NON-NULL DISTRIBUTION OF THE LIKELIHOOD-RATIO IN ANALYSIS OF DISPERSION

By J. ROY

Indian Statistical Institute, Calcutta

SUMMARY. In this paper the non-null distribution of Wilks' likelihood-ratio criterion for analysis of dispersion (multivariate analysis of variance) when the expectation-matrix is of unit rank is worked out in a form suitable for numerical evaluation of the power function when the deviation parameter is small and degrees of freedom for error moderately large. An illustrative table of the power function of the analysis of dispersion test at five percent level of significance is presented for $p = 1, 2, 3, 4$ variates when the degrees of freedom for the hypothesis are $m = 2, 3$ and the degrees of freedom for error are $n = 200$.

1. Introduction

It is well known that problems of analysis of dispersion (or, multivariate analysis of variance) can be reduced to the following form: The joint probability density function of the elements of the random matrices X of form $p \times n (n \geqslant p)$ and Y of form $p \times m$ is

$$(2\pi)^{-\frac{1}{2}p(n+m)} \; |\Sigma|^{-\frac{1}{2}(n+m)} \; \exp \left[-\tfrac{1}{2} \text{ tr } \Sigma^{-1}\{XX'+(Y-M)(Y-M)'\}\right] \quad \ldots \; (1.1)$$

where the matrix Σ of form $p \times p$ is positive-definite and unknown and the problem is to test the hypothesis H_0 that the matrix M of form $p \times m$ is a null-matrix, that is,

$$H_0: \;\; M = 0 \quad \ldots \; (1.2)$$

The likelihood-ratio statistic for testing this hypothesis is due to Wilks (1932) and can be put as

$$L = \frac{|XX'|}{|XX'+YY'|} \quad \ldots \; (1.3)$$

The sampling distribution of this statistic in the null case when H_0 is true has been investigated extensively; see Rao (1953) or Anderson (1957). In the null case, we shall say that L follows Wilks' distribution with degrees of freedom n, m and p and denote its probability density function by $W(L, n, m, p)$. In this paper, we shall be concerned with the non-null distribution of L when $M \neq 0$. The following properties of the null-distribution of L would be useful in this connection:

$$E(L^t | H_0) = \prod_{i=0}^{p-1} \frac{\left(\frac{n-i}{2}\right)_t}{\left(\frac{n+m-1-i}{2}\right)_t} \quad \ldots \; (1.4)$$

where $$(a)_t = \Gamma(a+t)/\Gamma(a).$$

Under H_0, the statistic L is distributed independently of the elements of $(XX'+YY')$. $\quad \ldots \; (1.5)$

The non-null distribution of L when $m = 1$ was derived by Bose and Roy (1938) and Hsu (1938) and the probability density function in this case can be written as :

$$\sum_{j=0}^{\infty} p_j(\tfrac{1}{2}\delta^2)\, B\left(L, \frac{n-p+1}{2}, \frac{p}{2}+j\right) \quad \ldots \quad (1.6)$$

where $p_j(\theta)$ is the Poisson probability function :

$$p_j(\theta) = e^{-\theta}\, \theta^j/j! \quad \ldots \quad (1.7)$$

and $B(L, r, s)$ is the Beta density function

$$B(L, r, s) = \frac{1}{B(r, s)} L^{r-1}(1-L)^{s-1} \quad \ldots \quad (1.8)$$

and $$\delta^2 = M'\Sigma^{-1}M. \quad \ldots \quad (1.9)$$

Anderson (1946) has shown that, in general, when the rank of the matrix M is q, $q \leqslant \min(p, m)$, the distribution of L can involve at most the q parameters $\delta_1, \delta_2, \ldots, \delta_q$ defined as the positive square-roots of the non-zero roots of the determinantal equation

$$|MM'-\delta^2\Sigma| = 0. \quad \ldots \quad (1.10)$$

He has also derived the moments of L when $t = 1, 2$.

2. A lemma

To obtain the non-null distribution of L we make the linear transformation

$$[X^* \vdots Y^*] = C_1B[X \vdots Y]C_2 \quad \ldots \quad (2.1)$$

where the matrix B of the form $p\times p$ satisfies

$$B\Sigma^{-1}B' = I \quad \ldots \quad (2.2)$$

and orthogonal matrices C_1 of the form $p\times p$ and C_2 of the form $m\times m$ are chosen to make

$$C_1\, BM\, C_2 = \begin{bmatrix} O_1 & \vdots & \begin{matrix} O_2 \\ \cdots\cdots \\ \Delta \end{matrix} \end{bmatrix} \quad \ldots \quad (2.3)$$

where O_1 and O_2 are null-matrices of the forms $p\times(m-q)$ and $(p-q)\times q$ respectively and Δ is a $q\times q$ diagonal matrix with diagonal elements $\delta_1, \delta_2, \ldots, \delta_q$ defined by (1.10). The existence of matrices C_1 and C_2 is proved by Deemer and Olkin (1951). Writing Y^* in the partitioned form

$$Y^* = [Y_1^* \vdots Y_2^*] \quad \ldots \quad (2.4)$$

where Y_1^* is of the form $p\times(m-q)$ and Y_2^* is of the form $p\times q$, it follows that

$$L = \frac{|X^*X^{*\prime}|}{|X^*X^{*\prime}+Y_1^*Y_1^{*\prime}+Y_2^*Y_2^{*\prime}|} \qquad \ldots \quad (2.5)$$

and that the joint probability density function of X^*, Y_1^*, Y_2^* is

$$(2\pi)^{-\frac{1}{2}p(n+m)} \exp\left[-\tfrac{1}{2}\operatorname{tr}\left\{X^*X^{*\prime}+Y_1^*Y_1^{*\prime}+\left(Y_2^*-\begin{bmatrix}O_2\\ \Delta\end{bmatrix}\right)\left(Y_2^*-\begin{bmatrix}O_2\\ \Delta\end{bmatrix}\right)'\right\}\right]. \qquad \ldots \quad (2.6)$$

Now let

$$L_1 = \frac{|X^*X^{*\prime}|}{|X^*X^{*\prime}+Y_1^*Y_1^{*\prime}|}, \quad L_2 = \frac{|Z_1Z_1'|}{|Z_1Z_1'+Z_2Z_2'|} \qquad \ldots \quad (2.7)$$

where $\quad Z_1 = [X^* \,\vdots\, Y_1^*], \qquad Z_2 = Y_2^* \qquad \ldots \quad (2.8)$

so that $$L = L_1 L_2. \qquad \ldots \quad (2.9)$$

Obviously L_1 follows Wilks' distribution with n, $m-q$ and p degrees of freedom and from (1.5) it is easy to see that the distribution of L_1 is independent of L_2. We thus have the

Lemma 2.1: *The probability distribution of the statistic L defined by* (1.3) *is the same as the product of two independent statistics L_1 and L_2 where L_1 follows Wilks' distribution with degrees of freedom n, $m-q$ and p and L_2 is defined as the determinantal ratio* (2.7) *where the joint probability density function of the elements of Z_1 of form $p\times(n+m-q)$ and Z_2 of form $p\times q$ is*

$$(2\pi)^{-\frac{1}{2}p(n+m)} \exp\left[-\tfrac{1}{2}\operatorname{tr}\{Z_1Z_1'+(Z_2-M_2)(Z_2-M_2)'\}\right] \qquad \ldots \quad (2.10)$$

where $\quad M_2 = \begin{bmatrix}O_2\\ \Delta\end{bmatrix}$.

The problem of deriving the distribution of L_2 is simpler than that of L because the value of m is now reduced to only q.

3. The case $q = 1$

In this case, the probability density function of L_2 is

$$\sum_{j=0}^{\infty} p_j(\tfrac{1}{2}\delta^2)B\left(L_2, \frac{n+m-p}{2}, \frac{p}{2}+j\right) \qquad \ldots \quad (3.1)$$

where $p_j(\theta)$ is defined by (1.7), and δ is the single parameter involved. Also, L_1 follows independently Wilks' distribution with degrees of freedom n, $m-1$ and p. Consequently, the probability that the product $L = L_1 L_2$ is less than a preassigned constant x, $0 < x < 1$, may be evaluated as

$$\text{Prob}\ (L < x) = \sum_{j=0}^{\infty} p_j(\tfrac{1}{2}\delta^2)P_j(x) \qquad \ldots \quad (3.2)$$

where

$$P_j(x) = \int_{L_1L_2<x} W(L_1, n, m-1, p)B\left(L_2, \frac{n+m-p}{2}, \frac{p}{2}+j\right)dL_1dL_2. \qquad \ldots \quad (3.3)$$

To evaluate $P_j(x)$ we note that the problem is the same as finding the cumulative distribution function of the statistic

$$L_{0j} = L_1 \,.\, L_{2j} \qquad \dots \quad (3.4)$$

where L_1 and L_{2j} are distributed independently,—L_1 having the probability density function $W(L_1, n, m-1, p)$ and L_{2j} having the probability density function $B\left(L_2, \frac{n+m-p}{2}, \frac{p}{2}+j\right)$. Consequently the t-th moment of L_{0j} is given by

$$E(L_{0j}^t) = E(L_{2j}^t)E(L_1^t) = \frac{\left(\frac{n}{2}+\frac{m-p}{2}\right)_t}{\left(\frac{n}{2}+j\right)_t} \cdot \prod_{i=0}^{p-1} \frac{\left(\frac{n}{2}-\frac{i}{2}\right)_t}{\left(\frac{n}{2}+\frac{m-1-i}{2}\right)_t} \qquad \dots \quad (3.5)$$

where $(a)_t$ is as defined in (1.4).

It is known, however, (Roy, 1951), that for any statistic V distributed in the interval (0, 1) with t-th moment about origin given by

$$E(V^t) = \prod_{j=1}^{s} \frac{\left(\frac{n}{2}+b_j\right)_t}{\left(\frac{n}{2}+c_j\right)_t} \qquad \dots \quad (3.6)$$

where $c_j > b_j$, $j = 1, 2, \dots, s$ are constants not involving n,

$$\text{Prob}\,(V < x) = Q_r(x^*) + \text{terms } O(n^{-2}) \qquad \dots \quad (3.7)$$

where $$x^* = -(n-\lambda) \log_e x, \quad r = 2r_1, \quad \lambda = (r_1 - r_2)/r_1 \qquad \dots \quad (3.8)$$

$$r_t = \sum_{j=1}^{s} (c_j^t - b_j^t) \quad t = 1, 2 \qquad \dots \quad (3.9)$$

and $Q_r(x)$ is the probability for a Chi-square with r degrees of freedom to exceed x, that is,

$$Q_r(x) = \int_x^\infty \frac{1}{2^{\frac{1}{2}r}\Gamma(\frac{1}{2}r)} e^{-\frac{1}{2}u} u^{\frac{1}{2}r-1} du. \qquad \dots \quad (3.10)$$

Thus $$P_j(x) = Q_{pm+2j}(X_j) + \text{ terms } O(n^{-2}) \qquad \dots \quad (3.11)$$

where $$X_j = -(n+m-\lambda_j) \log_e x \qquad \dots \quad (3.12)$$

$$\lambda_j = \frac{pm(p+m+1)+2j-4j^2}{2pm+2j}. \qquad \dots \quad (3.13)$$

Formulae (3.2) and (3.12) give a very convenient method for computation when n is large and δ small. Since

$$|\text{Prob}\,(L < x) - \sum_{j=0}^{k} p_j(\tfrac{1}{2}\delta^2)P_j(x)| < 1 - \sum_{j=0}^{k} p_j(\tfrac{1}{2}\delta^2) \qquad \dots \quad (3.14)$$

and when δ is small, even for comparatively small values of k the difference $1-\sum_{j=0}^{k} p_j(\frac{1}{2}\delta^2)$ is rather small, we need consider only a very few terms in the expansion (3.2).

If n is so large that terms $O(n^{-2})$ may be neglected, $P_j(x)$ can be computed from Tables of the incomplete Gamma function or from Hartley and Pearson's (1950) Tables of the Chi-square integral. The p_j's can be read off from Molina's (1943) Tables of Poisson probabilities.

4. Power function of analysis of dispersion test

Formula (3.2) can be used to evaluate the power function of the analysis of dispersion test when the alternative hypothesis is of rank one. To illustrate, consider the case where there are $p = 4$ variates, the degrees of freedom for error and hypotheses are $n = 200$ and $m = 2$ respectively, the level of significance is fixed at five percent ($\alpha = 0.05$) and the alternative hypothesis of unit rank specifies that $\delta^2 = 2$. Using Rao's (1951) approximation (see Rao, 1953) to the null distribution of L, namely that $-\left(n+\dfrac{m-p-1}{2}\right) \times \log_e L$ follows the Chi-square distribution with pm degrees of freedom, the lower five percent point of the distribution of L is found to be

$$x = 0.92486.$$

The power of the test is thus given by Prob ($L < x$). To compute this, we prepare first the following table of values of $pm+2j$, λ_j and X_j for $j = 0, 1, 2, 3, 4, 5$:

j	$pm+2j$	λ_j	X_j
0	8	3.5	15.5073
1	10	3.0	15.5464
2	12	2.3	15.6089
3	14	1.181818	15.6884
4	16	0	15.7807
5	18	−1.307692	15.8829

Writing $p_j = p_j(\frac{1}{2}\delta^2)$ and $Q_j = Q_{pm+2j}(X_j)$ we read off the values of p_j and Q_j from Hartley and Pearson's (1950) Tables :

j	p_j	Q_j
0	0.36788	0.05000
1	0.36788	0.11338
2	0.18394	0.20081
3	0.06131	0.33278
4	0.01533	0.46830
5	0.00307	0.60070

The power of the test is thus given by

$$\text{Prob } (L < x) = \Sigma\, p_j Q_j = 0.12812 \backsim 0.13$$

to two places of decimals.

The following table gives the power of analysis of dispersion tests at the five percent level of significance for $p = 1, 2, 3, 4$, $m = 2, 3$, $n = 200$, $\delta^2 = 0(1)8$.

POWER OF ANALYSIS OF DISPERSION TEST WHEN THE ALTERNATIVE HYPOTHESIS IS OF UNIT RANK

$n = 200$ $\alpha = .05$

δ^2	probability of rejection of the hypothesis							
	$m = 2$				$m = 3$			
	number of variates (p)				number of variates (p)			
	1	2	3	4	1	2	3	4
0	.05	.05	.05	.05	.05	.05	.05	.05
1	.13	.10	.09	.09	.11	.09	.08	.08
2	.22	.16	.14	.13	.19	.14	.12	.11
3	.31	.23	.20	.17	.26	.20	.17	.15
4	.40	.31	.26	.23	.35	.26	.22	.19
5	.49	.38	.33	.28	.43	.33	.27	.23
6	.57	.45	.39	.34	.50	.39	.32	.28
7	.64	.52	.45	.40	.57	.45	.38	.33
8	.70	.59	.52	.46	.64	.52	.43	.38

REFERENCES

ANDERSON, T. W. (1957): *Introduction to Multivariate Analysis*, John Wiley and Sons.

BOSE, R. C. and ROY S. N. (1938): The exact distribution of the Studentised D^2-statistic. *Sankhyā*, **4**, 19.

DEEMER, W. L. and OLKIN, I. (1951): The Jacobians of certain matrix transformations useful in multivariate analysis. *Biometrika*, **38**, 345.

HARTLEY and PEARSON (1950): Tables of χ^2-integral and the cumulative Poisson distribution. *Biometrika*, **37**, 313.

HSU, P. L. (1938): Notes on Hotelling's generalised T. *Ann. Math. Stat.*, **9**, 231.

RAO, C. R. (1953): *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons.

ROY, J. (1951): The distribution of certain likelihood-ratio criteria useful in multivariate analysis. *Proc. Int. Stat. Inst.*, **33**(2), 219.

WILKS, S. S. (1932): Certain generalisations in the analysis of variance. *Biometrika*, **24**, 471.

Paper received : December, 1959.