ON THE DISTRIBUTION OF WILKS' STATISTIC FOR TESTING THE INDEPENDENCE OF SEVERAL GROUPS OF VARIATES

By A. Wald[1] and R. J. Brookner[1]

Columbia University

1. **Introduction.** We consider p variates $x_1, x_2, \cdots, x_p$ which have a joint normal distribution. Let the variates be divided into k groups; group one containing $x_1, x_2, \cdots, x_{p_1}$, group two containing $x_{p_1+1}, x_{p_1+2}, \cdots, x_{p_2}$, etc. We are interested in testing the hypothesis that the set of all population correlation coefficients between any two variates which belong to different groups is zero

Wilks[2] has derived, by using the Neyman-Pearson likelihood ratio criterion, a statistic based on N independent observations on each variate with which one may test this hypothesis. Let $\| r_{ij} \|$ be the matrix of sample correlation coefficients; Wilks' statistic, λ, is the ratio of the determinant of the p-rowed matrix of sample correlations to the product of the p_1-rowed determinant of correlations of the variates of group one, the $(p_2 - p_1)$-rowed determinant of correlations of the second group, etc. That is

$$\lambda = \frac{| r_{ij} |}{| r_{\alpha_1\beta_1} | \cdot | r_{\alpha_2\beta_2} | \cdots | r_{\alpha_k\beta_k} |}$$

where $| r_{\alpha_i\beta_i} |$ is the principal minor of $| r_{ij} |$ corresponding to the ith group.

In order to use the test, the distribution function of λ must be known. Wilks has shown that in certain cases the exact distribution is a simple elementary function; in other cases it is an elementary function, but one which is rather unwieldy and which does not lend itself readily to practical use. It is our purpose in this paper (1) to show a method by which the exact distribution can be explicitly given as an elementary function for a certain class of groupings of the variates, and (2) to give an expansion of the exact cumulative distribution function in an infinite series which is applicable to any grouping.

2. **The exact distribution of λ.** By the method to be described, the exact distribution of λ can be found when the numbers of variates in the groups are such that there are an odd number in at most one group. If the number of variates is small, say at most eight, the method will increase only slightly the list of distribution functions that Wilks gives in his paper.

---

[1] Research under a grant-in-aid of the Carnegie Corporation of New York.

[2] S. S. Wilks, "On the independence of k sets of normally distributed statistical variables," *Econometrica*, Vol. 3 (1935), pp. 309-326. Other references to Wilks in this paper except where otherwise noted are to this publication.

For purposes of deriving the distribution of λ we may assume that $E(x_u) = 0$, $(u = 1, 2, \cdots, p)$; that there are $n = N - 1$ independent observations $x_{u\alpha}$ $(\alpha = 1, 2, \cdots, n)$ on each variate x_u; and that the sample covariance between x_i and x_j is given by $s_{ij} = \sum_{\alpha=1}^{n} x_{i\alpha}x_{j\alpha}/n$. We define u' (a function of u) to be the total number of variables in all the groups which precede the group in which x_u lies. The complete theory is independent of the ordering of the groups and of the ordering of the variates within the groups; hence without loss of generality, we may assume that if any group contains an odd number of variates, it will be the last group, hence u' is always an even integer.

Wilks has shown that λ is a product $\prod_{u=p_1+1}^{p} z_u$ where each z_u is distributed independently of the others, and that the distribution of z_u is

$$\frac{z_u^{\frac{1}{2}(n-u-1)}(1 - z_u)^{\frac{1}{2}(u'-2)}}{B[\frac{1}{2}(n - u + 1), u'/2]} dz_u . \tag{1}$$

Now let $y_u = \log z_u$, then the characteristic function of y_u is

$$\begin{aligned}\phi_u(t) &= \frac{1}{B[\frac{1}{2}(n - u + 1), u'/2]} \int_0^1 e^{t \log z_u} z_u^{\frac{1}{2}(n-u-1)}(1 - z_u)^{\frac{1}{2}(u'-2)} dz_u \\ &= \frac{1}{B[\frac{1}{2}(n - u + 1), u'/2]} \int_0^1 z_u^{\frac{1}{2}(n-u-1)+t}(1 - z_u)^{\frac{1}{2}(u'-2)} dz_u\end{aligned}$$

where t is a pure imaginary. It is known[3] that this integral, even with complex exponents, is the Beta-function so long as the real parts of both exponents are greater than minus one, so

$$\begin{aligned}\phi_u(t) &= \frac{B[\frac{1}{2}(n - u + 1) + t, u'/2]}{B[\frac{1}{2}(n - u + 1), u'/2]} \\ &= \frac{\Gamma[\frac{1}{2}(n - u + 1) + t\cdot]\Gamma[\frac{1}{2}(n - u + 1 + u')]}{\Gamma[\frac{1}{2}(n - u + 1 + u') + t]\cdot\Gamma[\frac{1}{2}(n - u + 1)]}.\end{aligned} \tag{2}$$

But here u' is always an even integer, hence by the well known recursion formula of the Gamma-function, which is valid for complex arguments excluding only negative integers

$$\phi_u(t) = c_u\{[\tfrac{1}{2}(n - u + 1) + t][\tfrac{1}{2}(n - u + 3) + t] \cdots [\tfrac{1}{2}(n - u + u' - 1) + t]\}^{-1}$$

where

$$c_u = [\tfrac{1}{2}(n - u + 1)][\tfrac{1}{2}(n - u + 3)] \cdots [\tfrac{1}{2}(n - u + u' - 1)].$$

[3] See Whittaker and Watson, *A Course in Modern Analysis*, Fourth edition 1927, Chap. 12.

Now set

$$y = \log \lambda = y_{p_1+1} + y_{p_1+2} + \cdots + y_p$$

and the characteristic function of y is

$$\phi(t) = \prod_{u=p_1+1}^{p} c_u\{[\tfrac{1}{2}(n - u + 1) + t][\tfrac{1}{2}(n - u + 3) + t] \cdots [\tfrac{1}{2}(n - u + u' - 1) + t]\}^{-1}.$$

From the characteristic function, we can obtain the distribution function, $g(y)$, of y by the relation

$$g(y) = \frac{c_n}{2\pi i} \int_{-i\infty}^{i\infty} \frac{e^{-yt}\, dt}{\prod_{u=p_1+1}^{p} [\frac{1}{2}(n - u + 1) + t] \cdots [\frac{1}{2}(n - u + u' - 1) + t]}$$

$$= \frac{c_n}{2\pi i} \int_{-i\infty}^{i\infty} \Phi(t)\, dt,$$

where

$$c_n = \prod_{u=p_1+1}^{p} c_u .$$

The integration can be carried out by the method of residues; since y is always negative (the range of λ is from 0 to 1), on a half circle with center at the origin in the negative half of the complex t-plane, the integral of the function $\Phi(t)$ converges to zero as the radius of the circle becomes infinite. Since $\Phi(t)$ is analytic except for a finite number of poles on the negative real axis, $g(y)$ is c_n times the sum of the residues at these points.

Now $\Phi(t)$ is of the form $\frac{e^{-yt}}{P(t)}$ where $P(t)$ is a polynomial in t as follows: suppose that the groups contain $r_1, r_2, \cdots, r_k$ variables respectively, then let $(k_j + 1)$ be the number of these r's which are greater than or equal to j; then

$$P(t) = [\tfrac{1}{2}(n - 2) + t]^{k_1}[\tfrac{1}{2}(n - 3) + t]^{k_2}[\tfrac{1}{2}(n - 4) + t]^{k_3+k_1}[\tfrac{1}{2}(n - 5) + t]^{k_4+k_2} [\tfrac{1}{2}(n - 6) + t]^{k_5+k_3+k_1} \cdots [\tfrac{1}{2}(n - p + 1) + t]^{k_{p-2}+k_{p-4}+\cdots+k_{[\frac{1}{2}p]-[\frac{1}{2}(p-3)]}}.$$

where

$$[\sigma/2] = \begin{array}{l} \sigma/2 \text{ if } \sigma \text{ is even} \\ (\sigma - 1)/2 \text{ if } \sigma \text{ is odd.} \end{array}$$

Then

$$g(y; r_1, r_2, \cdots, r_k) = c_n \sum_{\alpha=1}^{p-2} \frac{1}{\theta_\alpha!} \frac{d^{\theta_\alpha}}{dt^{\theta_\alpha}} [(t + \tfrac{1}{2}(n - \alpha - 1))^{\theta_\alpha+1}\Phi(t)]_{t=-\frac{1}{2}(n-\alpha-1)}$$

where

$$\theta_\alpha + 1 = k_\alpha + k_{\alpha-2} + \cdots + k_{[\frac{1}{2}(\alpha+2)]-[\frac{1}{2}(\alpha-1)]} .$$

It can be shown that θ_α is $\geqq 0$ for α between 1 and $p - 2$. Thus we have $g(y; r_1, r_2, \cdots, r_k)$ and from it we can calculate $f(\lambda; r_1, r_2, \cdots, r_k)$.

Suppose $p = 8$ and that the variables are divided into two groups of four each, then we will calculate the distribution function $f(\lambda; 4, 4)$. Now

$$g(y; 4, 4) = \frac{c_n}{2\pi i} \int_{-i\infty}^{i\infty} \frac{e^{-yt}\,dt}{[\frac{1}{2}(n-2)+t][\frac{1}{2}(n-3)+t][\frac{1}{2}(n-4)+t]^2 \cdot [\frac{1}{2}(n-5)+t]^2[\frac{1}{2}(n-6)+t][\frac{1}{2}(n-7)+t]}$$

and

$$c_n = \left(\frac{n-2}{2}\right)\left(\frac{n-3}{2}\right)\left(\frac{n-4}{2}\right)^2\left(\frac{n-5}{2}\right)^2\left(\frac{n-6}{2}\right)\left(\frac{n-7}{2}\right).$$

Then

$$g(y; 4, 4) = 16c_n\left[\frac{-e^{\frac{1}{2}(n-2)y}}{90} + e^{\frac{1}{2}(n-3)y} + \frac{8e^{\frac{1}{2}(n-4)y}}{9} - \frac{8e^{\frac{1}{2}(n-5)y}}{9} - e^{\frac{1}{2}(n-6)y} + \frac{e^{\frac{1}{2}(n-7)y}}{90} - \frac{ye^{\frac{1}{2}(n-4)y}}{3} + \frac{ye^{\frac{1}{2}(n-5)y}}{3}\right].$$

Since

$$y = \log\lambda, \qquad dy = \frac{d\lambda}{\lambda},$$

we have

$$f(\lambda; 4, 4,) = \frac{16c_n}{3}\left[-\frac{\lambda^{\frac{1}{2}(n-4)}}{30} + \frac{\lambda^{\frac{1}{2}(n-5)}}{2} - \frac{8\lambda^{\frac{1}{2}(n-6)}}{3} + \frac{8\lambda^{\frac{1}{2}(n-7)}}{3} - \frac{\lambda^{\frac{1}{2}(n-8)}}{2} + \frac{\lambda^{\frac{1}{2}(n-9)}}{30} - (\lambda^{\frac{1}{2}(n-7)} + \lambda^{\frac{1}{2}(n-6)})\log\lambda\right].$$

The cumulative distribution function is given by

$$\begin{aligned} J_w(4, 4) &= \text{Prob}\,[\lambda \leq w; 4, 4] \\ &= \frac{16c_n}{3} w^{\frac{1}{2}(n-7)}\left[\frac{1}{15(n-7)} - \frac{w^{\frac{1}{2}}}{n-6} - \frac{4(4n-23)w}{3(n-5)^2} + \frac{14(4n-13)w^{\frac{3}{2}}}{3(n-4)^2} + \frac{w^2}{n-3} - \frac{w^{\frac{5}{2}}}{15(n-2)} - \left(\frac{2w}{n-5} + \frac{2w^{\frac{3}{2}}}{n-4}\right)\log w\right]. \end{aligned}$$

Wilks' expression for the cumulative distribution function appears to be quite different, but if we substitute $n = N - 1$ and use the relation

$$\begin{aligned} \beta_{\sqrt{w}}(N-6; 4) &= \frac{\Gamma(N-2)}{\Gamma(N-6)\cdot\Gamma(4)} \int_0^{\sqrt{w}} x^{N-7}(1-x)^3\,dx \\ &= \tfrac{1}{6}(n-2)(n-3)(n-4)(n-5) \\ &\quad \cdot\left[\frac{w^{\frac{1}{2}(n-5)}}{n-5} - \frac{3w^{\frac{1}{2}(n-4)}}{n-4} + \frac{3w^{\frac{1}{2}(n-3)}}{n-3} - \frac{w^{\frac{1}{2}(n-2)}}{n-2}\right] \end{aligned}$$

it can be shown that the two formulas for the cumulative distribution are identical.

In cases where u' is not always an even integer, the exact distribution function of λ can still be obtained using this method. However, in such a case, the gamma functions do not cancel out and the integrand has an infinitude of poles, so the function is expressed by an infinite series. We will use a different method to obtain an infinite series expansion.

3. A series expansion of the cumulative distribution function. Let us put $v = -y$, and let the density function of v be $h(v)$, then from (2), we have

$$h(v)\,dv = dv\,\frac{c_n}{2\pi i}\int_{-i\infty}^{i\infty} e^{vt} \prod_{u=r_1+1}^{p} \frac{\Gamma[\frac{1}{2}(n-u+1)+t]\,dt}{\Gamma[\frac{1}{2}(n-u+1+u')+t]}.$$

Since v is a monotonic decreasing function of λ, and since the critical region for testing the null hypothesis is given by the inequality $\lambda < \lambda_0$, then the critical region will be defined by $v > v_0$, where v_0 is such that

$$\int_{v_0}^{\infty} h(v)\,dv$$

is equal to a chosen level of significance.

PROPOSITION 1.

$$h(v) = h_n(v)\bar{\psi}(v)$$

where $\bar{\psi}(v)$ does not depend on n, and $h_n(v) = c_n e^{-\frac{1}{2}v}$.

PROOF: Let

$$t' = t + \tfrac{1}{2}(n-p).$$

Then

$$h(v) = \frac{c_n}{2\pi i}\int_{-i\infty+\frac{1}{2}(n-p)}^{i\infty+\frac{1}{2}(n-p)} e^{v(t'-\frac{1}{2}(n-p))} \prod_{u} \frac{\Gamma[\frac{1}{2}(p-u+1)+t']\,dt'}{\Gamma[\frac{1}{2}(p-u+u'+1)+t']}.$$

Now the area in the complex plane bounded by the vertical line through $\frac{1}{2}(n-p)$, by the vertical line through the origin, and by arcs of a circle with center at the origin of arbitrary radius is one in which the integrand is everywhere regular. Furthermore, the integral along the arcs approaches zero as the radius of the circle approaches infinity, hence the integrals along the vertical line through $\frac{1}{2}(n-p)$ and along the vertical axis are equal. Then we may write

$$\begin{aligned}\frac{e^{\frac{1}{2}v}}{c_n}h(v) &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{v(t'+p/2)} \prod_{u} \frac{\Gamma[\frac{1}{2}(p-u+1)+t']\,dt'}{\Gamma[\frac{1}{2}(p-u+u'+1)+t']}\\ &= \bar{\psi}(v).\end{aligned}$$

Therefore

$$h(v) = c_n e^{-\frac{1}{2}v}\bar{\psi}(v).$$

PROPOSITION 2.

$$I = \lim_{n\to\infty} \int_0^\infty \frac{c_n e^{-\frac{n}{2}v} v^{r-1}\, dv}{\Gamma(r)} = 1$$

where we define

$$r = \sum_{j=i+1}^{k} \sum_{i=1}^{k-1} \frac{r_i r_j}{2}$$

so that

$$\begin{aligned} r &= \tfrac{1}{2}[r_2 r_1 + r_3(r_1 + r_2) + \cdots + r_k(r_1 + r_2 + \cdots + r_{k-1})] \\ &= \tfrac{1}{2} \sum_u u'. \end{aligned}$$

PROOF: Let

$$\frac{n}{2} v = v^*$$

then

$$\begin{aligned} \int_0^\infty c_n e^{-\frac{n}{2}v} v^{r-1}\, dv &= \int_0^\infty c_n e^{-v^*} \left(\frac{2}{n}\right)^r (v^*)^{r-1}\, dv^* \\ &= c_n \left(\frac{2}{n}\right)^r \Gamma(r). \end{aligned}$$

Hence

$$I = \lim_{n\to\infty} c_n \left(\frac{2}{n}\right)^r$$

but

$$c_n = \prod_u \frac{\Gamma\frac{1}{2}(n - u + 1 + u')}{\Gamma\frac{1}{2}(n - u + 1)}$$

and therefore

$$I_u = \lim_{n\to\infty} \frac{\Gamma\frac{1}{2}(n - u + 1 + u')}{\Gamma\frac{1}{2}(n - u + 1)} \left(\frac{2}{n}\right)^{u'/2} = 1$$

by an application of the Stirling approximation. Therefore

$$I = \prod_u I_u = 1.$$

We then write

$$\psi(v) = \frac{\bar{\psi}(v)\Gamma(r)}{v^{r-1}}$$

hence

$$h(v) = \frac{c_n e^{-\frac{1}{2}v} v^{r-1} \psi(v)}{\Gamma(r)}. \tag{3}$$

PROPOSITION 3. *For any positive integer s,*

$$\lim_{n\to\infty} \left\{ n^s \cdot \text{Prob}\left(v > \frac{1}{\sqrt{n}} \right) \right\} = 0.$$

PROOF: Since $v = -\log \lambda$, the inequality $v > 1/\sqrt{n}$ is equivalent to the inequality $\lambda < e^{-1/\sqrt{n}}$. Since $\lambda = \prod_{u=p_1+1}^{p} z_u$, the inequality $\lambda < e^{-1/\sqrt{n}}$ implies that there exists at least one value of u for which

$$z_u < e^{-1/(p-p_1)\sqrt{n}}.$$

Hence

$$\sum_{u=p_1+1}^{p} P(z_u < e^{-1/(p-p_1)\sqrt{n}}) \geq P(\lambda < e^{-1/\sqrt{n}}) = P(v > 1/\sqrt{n}).$$

Hence in order to prove Proposition 3 we have only to show that for each u and any arbitrary positive integer s

$$\lim_{n\to\infty} \{n^s \cdot P(z_u < e^{-1/(p-p_1)\sqrt{n}})\} = 0.$$

From (1) we have

$$P(z_u < e^{-1/(p-p_1)\sqrt{n}})$$

$$= \frac{1}{B[\frac{1}{2}(n-u+1); u'/2]} \int_0^{e^{-1/(p-p_1)\sqrt{n}}} z_u^{\frac{1}{2}(n-u-1)} (1-z_u)^{\frac{1}{2}(u'-2)}\, dz_u.$$

Over the range of integration, we have $z_u \leq e^{-1/(p-p_1)\sqrt{n}}$ so

$$P(z_u < e^{-1/(p-p_1)\sqrt{n}}) \leq \frac{e^{\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{B[\frac{1}{2}(n-u+1); u'/2]} \int_0^{e^{-1/(p-p_1)\sqrt{n}}} (1-z_u)^{\frac{1}{2}(u'-2)}\, dz_u$$

$$= \frac{e^{-\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{B[\frac{1}{2}(n-u+1); u'/2]} \left[-\frac{2}{u'} (1-z_u)^{u'/2} \right]_0^{e^{-1/(p-p_1)\sqrt{n}}}$$

$$= \frac{2e^{-\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{u' \cdot B[\frac{1}{2}(n-u+1); u'/2]} [1 - (1 - e^{-1/(p-p_1)\sqrt{n}})^{u'/2}].$$

It follows from the Stirling formula that

$$\lim_{n\to\infty} \left(\frac{n}{2}\right)^{u'/2} B[\tfrac{1}{2}(n-u+1); u'/2] = \lim_{n\to\infty} \frac{\Gamma\frac{1}{2}(n-u+1)\Gamma(u'/2)}{\Gamma\frac{1}{2}(n-u+u'+1)} \left(\frac{n}{2}\right)^{u'/2}$$

$$= \Gamma(u'/2).$$

Since

$$\lim_{n\to\infty} n^{\frac{1}{2}u'+s} e^{-\sqrt{n}/2(p-p_1)} = 0$$

and

$$\lim_{n\to\infty} (1 - (1 - e^{-1/\sqrt{n}})) = 1,$$

the proposition follows.

PROPOSITION 4. *The function $\psi(v)$ of formula* (3) *can be expanded in a power series, i.e.*

$$\psi(v) = \alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots$$

with a finite radius of convergence.

PROOF: Wilks[4] has considered the following integral equation:

$$\int_0^{\beta} w^t g(w)\, dw = CB^t \frac{\Gamma(b_1 + t)\cdot\Gamma(b_2 + t) \cdots \Gamma(b_q + t)}{\Gamma(c_1 + t)\cdot\Gamma(c_2 + t) \cdots \Gamma(c_q + t)},$$

where $C = \dfrac{\Gamma(c_1)\cdot\Gamma(c_2) \cdots \Gamma(c_q)}{\Gamma(b_1)\cdot\Gamma(b_2) \cdots \Gamma(b_q)}$, B and $g(w)$ are independent of t, and $b_i < c_i$ $(i = 1, 2, \cdots, q)$. Wilks has shown that the solution of the integral equation, $g(w)$, is given by the following expression:

$$
\begin{aligned}
g(w) = {} & \frac{kw^{b_q-1}\left(1 - \frac{w}{B}\right)^{\gamma_q-\beta_q-1}}{B^{b_q}} \int_0^1 \int_0^1 \cdots \int_0^1 v_1^{c_1-b_1-1} v_2^{c_2-b_2-1} \cdots v_{q-1}^{c_{q-1}-b_{q-1}-1} \\
& \times (1 - v_1)^{\gamma_{q-1}-\beta_{q-1}-1} (1 - v_2)^{\gamma_{q-2}-\beta_{q-2}-1} \cdots (1 - v_{q-1})^{\gamma_1-\beta_1-1} \\
& \times \left[1 - v_1\left(1 - \frac{w}{B}\right)\right]^{b_1-c_2} \left[1 - \{v_1 + v_2(1 - v_1)\}\left(1 - \frac{w}{B}\right)\right]^{b_2-c_3} \cdots \\
& \times \Big[1 - \{v_1 + v_2(1 - v_1) + \cdots \\
& \qquad + v_{q-1}(1 - v_1)(1 - v_2) \cdots (1 - v_{q-2})\}\left(1 - \frac{w}{B}\right)\Big]^{b_{q-1}-c_q} \\
& \times dv_1\, dv_2 \cdots dv_{q-1}
\end{aligned}
\tag{4}
$$

where

$$k = \prod_{i=1}^{q} \frac{\Gamma(c_i)}{\Gamma(b_i)\Gamma(c_i - b_i)}$$

and

$$\gamma_i = \sum_{j=0}^{i-1} c_{q-j} \qquad \beta_i = \sum_{j=0}^{i-1} b_{q-j}$$

---

[4] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 474–5.

the range of w being $0 \leqq w \leqq B$. Wilks has furthermore shown that

$$\{v_1 + v_2(1 - v_1) + \cdots + v_i(1 - v_1)(1 - v_2) \cdots (1 - v_{i-1})\}\left(1 - \frac{w}{B}\right) < 1 \tag{5}$$

for $w > 0$ and $0 \leqq v_i \leqq 1$ $(i = 1, 2, \cdots, q - 1)$.

We denote the left hand side of (5) by ζ_i. The factor $(1 - \zeta_i)^{b_i - c_{i+1}}$ can be expanded in a power series, i.e.

$$\begin{aligned}(1 - \zeta_i)^{b_i - c_{i+1}} &= (1 - \zeta_i)^{-(c_{i+1} - b_i)} \\ &= 1 + (c_{i+1} - b_i)\zeta_i + \tfrac{1}{2}(c_{i+1} - b_i)(c_{i+1} - b_i + 1)\zeta_i^2 + \cdots\end{aligned} \tag{6}$$

with a radius of convergence equal to one. Since we will show shortly that for the choices we make for the b_i's and c_i's, $c_{i+1} \geq b_i$, then all coefficients in this last expansion are non-negative. Substituting this series expansion (6) in (4), and ordering it according to powers of $(1 - w/B)$, the expression under the integral sign (in 4) becomes

$$\begin{aligned}&\theta_0(v_1, v_2, \cdots v_{q-1}) \\ &\quad + \theta_1(v_1, \cdots v_{q-1})\left(1 - \frac{w}{B}\right) + \theta_2(v_1, \cdots, v_{q-1})\left(1 - \frac{w}{B}\right)^2 + \cdots.\end{aligned} \tag{7}$$

This series is uniformly convergent over the domain defined by the inequalities $0 \leqq v_i \leqq 1$ $(i = 1, 2, \cdots, q - 1)$ and $|1 - w/B| < 1$. We can even say that (7) is uniformly convergent for $|1 - w/B| < 1$ if we substitute for each θ_i the maximum of θ_i with respect to $v_1, v_2, \cdots, v_{q-1}$. Hence we may integrate the series (7) with respect to $v_1, v_2, \cdots v_{q-1}$ term by term, i.e.

$$\int_0^1 \int_0^1 \cdots \int_0^1 (7)\, dv_1\, dv_2 \cdots dv_{q-1} = \sigma_0 + \sigma_1\left(1 - \frac{w}{B}\right) + \sigma_2\left(1 - \frac{w}{B}\right)^2 + \cdots \tag{8}$$

and the series (8) is uniformly convergent for $|1 - w/B| < 1$. The coefficients $\sigma_0, \sigma_1, \cdots$ are non-negative.

The case of the λ statistic which we are considering is a special case of this integral equation which we obtain by making the following substitutions:

$$w = \lambda, \quad B = 1, \quad u = r + p_1, \quad q = p - p_1$$

$$b_r = \tfrac{1}{2}(n - u + 1), \quad c_r = \tfrac{1}{2}(n - u + u' + 1), \quad (r = 1, 2, \cdots, p - p_1)$$

Note that then

$$c_{r+1} - b_r = \tfrac{1}{2}[(u + 1)' - 1] \geqq 0.$$

Hence, according to (4)

$$g(\lambda)\, d\lambda = k \cdot \lambda^{\frac{1}{2}(n-p-1)}(1 - \lambda)^{\frac{1}{2}\Sigma u' - 1}\{\sigma_0 + \sigma_1(1 - \lambda) + \sigma_2(1 - \lambda^2) + \cdots\}\, d\lambda$$

where the infinite series converges for $|1 - \lambda| < 1$.

Now $v = -\log \lambda$, or $\lambda = e^{-v}$, hence

$$h(v)\, dv = k \cdot e^{-\frac{1}{2}(n-p+1)v} v^{r-1}\left(\frac{1 - e^{-v}}{v}\right)^{r-1}\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}\, dv$$

where the series $\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$ is obtained from the series $\{\sigma_0 + \sigma_1(1 - \lambda) + \cdots\}$ by substituting for $(1 - \lambda)$ the Taylor expansion of $(1 - e^{-v})$. The series $\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$ has a finite radius of convergence.[5]

Hence the function $\psi(v)$ can be written as

$$\psi(v) = A \cdot e^{\frac{1}{2}(p-1)v}\left(\frac{1 - e^{-v}}{v}\right)^{r-1}\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$$

where A denotes a constant factor. Then since $e^{\frac{1}{2}(p-1)v}\left(\frac{1 - e^{-v}}{v}\right)^{r-1}$ can be expanded in a Taylor series around $v = 0$, Proposition 4 is proved.

4. **Evaluation of the coefficients in the expansion of $\psi(v)$.** Let the series expansion of $\psi(v)$ be

$$\psi(v) = \alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots$$

Then we have

$$\int_0^\infty \frac{c_n e^{-\frac{1}{2}nv} v^{r-1}}{\Gamma(r)} (\alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots)\, dv \equiv 1.$$

Now let $v^* = \frac{n}{2} v$, then

$$\int_0^\infty \left(\frac{2}{n}\right)^r \frac{c_n e^{-v^*} v^{*^{r-1}}}{\Gamma(r)} \left(\alpha_0 + \frac{2\alpha_1 v^*}{n} + \frac{4\alpha_2 v^{*^2}}{n^2} + \cdots\right) dv^* \equiv 1.$$

Suppose that the asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$ is given by

$$\beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots.$$

On account of Proposition 3, we have that the asymptotic expansion in powers of $1/n$ of

$$\int_0^{\sqrt{n}} \frac{e^{-v^*} v^{*^{r-1}}}{\Gamma(r)} \left(\alpha_0 + \frac{2\alpha_1}{n} v^* + \frac{4\alpha_2}{n^2} v^{*^2} + \cdots\right) dv^* \tag{9}$$

must be equal to the asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$. Since we may integrate in (9) term by term for sufficiently large n, we easily obtain

$$\alpha_0 = \beta_0, \qquad \alpha_1 = \frac{\beta_1}{2r}, \cdots \alpha_k = \frac{\beta_k}{2^k \cdot r(r+1) \cdots (r+k-1)}.$$

[5] See A. Gutzmer, *Theorie der Eindeutigen Analytischen Funktionen*, 1906, pp. 91–2.

The asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$ can be calculated in the following manner:

$$\left(\frac{n+2}{n}\right)^r \frac{c_n}{c_{n+2}} = \frac{\beta_0 + \frac{\beta_1}{n+2} + \frac{\beta_2}{(n+2)^2} + \cdots}{\beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots}$$

and

$$\left(\frac{n+2}{n}\right)^r \frac{c_n}{c_{n+2}} = (1 + 2/n)^r \prod_u \frac{n - u + 1}{n - u + u' + 1}.$$

Equating the right hand members of these last two equations, and taking logs, we obtain

$$\log\left[\beta_0 + \frac{\beta_1}{n+2} + \frac{\beta_2}{(n+2)^2} + \cdots\right] = r \log(1 + 2/n) + \sum_u \log\left(1 - \frac{u-1}{n}\right) - \sum_u \log\left(1 - \frac{u - u' - 1}{n}\right) + \log\left(\beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots\right).$$

Then we expand each term in a series of powers of $1/n$ and equate coefficients of $1/n^i$ for each i. We obtain the following formulae for the first five β's:

$$\beta_0 = 1$$

$$\beta_1 = r + \tfrac{1}{4}\sum_u (u-1)^2 - \tfrac{1}{4}\sum_u (u - u' - 1)^2$$

$$\beta_2 = \beta_1 + \frac{\beta_1^2}{2} - \frac{2r}{3} + \frac{1}{12}\sum_u (u-1)^3 - \frac{1}{12}\sum_u (u - u' - 1)^3$$

$$\beta_3 = -\tfrac{4}{3}\beta_1 - \beta_1^2 - \tfrac{1}{3}\beta_1^3 + \beta_1\beta_2 + 2\beta_2 + \tfrac{2}{3}r + \tfrac{1}{24}\sum_u (u-1)^4 - \tfrac{1}{24}\sum_u (u - u' - 1)^4$$

$$\beta_4 = 2\beta_1 + 2\beta_1^2 + \beta_1^3 + \frac{\beta_1^4}{4} - 3\beta_1\beta_2 + \beta_1\beta_3 - \beta_1^2\beta_2 - 4\beta_2 + \frac{\beta_2^2}{2} + 3\beta_3 - \frac{4}{5}r + \frac{1}{40}\sum_u (u-1)^5 - \frac{1}{40}\sum_u (u - u' - 1)^5.$$

5. **Practical use of the series.** In practical applications, the value of the statistic, say λ_0, is calculated, and it is desired that we determine whether or not this value of the statistic falls into the critical region. That is, for a particular grouping of the variates, for a particular number of degrees of freedom, and for a chosen level of significance α, there is determined from the distribution of λ, a value λ^* such that

$$\text{Prob}\,[\lambda < \lambda^*] = \alpha,$$

and if $\lambda_0 < \lambda^*$ we reject the hypothesis that in the population from which the sample is taken all the correlation coefficients between variates in different groups are zero.

Since v is a monotonic decreasing function of λ we make the test by computing $v_0 = -\log \lambda_0$ and we reject the hypothesis if $v_0 > v^*$ where $v^* = -\log \lambda^*$. But this is equivalent to computing Prob $[v > v_0]$ and if this value is less than α we reject the hypothesis. Now

$$\begin{aligned} \text{Prob } [v > v_0] &= J_{v_0}(r_1, r_2, \cdots, r_k) \\ &= \frac{C_n}{\Gamma(r)} \int_{v_0}^{\infty} e^{-\frac{1}{2}nv} v^{r-1}(1 + \alpha_1 v + \alpha_2 v^2 + \cdots)\, dv. \end{aligned}$$

Setting $\frac{nv}{2} = z$

$$\text{Prob } [v > v_0] = \left(\frac{2}{n}\right)^r \frac{C_n}{\Gamma(r)} \int_{nv_0/2}^{\infty} e^{-z} z^{r-1}\left[1 + \alpha_1 \frac{2z}{n} + \alpha_2 \left(\frac{2}{n}\right)^2 z^2 + \cdots\right] dz.$$

On account of Proposition 3 we obtain an asymptotic expansion of Prob $[v > v_0]$ by integrating the right hand member of the above equation term by term. This can be expressed by means of the incomplete gamma function, which is tabulated[6] in the form

$$I(u, p) = \frac{\int_0^{u\sqrt{p+1}} v^p e^{-v}\, dv}{\Gamma(p + 1)}.$$

We obtain

$$\begin{aligned} \text{Prob } [v > v_0] = \left(\frac{2}{n}\right)^r c_n \Bigg\{ &\left[1 - I\left(\frac{nv_0}{2\sqrt{r}}, r - 1\right)\right] \\ &+ \frac{\beta_1}{n}\left[1 - I\left(\frac{nv_0}{2\sqrt{r+1}}, r\right)\right] + \frac{\beta_2}{n^2}\left[1 - I\left(\frac{nv_0}{2\sqrt{r+2}}, r + 1\right)\right] + \cdots \Bigg\}. \end{aligned}$$

The values of the constant $K = \left(\frac{2}{n}\right)^r c_n$ and the values of $\beta_1, \beta_2, \beta_3, \beta_4$ are herein tabulated for any grouping which might be made on six or fewer variates. Some cases, such as groupings $(1, p - 1)$, in which case the distribution of λ is the distribution of the multiple correlation coefficient; and as the groupings $(2, p - 2)$, the exact distribution for which was given by Wilks as an incomplete Beta-function, are superfluous here. These cases are included only for the sake of completeness.

---

[6] K. Pearson (Editor), *Tables of the Incomplete Gamma Function*, Biometric Laboratory, London, 1922.

Table of the First Four β's

Grouping	r	β_1	β_2	β_3	β_4
2,1	1	2	4	8	16
1,1,1	1.5	2.75	6.28125	13.38281	27.57568
3,1	1.5	3.75	12.03125	36.91406	111.55225
2,2	2	5	19	65	211
2,1,1	2.5	5.75	23.53125	83.97656	279.50538
1,1,1,1	3	6.5	28.625	106.9375	366.39844
4,1	2	6	28	120	496
3,2	3	9	55	285	1351
3,1,1	3.5	9.75	62.53125	334.10156	1615.91163
2,2,1	4	11	77	439	2229
2,1,1,1	4.5	11.75	86.03125	506.16406	2628.23974
1,1,1,1,1	5	12.5	95.625	580.6875	3085.52344
5,1	2.5	8.75	55.78125	315.82031	1690.65282
4,2	4	14	125	910	5901
3,3	4.5	15.75	154.03125	1205.03906	8277.55226
4,1,1	4.5	14.75	136.28125	1015.50781	6693.45068
3,2,1	5.5	17.75	189.53125	1584.10156	11445.75538
2,2,2	6	19	214	1866	13947
3,1,1,1	6	18.5	203.625	1740.9375	12797.27344
2,2,1,1	6.5	19.75	229.03125	2042.16406	15530.08351
2,1,1,1,1	7	20.5	244.625	2230.1875	17257.64836
1,1,1,1,1,1	7.5	21.25	260.78125	2430.49219	19139.02892

Tables of the Constant $K = \left(\frac{2}{n}\right)^r C_n$

n	21	111	31	22	211	1111	41	311
10	.800	.738	.646	.560	.517	.477	.480	.310
11	.818	.761	.676	.595	.553	.515	.521	.352
12	.833	.780	.702	.625	.585	.548	.556	.390
13	.846	.796	.724	.651	.612	.576	.586	.424
14	.857	.810	.743	.674	.637	.602	.612	.455
15	.867	.822	.759	.693	.658	.624	.636	.482
16	.875	.833	.774	.711	.677	.645	.656	.508
17	.882	.843	.787	.727	.694	.663	.675	.531
18	.889	.851	.798	.741	.709	.679	.691	.552
19	.895	.859	.808	.754	.723	.694	.706	.571
20	.900	.866	.818	.765	.736	.708	.720	.589
22	.909	.878	.834	.785	.758	.732	.744	.620
24	.917	.888	.847	.802	.777	.752	.764	.647
26	.923	.896	.859	.817	.793	.770	.781	.671
28	.929	.903	.869	.829	.807	.785	.796	.691
30	.933	.910	.877	.840	.819	.798	.809	.710
35	.943	.922	.894	.862	.843	.825	.835	.747
40	.950	.932	.908	.879	.862	.846	.855	.776
45	.956	.940	.918	.892	.877	.862	.871	.799
50	.960	.946	.926	.902	.889	.875	.883	.818
55	.964	.950	.932	.911	.899	.886	.894	.833
60	.967	.954	.938	.918	.907	.895	.902	.846
65	.969	.958	.943	.924	.914	.903	.910	.858
70	.971	.961	.947	.930	.920	.910	.916	.867
80	.975	.966	.953	.938	.930	.921	.926	.883
90	.978	.970	.959	.945	.937	.929	.934	.896
100	.980	.973	.963	.951	.943	.936	.941	.906

Tables of the Constant K (ii)

n	221	2111	32	11111	51	42	33
10	.269	.248	.336	.229	.323	.168	.136
11	.310	.288	.379	.268	.369	.206	.171
12	.347	.325	.417	.304	.410	.243	.205
13	.381	.359	.451	.338	.445	.277	.237
14	.412	.390	.481	.368	.478	.309	.268
15	.441	.418	.508	.397	.506	.339	.297
16	.467	.444	.533	.423	.532	.367	.324
17	.490	.468	.556	.447	.555	.392	.350
18	.512	.490	.576	.470	.576	.416	.374
19	.532	.511	.595	.490	.596	.438	.396
20	.551	.530	.612	.510	.613	.459	.417
22	.584	.564	.642	.544	.644	.496	.455
24	.613	.593	.668	.575	.671	.529	.489
26	.638	.619	.691	.601	.694	.558	.519
28	.660	.642	.711	.625	.714	.584	.546
30	.680	.662	.728	.646	.731	.607	.570
35	.720	.704	.764	.689	.767	.654	.621
40	.751	.737	.791	.723	.794	.692	.661
45	.776	.763	.813	.751	.816	.722	.694
50	.797	.785	.830	.773	.833	.747	.721
55	.814	.803	.845	.792	.848	.768	.743
60	.828	.818	.857	.808	.860	.786	.762
65	.841	.831	.868	.822	.870	.801	.779
70	.852	.842	.877	.833	.879	.814	.793
80	.869	.861	.892	.853	.894	.836	.817
90	.883	.876	.903	.869	.905	.853	.836
100	.894	.888	.913	.881	.915	.867	.852

Tables of the Constant K (iii)

n	411	321	222	3111	2211	21111	111111
10	.155	.108	.094	.100	.087	.080	.076
11	.192	.140	.123	.130	.114	.106	.099
12	.228	.171	.152	.160	.142	.133	.125
13	.261	.201	.180	.189	.170	.160	.150
14	.292	.230	.208	.217	.197	.186	.176
15	.322	.257	.235	.244	.223	.212	.201
16	.349	.284	.261	.270	.248	.236	.225
17	.375	.309	.285	.295	.272	.260	.248
18	.398	.332	.308	.318	.295	.283	.271
19	.421	.354	.330	.340	.317	.304	.292
20	.442	.375	.351	.361	.338	.325	.313
22	.479	.414	.390	.400	.376	.363	.351
24	.512	.448	.424	.434	.411	.398	.385
26	.542	.479	.456	.465	.442	.430	.417
28	.568	.507	.484	.493	.471	.458	.446
30	.591	.532	.510	.519	.497	.484	.472
35	.640	.585	.564	.573	.552	.540	.528
40	.679	.628	.608	.616	.597	.585	.574
45	.710	.663	.644	.652	.633	.623	.612
50	.736	.692	.674	.681	.664	.654	.644
55	.758	.716	.700	.706	.690	.681	.671
60	.776	.737	.722	.728	.712	.704	.695
65	.792	.755	.740	.746	.732	.723	.715
70	.805	.771	.757	.762	.749	.741	.733
80	.828	.797	.784	.789	.777	.770	.762
90	.846	.818	.806	.811	.800	.793	.786
100	.860	.835	.824	.828	.818	.812	.806