The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1491-1497

ON THE TEST OF INDEPENDENCE BETWEEN TWO SETS OF VARIATES[1]

BY R. J. MUIRHEAD

Yale University

1. Summary. In this paper an asymptotic expansion is derived for the power function of the likelihood ratio criterion for testing independence between two sets of variates for the case when the population canonical correlation coefficients are small. The method used can theoretically give the expansion up to any order of N where N is the sample size. Here the expansion is given up to N^{-3} and is an extension of an expansion obtained independently by Sugiura [10] using a different method. The theorem in Section 3 summarizes the final result. In Section 4 the expansion is compared numerically with a different approximation obtained by Sugiura and Fujikoshi [11] and with exact results obtained by Pillai and Jayachandran [9].

2. Introduction. The moments of the likelihood ratio criterion λ for testing the independence between p and q sets of variates ($p \leqq q$) in an s-variate ($s = p + q$) normal population have been expressed, under alternative hypothesis, by Sugiura and Fujikoshi [11] as

$$E[\gamma^{(2/N)h}] = \frac{\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}(n+2h-q))}{\Gamma_p(\frac{1}{2}(n+2h))\Gamma_p(\frac{1}{2}(n-q))} \det(I-P^2)^{\frac{1}{2}n} \times {}_2F_1(\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}(n+2h); P^2) \tag{2.1}$$

where

$$\Gamma_p(a) = \pi^{p(p-1)/4} \textstyle\prod_{i=1}^{p} \Gamma(a - \frac{1}{2}(i-1)),$$

N is the sample size, $n = N - 1$, $P^2 = \operatorname{diag}(\rho_1^2, \rho_2^2, \cdots, \rho_p^2)$ with $\rho_1, \cdots, \rho_p$ the population canonical correlation coefficients, and ${}_2F_1$ is the hypergeometric function of matrix argument (see Constantine [2]). When $P^2 \neq 0$ we can consider two limiting distributions.

(i) Olkin and Siotani [8] have shown that $\lambda^{2/N}$ is asymptotically normal with mean $\det(I - P^2)$ and variance $(4/N)(\operatorname{tr} P^2)\det(I - P^2)^2$, from which it follows that the limiting distribution of the statistic

$$-\left(\frac{N}{4 \operatorname{tr} P^2}\right)^{\frac{1}{2}} \{\ln \lambda^{2/N} - \ln \det(I - P^2)\}$$

is that of a standard normal variate. Using this, Sugiura and Fujikoshi [11] have obtained an expansion up to order N^{-1} of the power function in terms of the standard normal distribution function and its derivatives.

Received January 12, 1970; revised January 26, 1971.

[1] This research was supported by a C.S.I.R.O. Post-graduate Studentship at the University of Adelaide, South Australia.

(ii) Assume that $\Omega = NP^2$ is a fixed matrix, i.e., we assume that $\rho_1^2, \cdots, \rho_p^2$ are so small that in effect $P^2 \to 0$ as $N \to \infty$. Then Sugiura [10] has shown that the limiting distribution of $-2 \log \lambda$ is noncentral χ^2 on pq degrees of freedom and noncentrality parameter $\operatorname{tr} \Omega$. We will obtain an expansion up to order N^{-3} of the power function in terms of noncentral χ^2 distribution functions.

It is of interest to compare the limiting behaviors given in (i) and (ii) when $p = 1$. Let R and $\bar{R}$ denote the sample and population multiple correlation coefficients respectively. Wishart [12] has shown that the distribution of R^2 is asymptotically normal with mean $\bar{R}^2$ and variance $4\bar{R}^2(1 - \bar{R}^2)^2/N$ (this corresponds to case (i)), and Fisher [3] has shown that if $\omega = N\bar{R}^2$ is fixed then the limiting distribution of NR^2 is noncentral χ^2 on q degrees of freedom and noncentrality parameter ω (case (ii)).

3. The asymptotic expansion of the distribution. We now consider the asymptotic distribution of the modified test statistic $-2\rho \ln \lambda$ under the alternative hypothesis $P^2 = m^{-1}\Omega$, where $m = \rho N = n - \delta$ with $\delta = \frac{1}{2}(p + q + 1)$ (see Anderson [1]). Substituting in (2.1) yields an expression for the moment generating function (mgf) $g(t, \Omega)$ of $-2\rho \ln \lambda$ as

$$
\begin{aligned}
g(t, \Omega) &= \frac{\Gamma_p(\frac{1}{2}(m + \delta))\Gamma_p(\frac{1}{2}m(1 + 2t) + \frac{1}{2}(\delta - q))}{\Gamma_p(\frac{1}{2}m(1 + 2t) + \frac{1}{2}\delta)\Gamma_p(\frac{1}{2}(m + \delta - q))} \\
&\quad \times \det\left(I - \frac{1}{m}\Omega\right)^{\frac{1}{2}(m+\delta)} \\
&\quad \times {}_2F_1\left(\tfrac{1}{2}(m + \delta), \tfrac{1}{2}(m + \delta); \tfrac{1}{2}m(1 + 2t) + \tfrac{1}{2}\delta; \frac{1}{m}\Omega\right) \\
&= g_1(t) \cdot g_2(t, \Omega)
\end{aligned} \tag{3.1}
$$

$g_1(t)$ denotes the mgf under the null hypothesis ($P^2 = 0$) given by the term in (3.1) involving the Γ_p functions, and has the expansion (Anderson [1] page 239)

$$
g_1(t) = (1 + 2t)^{\frac{1}{2}pq}\left\{1 + \frac{\gamma}{m^2}[(1 + 2t)^{-2} - 1] + O(m^{-4})\right\} \tag{3.2}
$$

where $\gamma = pq(p^2 + q^2 - 5)/48$. $g_2(t, \Omega)$ denotes the product of the determinant term and the ${}_2F_1$ function in (3.1) and may be expanded for large m using partial differential equations (pde's). The method outlined here is essentially the same as that used by Muirhead [7]. A system of pde's satisfied by the ${}_2F_1$ function is known (see [6], Equation 1.3). From this we may easily derive a system of pde's satisfied by $g_2(t, \Omega)$. Since

$$
g_2(t, \Omega) \to \operatorname{etr}\left(\frac{-t}{1 + 2t}\Omega\right)
$$

as $m \to \infty$ we may extract this limit to write

$$
g(t, \Omega) = \operatorname{etr}\left(-\frac{t}{1 + 2t}\Omega\right) \cdot e^{H(t,\Omega)}. \tag{3.3}
$$

The system of pde's satisfied by $H(t, \Omega)$ is readily found to be

$$
\begin{aligned}
&2\omega_i\left(1-\frac{1}{m}\omega_i\right)^2\left\{\frac{\partial^2 H}{\partial\omega_i^2}+\left(\frac{\partial H}{\partial\omega_i}\right)^2\right\}\\
&\quad+\left\{m(1+2t)+\delta-p+1-\omega_i\left[1+2t-\frac{4t}{1+2t}+\frac{1}{m}(\delta-2p+4)\right]\right.\\
&\quad\left.+\,\omega_i^2\left[\frac{1}{m}\left(\frac{8t}{1+2t}\right)+\frac{1}{m^2}(3-p)\right]-\frac{4t\omega_i^3}{m^2(1+2t)}\right.\\
&\quad\left.+\sum_{j=1,j\neq i}^{p}\frac{\omega_i(1-\omega_i/m)^2}{\omega_i-\omega_j}\right\}\frac{\partial H}{\partial\omega_i}-\sum_{j=1,j\neq i}^{p}\frac{\omega_j(1-\omega_j/m)(1-\omega_i/m)}{\omega_i-\omega_j}\frac{\partial H}{\partial\omega_j}\\
&=-\frac{2t^2\delta}{1+2t}-\omega_i\left[\frac{2t^2}{(1+2t)^2}+t+\frac{1}{m}\left(\frac{t\delta}{1+2t}+\frac{2t}{1+2t}\right)\right]\\
&\quad+\omega_i^2\left[\frac{4t^2}{m(1+2t)^2}+\frac{2t}{m^2(1+2t)}\right]-\frac{2t^2\omega_i^3}{m^2(1+2t)^2}\\
&\quad-\frac{t}{1+2t}\left(\frac{1}{m}-\frac{\omega_i}{m^2}\right)\sum_{j=1,j\neq i}^{p}\omega_j \qquad (i=1,2,\cdots,p)
\end{aligned}
\tag{3.4}
$$

where $\omega_1, \omega_2, \cdots, \omega_p$ are the latent roots of Ω.

We now look for a solution of (3.4) of the form

$$H(t, \Omega) \sim \sum_{k=1}^{\infty} m^{-k} Q_k(t, \Omega) \tag{3.5}$$

where $Q_k(t, 0) = 0$ for all k, so that $g_2(t, \Omega) = 1$. We substitute the series (3.5) into the system (3.4) and equate like powers of m^{-1} on both sides. For example, equating the constant terms on both sides of (3.4) we obtain the pde's

$$\frac{\partial Q_1}{\partial\omega_i} = \frac{-2t^2\delta}{(1+2t)^2} - \omega_i\left[\frac{2t^2}{(1+2t)^3} + \frac{t}{1+2t}\right] \qquad (i = 1, 2, \cdots, p)$$

so that upon integrating and using the fact that the Q_k are symmetric functions of $\omega_1, \cdots, \omega_p$, we obtain

$$Q_1 = -\frac{2t^2\delta}{(1+2t)^2}\sigma_1 - \left[\frac{2t^2}{(1+2t)^3} + \frac{t}{1+2t}\right]\frac{\sigma_2}{2}$$

where $\sigma_r = \omega_1^r + \cdots + \omega_p^r$. Similarly, equating coefficients of m^{-1}, and of m^{-2}, and integrating gives us Q_2 and Q_3 as polynomials in the latent roots of Ω of degrees 3 and 4 respectively. The coefficients of higher powers of m^{-1} may also be obtained in a similar manner if required. The expansion for $\ln g(t, \Omega)$ is obtained by adding in the terms for $\ln g_1(t)$ obtained from (3.2). Exponentiating now gives an expansion for the mgf $g(t, \Omega)$ which may then be inverted term by term to give $\Pr(-2\rho \ln \lambda > x)$ in terms of noncentral χ^2 distributions. The final result is given in the

THEOREM. *The non-null distribution of the likelihood ratio criterion λ can, in the case when $\Omega = mP^2$ is a fixed matrix, be approximated asymptotically up to the* *order* m^{-3} *by*

$$
\begin{aligned}
&\Pr(-2\rho \ln \lambda > x) \\
(3.6) \qquad &= \Pr(\chi^2_{pq}(\sigma_1) > \kappa) + \frac{1}{2m} \textstyle\sum_{j=1}^{3} a_j \, \phi_{pq+2j}(x) \\
&\quad + \frac{1}{48m^2} \textstyle\sum_{j=1}^{6} b_j \, \phi_{pq+2j}(x) + \frac{1}{192m^3} \textstyle\sum_{j=1} c_j \, \phi_{pq+2j}(x) + O(m^{-4})
\end{aligned}
$$

where

$$
\begin{aligned}
a_1 &= \sigma_2 + 2\delta\sigma_1 , \qquad a_2 = \sigma_2 - 2\delta\sigma_1 , \qquad a_3 = -\sigma_2 , \\
b_1 &= -(3\sigma_2^2 + \sigma_1\sigma_2 - 16\sigma_3 + 12\delta^2\sigma_1^2 - 24\delta\sigma_2 - 96\gamma) , \\
b_2 &= -(3\sigma_2^2 - \sigma_1\sigma_2 - 16\sigma_3 - 36\delta^2\sigma_1^2 - 24\delta\sigma_2 + 48\delta^2\sigma_1 - 96\gamma) , \\
b_3 &= 9\sigma_2^2 + 2\sigma_1\sigma_2 + 16\sigma_3 - 12(3\delta^2 + 2)\sigma_1^2 - 24(5\delta + 1)\sigma_2 , \\
b_4 &= 3\sigma_2^2 - 3\sigma_1\sigma_2 - 64\sigma_3 + 12(\delta^2 + 2)\sigma_1^2 + 24(3\delta + 1)\sigma_2 , \\
b_5 &= -(9\sigma_2^2 - \sigma_1\sigma_2 - 32\sigma_3) , \qquad b_6 = 3\sigma_2^2 , \\
c_1 &= \sigma_2^3 - 16\sigma_2\sigma_3 + 6\delta\sigma_1\sigma_2^2 + 12\delta^2\sigma_1^2\sigma_2 - 24\delta\sigma_2^2 - 32\delta\sigma_1\sigma_3 + 48\sigma_4 \\
&\quad + 8\delta^3\sigma_1^3 - 48\delta^2\sigma_1\sigma_2 + 64\delta\sigma_3 + 12(3 - 8\gamma)\sigma_2 - 192\delta\gamma\sigma_1 , \\
c_2 &= \sigma_2^3 - 16\sigma_2\sigma_3 - 6\delta\sigma_1\sigma_2^2 - 36\delta^2\sigma_1^2\sigma_2 - 24\delta\sigma_2^2 + 32\delta\sigma_1\sigma_3 + 48\sigma_4 \\
&\quad - 40\delta^3\sigma_1^3 + 96\delta^2\sigma_1\sigma_2 + 64\delta\sigma_3 + 96\delta^3\sigma_1^2 - 48(3 + 2\gamma)\sigma_2 + 192\delta\gamma\sigma_1 , \\
c_3 &= -(2\sigma_2^3 - 16\sigma_2\sigma_3 + 24\delta\sigma_1\sigma_2^2 - 12(\delta^2 + 2)\sigma_1^2\sigma_2 - 24(7\delta + 1)\sigma_2^2 - 48\sigma_4 \\
&\quad - 16\delta(5\delta^2 + 3)\sigma_1^3 - 48\delta(5\delta + 1)\sigma_1\sigma_2 - 64\delta\sigma_3 + 96\delta(3\delta^2 + 2)\sigma_1^2 \\
&\quad + 24(24\delta^2 + 8\delta - 9 + 8\gamma)\sigma_2 - 192\delta(\delta^2 + \gamma)\sigma_1) , \\
c_4 &= -(14\sigma_2^3 - 80\sigma_2\sigma_3 - 36\delta\sigma_1\sigma_2^2 - 24(3\delta^2 - 1)\sigma_1^2\sigma_2 + 24(2\delta + 1)\sigma_2^2 \\
&\quad - 160\delta\sigma_1\sigma_3 - 48\sigma_4 + 16\delta(5\delta^2 + 9)\sigma_1^3 + 48(16\delta^2 + 3\delta + 8)\sigma_1\sigma_2 \\
&\quad + 64(19\delta + 6)\sigma_3 - 96(3\delta^2 + 6\delta + 1)\sigma_1^2 \\
&\quad - 48(24\delta^2 + 12\delta - 3 + 2\gamma)\sigma_2 + 192\delta(\delta^2 + \gamma)\sigma_1) , \\
c_5 &= 28\sigma_2^3 - 16\sigma_2\sigma_3 + 12\delta\sigma_1\sigma_2^2 - 12(9\delta^2 + 4)\sigma_1^2\sigma_2 - 24(11\delta + 2)\sigma_2^2 \\
&\quad - 352\delta\sigma_1\sigma_3 + 624\sigma_4 + 8\delta(5\delta^2 + 18)\sigma_1^3 + 48(14\delta^2 + 3\delta + 16)\sigma_1\sigma_2 \\
&\quad + 128(13\delta + 6)\sigma_3 - 96(\delta^3 + 4\delta + 1)\sigma_1^2 - 12(48\delta^2 + 36\delta - 3 + 8\gamma)\sigma_2 , \\
c_6 &= -(14\sigma_2^3 + 144\sigma_2\sigma_3 + 48\delta\sigma_1\sigma_2^2 - 12(5\delta^2 + 6)\sigma_1^2\sigma_2 - 24(11\delta + 3)\sigma_2^2 \\
&\quad - 256\delta\sigma_1\sigma_3 - 720\sigma_4 + 8\delta(\delta^2 + 6)\sigma_1^3 + 48(4\delta^2 + \delta + 8)\sigma_1\sigma_2 \\
&\quad + 128(5\delta + 3)\sigma_3) , \\
c_7 &= -(4\sigma_2^3 - 128\sigma_2\sigma_3 - 30\delta\sigma_1\sigma_2^2 + 12(\delta^2 + 2)\sigma_1^2\sigma_2 + 24(3\delta + 1)\sigma_2^2 \\
&\quad + 64\delta\sigma_1\sigma_3 + 240\sigma_4) , \\
c_8 &= 5\sigma_2^3 - 32\sigma_2\sigma_3 - 6\delta\sigma_1\sigma_2^2 , \qquad c_9 = -\sigma_2^3 ,
\end{aligned}
$$

$\sigma_r = \omega_1^r + \cdots + \omega_p^r$ *where* $\omega_1, \cdots, \omega_p$ *are the latent roots of* Ω,

$$\delta = \tfrac{1}{2}(p + q + 1) , \qquad \gamma = \tfrac{1}{48}pq(p^2 + q^2 - 5) , \qquad m = \rho N = N - 1 - \delta ,$$

$\Pr(\chi_k^2(\sigma_1) > x)$ *denotes the probability that a noncentral* χ^2 *variate on* k *degrees of* *freedom and noncentrality parameter* σ_1 *exceeds* x *and* $\phi_k(x)$ *denotes the probability density function, evaluated at* x, *of a noncentral* χ^2 *variate on* k *degrees of freedom and noncentrality parameter* σ_1.

The first three terms in (3.6) give the expansion obtained by Sugiura [10].

4. Numerical comparisons. In this section the results of some numerical calculations are tabulated, when $p = 1$, in order to obtain an indication of the ranges of effectiveness of the expansion (3.6) and the expansion obtained by Sugiura and Fujikoshi [11] in terms of the normal distribution function. The entries in Table 1 are $\Pr(R^2 > y \mid \bar{R}^2)$ for the case $n - q = 40$, where y is the upper 5% point of the distribution of R^2 tabulated by Mijares [5]. For each pair $(q, \bar{R}^2)$, three values are given. The top value is exact (correct to four decimal places) obtained using an expression for the distribution function of R^2 given by Gurland [4] in terms of incomplete beta functions; the second value is obtained using (3.6),

TABLE 1
$p = 1$

$n - q = 40$					
q	2	3	5	10	15
$\bar{R}^2$ \ y	.1391	.1755	.2344	.3418	.4192
0.03	.1515	.1313	.1110	.0916	.0837
	.1517	.1309	.1102	.0901	.0813
	.2087	.1319	.0051	.0000	.0000
0.05	.2307	.1973	.1623	.1272	.1125
	.2303	.1967	.1611	.1246	.1083
	.2271	.2351	.0940	.0001	.0000
0.07	.3147	.2699	.2207	.1690	.1467
	.3142	.2691	.2193	.1655	.1406
	.3074	.2846	.2192	.0039	.0000
0.10	.4423	.3853	.3184	.2425	.2077
	.4418	.3845	.3169	.2380	.1992
	.4396	.3886	.3488	.0599	.0008
0.20	.7875	.7343	.6570	.5426	.4774
	.7897	.7370	.6603	.5437	.4711
	.7865	.7361	.6786	.6122	.2833
0.30	.9482	.9260	.8868	.8110	.7564
	.9544	.9349	.9007	.8323	.7758
	.9477	.9272	.8967	.9347	.9801
0.40	.9925	.9877	.9777	.9527	.9302
	.9993	.9989	.9986	.9961	.9868
	.9931	.9883	.9781	.9924	1.1754
0.50	.9994	.9989	.9978	.9940	.9900
	1.0043	1.0076	1.0161	1.0428	1.0667
	.9995	.9984	.9942	.9717	.9935

Entries are $\Pr(R^2 > y/\bar{R}^2)$. Top value is correct (to four decimal places). Second value is obtained using (3.6). Third value is obtained using the normal approximation in [11].

and the third value is obtained using the normal approximation. It can be seen that for small values of $\bar{R}^2$ the approximation (3.6) is more accurate, as we should expect, because this case is very close to the null distribution where the normal approximation becomes invalid. However for larger values of $\bar{R}^2$ the normal approximation becomes more accurate than (3.6). The values beneath the line drawn on Table 1 fall into this category. It can also be noted that both approximations lose accuracy as q increases, although the normal expansion appears to be far worse in this respect.

Finally, some calculations are given for the case $p = 2$ in Table 2. The top values in this table are exact powers of the likelihood ratio test given by Pillai and Jayachandran [9]. The second value is the power obtained using (3.6). The agreement here is also seen to be quite good.

TABLE 2
$p = 2$

		n	20	46	90
		q	7	13	7
ρ_1^2	ρ_2^2				
.0001	0		.0500	.0501	.0502
			.0496	.0498	.0502
.01	0		.0530	.0559	.0713
			.0518	.0552	.0712
.005	.005		.0530	.0559	.0713
			.0518	.0552	.0712
.05	.001		.0668	.0855	.2018
			.0623	.0821	.2009

Entries are the powers of the likelihood ratio test. Top value is from Pillai and Jayachandran [9]. Second value is obtained using (3.6).

REFERENCES

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

[2] CONSTANTINE, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

[3] FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Statist. Soc. Ser. A* **121** 654–673.

[4] GURLAND, J. (1968). A relatively simple form of the distribution of the multiple correlation coefficient. *J. Roy. Statist. Soc. Ser. B* **30** 276–283.

[5] MIJARES, T. A. (1964). Percentage points of the sum $V_1^{(s)}$ of s roots ($s = 1 - 50$). Univ. of the Philippines, Manila.

[6] MUIRHEAD, R. J. (1970a). Systems of partial differential equations for hypergeometric functions of matrix argument. *Ann. Math. Statist.* **41** 991–1001.

[7] MUIRHEAD, R. J. (1970b). Asymptotic distributions of some multivariate tests. *Ann. Math. Statist.* **41** 1002–1010.

[8] OLKIN, I. and SIOTANI, M. (1964). Asymptotic distributions of functions of a correlation matrix. Stanford Univ. Tech. Report No. 6.

[9] PILLAI, K. C. S. and JAYACHANDRAN, K. (1967). Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika* **54** 195–210.

[10] SUGIURA, N. (1970). Asymptotic non-null distributions of the likelihood ratio criteria for covariance matrix under local alternatives. To appear in *Ann. Math. Statist.*

[11] SUGIURA, N. and FUJIKOSHI, Y. (1969). Exact expansions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Statist.* **40** 942-952.

[12] WISHART, J. (1931). The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population. *Biometrika* **22** 353-361.

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520