u(a, b) = B^{-1}(a, b)\int_0^u x^{a-1}(1 - x)^{b-1}\,dx$, has been tabulated by Pearson (1932).

It can be easily shown (see Appendix) that the result (3.2.2) is a simplified form of the results obtained by Wilks (1935) and Anderson (1958) by other methods.

CASE II. For $m = 4$, the exact distribution of $U_{3,4,n}$ becomes

$$(3.2.3)\quad f(U_{3,4,n}) = CU^{n/2-2}[1 - U^2 + 8U^{\frac{1}{2}}(1 - U) - 6U \log U]$$

where $0 \leqq U \leqq 1$ and

$$C = [\Gamma(n + 3)\Gamma(\tfrac{1}{2}n + 1)/\Gamma(n - 1)\Gamma(\tfrac{1}{2}n - 1) \times 24]$$

and the cumulative distribution of $U_{3,4,n}$ is

$$\text{(3.2.4)} \quad \Pr(U_{3,4,n} \leqq u) = 2C \cdot u^{n/2-1}[1/(n-2) - u^2/(n+2) - 8u^{\frac{1}{2}}/(n-1) + 8u^{\frac{3}{2}}/(n+1) - (6u/n)\log u + 12u/n^2]$$

which is the same as that obtained by Anderson (1958) by another method. Thus the expression obtained by Wilks (1935) was definitely incorrect.

CASE III. For $m = 5$, the exact and cumulative distributions of $U_{3,5,n}$ take the following respective forms:

$$\text{(3.2.5)} \quad \begin{aligned} f(U_{3,5,n}) = C \cdot U^{n/2-2}[&(1-U)^{\frac{5}{2}} - 45U(1-U)^{\frac{1}{2}} \\ &+ (30U^{\frac{3}{2}} - \tfrac{15}{2}U^{\frac{5}{2}}) \sin^{-1}(1-U)^{\frac{1}{2}} \\ &+ (30U - \tfrac{15}{2}U^2) \log\{U^{-\frac{1}{2}} + U^{-\frac{1}{2}}(1-U)^{\frac{1}{2}}\}] \end{aligned}$$

and

$$\text{(3.2.6)} \quad \begin{aligned} \Pr(U_{3,5,n} \leqq u) = I_u(\tfrac{1}{2}n - 1, \tfrac{5}{2}) + C[&2u^{n/2-1}(1-u)^{\frac{5}{2}}/(n+3) \\ &- [15(3n+7)/(n+1)(n+2)]U^{n/2}(1-U)^{\frac{1}{2}} \\ &- [15(3n-4)/n(n-1)]u^{n/2-1}(1-u)^{\frac{1}{2}} \\ &+ [5(9n^3+17n^2-16n-16)/(n-1)4]u^{n/2-1}(1-u)^{\frac{3}{2}} \\ &+ \{4u/(n+1) - 1/(n-1)\} \\ &\quad \cdot 15u^{(n-1)/2} \sin^{-1}(1-u)^{\frac{1}{2}} + \{4/n - u/(n+2)\} \\ &\quad \cdot 15u^{n/2} \log\{u^{-\frac{1}{2}} + u^{-\frac{1}{2}}(1-u)^{\frac{1}{2}}\}] \end{aligned}$$

where $0 \leqq U \leqq u \leqq 1$ and $I_u(a, b)$ is the incomplete beta function and

$$C = [\Gamma(n+4)\Gamma[\tfrac{1}{2}(n+3)]/\Gamma(n-1)\Gamma(\tfrac{1}{2}n-1) \times 90\pi^{\frac{1}{2}}].$$

CASE IV. For $m = 6$, the exact and the cumulative distributions of $U_{3,6,n}$ respectively become

$$\text{(3.2.7)} \quad f(U_{3,6,n}) = C \cdot U^{n/2-2}[1 - 16U^{\frac{1}{2}} - 65U + 160U^{\frac{3}{2}} - 65U^2 - 16U^{\frac{5}{2}} + U^3 - 30U(1-U)\log U]$$

and

$$\text{(3.2.8)} \quad \begin{aligned} \Pr(U_{3,6,n} \leqq u) = 2C \cdot u^{n/2-1}[&1/(n-2) - 16u^{\frac{1}{2}}/(n-1) - 65u/n \\ &+ 60u/n^2 + 160u^{\frac{3}{2}}/(n+1) - 65u^2/(n+2) \\ &- 60u^2/(n+2)^2 - 16u^{\frac{5}{2}}/(n+3) + u^3/(n+4) \\ &+ \{30u^2/(n+2) - 30u/n\} \log u] \end{aligned}$$

where $0 \leqq U \leqq u \leqq 1$ and

$$C = [\Gamma(n+5)\Gamma(\tfrac{1}{2}n+2)/\Gamma(n-1)\Gamma(\tfrac{1}{2}n-1) \times 1440].$$

CASE V. For $m = 7$, the algebraic forms of the exact and cumulative distributions of $U_{3,7,n}$ respectively are

$$\begin{aligned} f(U_{3,7,n}) = CU^{(n/2)-2}[(1 - U)^{7/2} &- (1855/8)U(1 - U)^{\frac{3}{2}} \\ &- (105/8)(1 - 20U + 8U^2)U^{\frac{1}{2}} \sin^{-1} (1 - U)^{\frac{1}{2}} \\ &+ (105/8)(8 - 20U + U^2)U \log (\{1 + (1 - U)^{\frac{1}{2}}\}U^{-\frac{1}{2}})] \end{aligned} \tag{3.2.9}$$

and

$$\begin{aligned} \Pr (U_{3,7,n} \leqq u) = I_u(n/2, \tfrac{5}{2}) &+ C \cdot u^{\frac{1}{2}n}[2u^{-1}(1 - u)^{7/2}/(n - 2) \\ &+ 14(1 - u)^{\frac{5}{2}}/(n - 2)(n + 5) + (105/4) \\ &\cdot(8/n - 20u/(n + 2) + u^2/(n + 4)) \\ &\cdot\log (\{1 + (1 - u)^{\frac{1}{2}}\}u^{-\frac{1}{2}}) - (105/4) \\ &\cdot(1/(n - 1) - 20u/(n + 1) + 8u^2/(n + 3)) \\ &\cdot u^{-\frac{1}{2}} \sin^{-1} (1 - u)^{\frac{1}{2}} + [35(1 - u)^{\frac{3}{2}}/4] \\ &\cdot\{(7n+29)(n+2+u-nu)/(n+3)(n+4) \\ &- (7n - 8)(n + 4 - nu - u)/n(n - 1) \\ &- 20(n + 3 - nu)/(n + 1)(n + 3)\}] \end{aligned} \tag{3.2.10}$$

where $0 \leqq U \leqq u \leqq 1$ and $I_u(n/2, \frac{5}{2})$ is the incomplete beta function tabulated by Pearson and

$$C = [\Gamma(n + 6)\Gamma\{\tfrac{1}{2}(n + 5)\}/\Gamma(n - 1)\Gamma(\tfrac{1}{2}n - 1)\ 2\Gamma(7)\Gamma(\tfrac{9}{2})].$$

CASE VI. For $m = 8$, the exact and cumulative distributions of $U_{3,8,n}$ respectively become

$$\begin{aligned} f(U_{3,8,n}) = CU^{\frac{1}{2}n-2}[1 &- (128/5)U^{\frac{1}{2}}(1 - U^3) - (1428/5)U(1 - U^2) \\ &+ 896U^{\frac{3}{2}}(1 - U) - U^4 - 84U(1 - 5U + U^2) \log U] \end{aligned} \tag{3.2.11}$$

and

$$\begin{aligned} \Pr (U_{3,8,n} \leqq u) = 2C \cdot u^{\frac{1}{2}n-1}[1/(n - 2) &- 128u^{\frac{1}{2}}/5(n - 1) \\ &- 1428u/5n + 896u^{\frac{3}{2}}/(n + 1) - 896u^{\frac{5}{2}}/(n + 3) \\ &+ 1428u^3/5(n + 4) + 128u^{7/2}/5(n + 5) \\ &- u^4/(n + 6) \\ &- 336\{u/n^2 - 5u^2/(n + 2)^2 + u^3/(n + 4)^3\} \\ &- 168\{u/n - 5u^2/(n + 2) + u^3/(n + 4)\} \log u] \end{aligned} \tag{3.2.12}$$

where $0 \leqq U \leqq u \leqq 1$, and

$$C = [\Gamma(n + 7)\Gamma(\tfrac{1}{2}n + 3)/\Gamma(n - 1)\Gamma(\tfrac{1}{2}n - 1)\ 2\Gamma(8)\Gamma(5)]$$

3.3 *Exact probability distribution of* $U_{4,m,n}$. When $p = 4$, we see that the exact probability distribution of $U_{4,m,n}$ is given by the expression (3.0.2) in the form

$$f(U_{4,m,n}) = C\cdot U^{(n-5)/2}\cdot(1/2\pi i)\int_{c-i\infty}^{c+i\infty} U^{-t}\cdot[\Gamma(t)\Gamma(t+\tfrac{1}{2})\Gamma(t+1)\Gamma(t+\tfrac{3}{2})/ \Gamma(t+\tfrac{1}{2}m)\Gamma(t+\tfrac{1}{2}m+\tfrac{1}{2})\Gamma(t+\tfrac{1}{2}m+1)\Gamma(t+\tfrac{1}{2}m+\tfrac{3}{2})]\,dt$$

where

$$C = \Gamma[\tfrac{1}{2}(m+n-3)]\Gamma[\tfrac{1}{2}(m+n-2)]\Gamma[\tfrac{1}{2}(m+n-1)]\Gamma[\tfrac{1}{2}(m+n)]/ \Gamma[\tfrac{1}{2}(n-3)]\Gamma[\tfrac{1}{2}(n-2)]\Gamma[\tfrac{1}{2}(n-1)]\Gamma(\tfrac{1}{2}n)$$

which, by the successive application of Legendre's duplication formula becomes

$$f(U_{4,m,n}) = [\Gamma(m+n-3)\Gamma(m+n-1)/\Gamma(n-3)\Gamma(n-1)]\cdot U^{(n-5)/2} \cdot(1/2\pi i)\int_{c-i\infty}^{c+i\infty} U^{-t}[\Gamma(2t)\Gamma(2t+2)/\Gamma(2t+m)\Gamma(2t+m+2)]\cdot dt.$$

By applying the modified form of Consul's inverse Mellin transform (2.3) to the above integral, we find that the exact distribution of $U_{4,m,n}$ is given by

$$f(U_{4,m,n}) = [\Gamma(m+n-3)\Gamma(m+n-1)/\Gamma(n-3)\Gamma(n-1)\cdot 2\Gamma(2m)] \cdot U^{(n-5)/2}(1-U^{\frac{1}{2}})^{2m-1}F(m-2, m; 2m; 1-U^{\frac{1}{2}}) \tag{3.3.0}$$

where $0 \leqq U \leqq 1$.

The result (3.3.0) clearly proves that for all values of m the exact probability distribution of $U_{4,m,n}$ is in the form of Gauss hypergeometric functions. However, for particular values of m, the hypergeometric function in the distribution of $U_{4,m,n}$ can be expressed in terms of algebraic functions by first reducing the values of the parameters with the help of Consul's formula (2.6) and then by replacing the resulting hypergeometric functions with their known values by the formulae (2.7) or (2.8).

The cumulative distribution function Pr $(U_{4,m,n} \leqq u)$ can be obtained, for different values of m, by integrating the resulting algebraic forms of the exact distribution of $U_{4,m,n}$ with respect to U between the limits 0 to u ($\leqq 1$).

The algebraic forms of the exact distribution and cumulative distribution functions of $U_{4,m,n}$ are being listed below, for convenience, for some special values of m viz. $m = 3, 4, 5, 6, 7$ and 8.

CASE I. For $m = 3$, the exact and cumulative distribution functions of $U_{4,3,n}$ become

$$f(U_{4,3,n}) = [\Gamma(n)\Gamma(n+2)/\Gamma(n-3)\Gamma(n-1)\times 96] \cdot U^{(n-5)/2}[1-U^2-8U^{\frac{1}{2}}(1-U)-6U\log U] \tag{3.3.1}$$

where $0 \leqq U \leqq 1$, and

$$\Pr(U_{4,3,n} \leqq u) = [\Gamma(n)\Gamma(n+2)/\Gamma(n-3)\Gamma(n-1) \times 48] \cdot u^{(n-3)/2}[1/(n-3) - 8u^{\frac{1}{2}}/(n-2) + 8u^{\frac{3}{2}}/n - u^2/(n+1) + 12u/(n-1)^2 - (6u/(n-1)) \log u] \tag{3.3.2}$$

The expression (3.3.2) is the same as that obtained by Anderson (1958) by another method.

CASE II. For $m = 4$, the algebraic forms of the exact and cumulative distributions of $U_{4,4,n}$ are respectively given by

$$f(U_{4,4,n}) = C \cdot U^{(n-5)/2}[1 - 15U^{\frac{1}{2}} - 80U + 80U^{\frac{3}{2}} + 15U^2 - U^{\frac{5}{2}} - 30(U + U^{\frac{3}{2}}) \log U] \tag{3.3.3}$$

where $0 \leqq U \leqq 1$, and C is given by $C = \Gamma(n+1)\Gamma(n+3)/\Gamma(n-3) \cdot \Gamma(n-1) \cdot 6! \cdot 2$ and

$$\Pr(U_{4,4,n} \leqq u) = 2C \cdot U^{(n-3)/2}[1/(n-3) - 15U^{\frac{1}{2}}/(n-2) - 80u/(n-1) + 80u^{\frac{3}{2}}/n + 15u^2/(n+1) - u^{\frac{5}{2}}/(n+2) + 60u/(n-1)^2 + 60u^{\frac{3}{2}}/n^2 - \{30u/(n-1) + 30u^{\frac{3}{2}}/n\} \log u]. \tag{3.3.4}$$

The expression (3.3.4) is much more simpler than the result obtained by Wilks (1935) and Anderson (1958). We have shown in Appendix that Anderson's result can be simplified to (3.3.4).

CASE III. For $m = 5$, the exact distribution and the cumulative distribution functions of $U_{4,5,n}$ can be respectively expressed in the form

$$f(U_{4,5,n}) = C \cdot \tfrac{1}{2} U^{(n-5)/2}[1 - 24U^{\frac{1}{2}} - 375U + 375U^2 + 24U^{\frac{5}{2}} - U^3 - 30(3U + 8U^{\frac{3}{2}} + 3U^2) \log U] \tag{3.3.5}$$

where $0 \leqq U \leqq 1$, and the constant C is given by $C = \Gamma(n+2)\Gamma(n+4)/\Gamma(n-3)\Gamma(n-1) \cdot 4!\,6!$ and

$$\Pr(U_{4,5,n} \leqq u) = C \cdot u^{(n-3)/2}[1/(n-3) - 24u^{\frac{1}{2}}/(n-2) - 375u/(n-1) + 375u^2/(n+1) + 24u^{\frac{5}{2}}/(n+2) - u^3/(n+3) + 180u/(n-1)^2 + 480u^{\frac{3}{2}}/n^2 + 180u^2/(n+1)^2 - \{90u/(n-1) + 240u^{\frac{3}{2}}/n + 90u^2/(n+1)\} \log u]. \tag{3.3.6}$$

CASE IV. For $m = 6$, the exact and cumulative distribution functions of $U_{4,6,n}$ respectively become

$$f(U_{4,6,n}) = (C/2)U^{(n-5)/2}[1 - 35U^{\frac{1}{2}} - 1099U - 1575U^{\frac{3}{2}} + 1575U^2 + 1099U^{\frac{5}{2}} + 35U^3 - U^{7/2} - 210(U + 5U^{\frac{3}{2}} + 5U^2 + U^{\frac{5}{2}}) \log U] \tag{3.3.7}$$

and

$$\Pr(U_{4,6,n} \leqq u) = C\cdot u^{(n-3)/2}[1/(n-3) - 35u^{\frac{1}{2}}/(n-2) - 1099u/(n-1) - 1575u^{\frac{3}{2}}/n + 1575u^2/(n+1) + 1099u^{\frac{5}{2}}/(n+2) + 35u^3/(n+3) - u^{7/2}/(n+4) + 420\{u/(n-1)^2 + 5u^{\frac{3}{2}}/n^2 + 5u^2/(n+1)^2 + u^{\frac{5}{2}}/(n+2)^2\} - 210\,\{u/(n-1) + 5u^{\frac{3}{2}}/n + 5u^2/(n+1) + u^{\frac{5}{2}}/(n+2)\}\log u] \tag{3.3.8}$$

where $0 \leqq U \leqq u \leqq 1$, and the constant $C = \Gamma(n+3)\Gamma(n+5)/\Gamma(n-3)\cdot \Gamma(n-1)\cdot 5!\,7!$.

CASE V. For $m = 7$, the algebraic forms of the exact and cumulative distribution functions of $U_{4,7,n}$ respectively are

$$f(U_{4,7,n}) = (\tfrac{1}{2}C)\cdot U^{\frac{1}{2}(n-5)}\cdot[1 - U^4 - 48U^{\frac{1}{2}}(1-U^3) - 2548U(1-U^2) - 8624U^{\frac{3}{2}}(1-U) - 420(U + 8U^{\frac{3}{2}} + 15U^2 + 8U^{\frac{5}{2}} + U^3)\log U] \tag{3.3.9}$$

where $0 \leqq U \leqq 1$, and $C = \Gamma(n+4)\Gamma(n+6)/\Gamma(n-3)\Gamma(n-1)\cdot 6!\,8!$ and

$$\Pr(U_{4,7,n} \leqq u) = C\cdot u^{\frac{1}{2}(n-3)}[1/(n-3) - 48u^{\frac{1}{2}}/(n-2) - 2548u/(n-1) - 8624u^{\frac{3}{2}}/n + 8624u^{\frac{5}{2}}/(n+2) + 2548u^3/(n+3) + 48u^{7/2}/(n+4) - u^4/(n+5) - 420\{u/(n-1) + 8u^{\frac{3}{2}}/n + 15u^2/(n+1) + 8u^{\frac{5}{2}}(n+2) + u^3/(n+3)\}\log u + 840\{u/(n-1)^2 + 8u^{\frac{3}{2}}/n^2 + 15u^2/(n+1)^2 + 8u^{\frac{5}{2}}/(n+2)^2 + u^3/(n+3)^2\}]. \tag{3.3.10}$$

CASE VI. For $m = 8$, the exact and cumulative distribution functions of $U_{4,8,n}$ take the respective forms given below:

$$f(U_{4,8,n}) = (\tfrac{1}{2}C)\cdot U^{\frac{1}{2}(n-5)}[1 - U^{9/2} - 63(U^{\frac{1}{2}} - U^4) - (5104 + \tfrac{4}{5})\cdot(U - U^{7/2}) - 29988(U^{\frac{3}{2}} - U^3) - 28244(U^2 - U^{\frac{5}{2}}) - 252(3U + 35U^{\frac{3}{2}} + 105U^2 + 105U^{\frac{5}{2}} + 35U^3 + 3U^{7/2})\log U] \tag{3.3.11}$$

where $0 \leqq U \leqq 1$, and $C = \Gamma(n+5)\Gamma(n+7)/\Gamma(n-3)\Gamma(n-1)\cdot 7!\,9!$ and

$$\Pr(U_{4,8,n} \leqq u) = C\cdot u^{\frac{1}{2}(n-3)}\cdot[1/(n-3) - 63u^{\frac{1}{2}}/(n-2) - (25524u/5(n-1)) - 29988u^{\frac{3}{2}}/n$$

$$
\begin{aligned}
&+ 29988u^{3}/(n+3) - 28224u^{2}/(n+1) \\
&+ 28224u^{\frac{5}{2}}/(n+2) + 25524u^{7/2}/5(n+4) \\
&+ 63u^{4}/(n+5) - u^{9/2}/(n+6) \qquad (3.3.12) \\
&+ 504\{3u/(n-1)^2 + 35u^{\frac{3}{2}}/n^2 + 105u^2/(n+1)^2 \\
&+ 105u^{\frac{5}{2}}/(n+2)^2 + 35u^3/(n+3)^2 \\
&+ 3u^{7/2}/(n+4)^2\} - 252\{3u/(n-1) \\
&+ 35u^{\frac{3}{2}}/n + 105u^2/(n+1) + 105u^{\frac{5}{2}}/(n+2) \\
&+ 35u^3/(n+3) + 3u^{7/2}/(n+4)\} \log u].
\end{aligned}
$$

3.4 *Exact distributions of $U_{p,3,n}$ and $U_{p,4,n}$ for $p = 5, 6, 7, 8$.* Wilks (1935) has already shown that, when the hypothesis is true, the exact distribution of $U_{p,m,n}$ is the same as that of $U_{m,p,n+m-p}$. Using this theorem we easily get the exact distributions and the cumulative distributions of $U_{p,3,n}$ and $U_{p,4,n}$ for $p = 5, 6, 7, 8$ from $U_{3,m,n}$ and $U_{4,m,n}$ for $m = 5, 6, 7, 8$.

Acknowledgment. I thank the editors and the referees for drawing my attention to the recent work of Schatzoff on this subject and for their valuable suggestions for improving the utility of the paper.

APPENDIX

Cumulative distribution of $U_{3,3,n}$.

(i) $-2\sin^{-1}(1-u)^{\frac{1}{2}} = \sin^{-1}\{-2u^{\frac{1}{2}}(1-u)^{\frac{1}{2}}\} = \sin^{-1}[-\{1-(2u-1)^2\}^{\frac{1}{2}}]$

$= \sin^{-1}(2u-1) - \sin^{-1}(1) = \arcsin(2u-1) - \frac{1}{2}\pi;$

(ii)
$$
\begin{aligned}
I_u(\tfrac{1}{2}n, \tfrac{1}{2}) &= B^{-1}(\tfrac{1}{2}n, \tfrac{1}{2}) \cdot \textstyle\int_0^u t^{n/2-1}(1-t)^{-\frac{1}{2}}\, dt \\
&= B^{-1}(\tfrac{1}{2}n, \tfrac{1}{2})[-2u^{n/2-1}(1-u)^{\frac{1}{2}} \\
&\quad + (n-2) \textstyle\int_0^u t^{n/2-2}(1-t)^{\frac{1}{2}}\, dt] \\
&= -2B^{-1}(\tfrac{1}{2}n, \tfrac{1}{2}) \cdot u^{n/2-1}(1-u)^{\frac{1}{2}} + I_u(\tfrac{1}{2}n - 1, \tfrac{3}{2});
\end{aligned}
$$

(iii)
$$
\begin{aligned}
I_u(\tfrac{1}{2}n - 1, \tfrac{5}{2}) &= B^{-1}(\tfrac{1}{2}n - 1, \tfrac{5}{2}) \textstyle\int_0^u t^{n/2-2}(1-t)^{\frac{3}{2}} \cdot dt \\
&= B^{-1}(\tfrac{1}{2}n - 1, \tfrac{5}{2})(1/(n+1))[2u^{n/2-1}(1-u)^{\frac{3}{2}} \\
&\quad + 3 \textstyle\int_0^u t^{n/2-2}(1-t)^{\frac{1}{2}} \cdot dt] \\
&= \{\Gamma[\tfrac{1}{2}(n+1)]/\Gamma[\tfrac{1}{2}(n-2)] \cdot \Gamma(\tfrac{5}{2})\} \cdot u^{n/2-1}(1-u)^{\frac{3}{2}} \\
&\quad + I_u(\tfrac{1}{2}n - 1, \tfrac{3}{2}).
\end{aligned}
$$

By putting the values from (i), (ii) and (iii) in the expression (3.2.2), we get the cumulative distribution of $U_{3,3,n}$ in the same form as obtained by Anderson (1958).

Cumulative distribution of $U_{4,4,n}$. The different terms of Anderson's result are equivalent to

$$\text{1st term} = B^{-1}(n-1, 4) \int_0^{u^{\frac{1}{2}}} t^{n-2}(1-t)^3\, dt = [Cu^{(n-3)/2}/(n-3)_4]$$
$$[6u/(n-1) - 8u^{\frac{3}{2}}/n + 18u^2/(n+1) - 6u^{\frac{5}{2}}/(n+2)],$$

$$\text{2nd term} = [Cu^{(n-3)/2}/20(n-3)][1 - 10u + 20u^{\frac{3}{2}} - 15u^2 + 4u^{\frac{5}{2}}],$$

$$\text{3rd term} = -[Cu^{(n-3)/2}/20(n-2)][15u^{\frac{1}{2}} - 60u + 90u^{\frac{3}{2}} - 60u^2 + 15u^{\frac{5}{2}}],$$

$$\text{4th term} = -[3C\cdot u^{(n-1)/2}/2(n-1)] \log u + [Cu^{(n-3)/2}/20(n-1)][-110u$$
$$+ 180u^{\frac{3}{2}} - 90u^2 + 20u^{\frac{5}{2}}],$$

$$\text{5th term} = -[3Cu^{n/2}/2n] \log u - [C\cdot u^{(n-3)/2}/20n][20u + 30u^{\frac{3}{2}} - 60u^2$$
$$+ 10u^{\frac{5}{2}}],$$

where $C = \Gamma(n+1)\Gamma(n+3)/\{36\Gamma(n-3)\Gamma(n-1)\}$.

By adding together all the terms, we get the same expression as was obtained by us in (3.3.4). Thus Anderson's result can be simplified into the one obtained by us.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] BARTLETT, M. S. (1934). The vector representation of a sample. *Proc. Cambridge Phil. Soc.* **30** 327–340.

[3] BARTLETT, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Cambridge Phil. Soc.* **34** 33–40.

[4] BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–346.

[5] CONSUL, P. C. (1965). The exact distribution of certain likelihood criteria useful in multivariate analysis. *Bull. Class. des Sc.* **51** 449–457.

[6] CONSUL, P. C. (1966). On some inverse Mellin transforms. *Bull. Class. des Sc.* **52** 547–561.

[7] CONSUL, P. C. (1966). On some reduction formulae for hypergeometric functions. *Bull. Class. des Sc.* **52** 562–576.

[8] RAO, C. R. (1951). An asymptotic expansion of the distribution of Wilks Λ-criteria. *Bull. Inst. Internat. Statist.* **33** 177–180.

[9] SCHATZOFF, M. (1964). Exact distribution of Wilks' likelihood ratio criterion and comparisons with competitive tests. Ph.D. Thesis, Harvard University; Abstract, *Ann. Math. Statist.* **35** 1397.

[10] TITCHMARSH, E. C. (1948). *Introduction to the Theory of Fourier Integrals.* Oxford Univ. Press, London.

[11] WALD, A. and BROOKNER, R. J. (1941). On the distribution of Wilks statistic for testing independence of several groups of variables. *Ann. Math. Statist.* **12** 137–152.

[12] WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24** 471–494.

[13] WILKS, S. S. (1934). Moment generating operators for determinants of product moments in samples from a normal system. *Ann. Math.* **35** 312–340.

[14] WILKS, S. S. (1935). On the independence of *k* sets of normally distributed statistical variables. *Econometrica* **3** 309–326.