On the Distribution of the Largest of Seven Roots of a Matrix in Multivariate Analysis

by

K. C. Sreedharan Pillai
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series Number 9
August 1963

On the Distribution of the Largest of Seven Roots of a Matrix in Multivariate Analysis[1]

by K. C. Sreedharan Pillai
Purdue University

1. INTRODUCTION

The joint distribution of s non-null characteristic roots of a matrix in multivariate analysis given by Fisher (1939), Girshick (1939), Hsu (1939) and Roy (1939) can be expressed in the form

$$f(\theta_1, \ldots, \theta_s) = C(s,m,n) \prod_{i=1}^{s} \theta_i^m (1-\theta_i)^n \prod_{i>j} (\theta_i - \theta_j) \tag{1}$$

$$(0 < \theta_1 \leq \ldots\ldots \leq \theta_s < 1) ,$$

where

$$C(s,m,n) = \frac{\pi^{\frac{1}{2}s} \prod_{i=1}^{s} \Gamma\left(\frac{2m+2n+s+i+2}{2}\right)}{\prod_{i=1}^{s} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{2n+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right)} , \tag{2}$$

and the parameters m and n are defined differently for various situations as described in detail by Pillai (1955, 1957).

[1]Research supported by XL Grant from the Purdue Research Foundation.

Roy (1945, 1953, 1957) has shown that tests of certain hypotheses in multivariate analysis and associated confidence interval estimation can be based on the extreme characteristic roots, θ_1 and θ_s, and Roy and Mikhail (1961) have demonstrated the monotonic character of the power functions of two multivariate tests using θ_s, the largest characteristic root.

In this paper, the cumulative distribution function (c.d.f.) of the largest of seven characteristic roots has been studied with a view to obtaining an approximation to the c.d.f. and computing the upper percentage points.

2. C.D.F. OF THE LARGEST OF SEVEN ROOTS

The c.d.f. of θ_s can be presented in the following determinantal form:

$$P_r(\theta_s \leq x) = C(s,m,n) \begin{vmatrix} \int_0^x \theta_s^{m+s-1}(1-\theta_s)^n \, d\theta_s \ldots\ldots \int_0^x \theta_s^m (1-\theta_s)^n \, d\theta_s \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \int_0^{\theta_2} \theta_1^{m+s-1}(1-\theta_1)^n \, d\theta_1 \ldots \int_0^{\theta_2} \theta_1^m (1-\theta_1)^n \, d\theta_1 \end{vmatrix} . \qquad (3)$$

Now put $s=7$ in (3), and thus the c.d.f. of the largest of seven roots is obtained as a determinant of order seven which has to be expanded keeping in mind the order of integration of the variables.

In order to overcome the difficulty of integrating each of the $s!$ multiple integrals in the expansion of the determinant in (3), Pillai (1954, 1956) suggested a reduction formula and gave exact expressions for the c.d.f. of the largest root in terms of incomplete beta functions or functions of incomplete beta functions for values of s from 2 to 10. In addition, Pillai (1954) suggested a method of approximating the c.d.f. of the largest root for computing the upper percentage points and this method which was used to obtain approximations to the c.d.f. and upper percentage points for $s=2,3,4,5$ and 6 (Pillai, 1954, 1960) is once again used to derive the approximation suggested in the following section.

3. AN APPROXIMATION TO THE C.D.F. OF THE LARGEST OF SEVEN ROOTS

Following Pillai's method (1954, 1956a), an approximation to the c.d.f. of the largest of seven roots, useful in computing the upper percentage points, is obtained and presented below:

$$1-P_r(\theta_7\leq x) = K\Big\{a_0 I_0(m+6) - a_1 I_0(m+5) + a_2 I_0(m+4) - a_3 I_0(m+3) + a_4 I_0(m+2)$$

$$+ b_0 J(m) + b_1 J(m+1) - b_2 J(m+2) + b_3 J(m+3)\Big\}, \qquad (4)$$

where $K = \dfrac{\Gamma(m+n+8)}{45\ (m+n+5)(m+n+6)(m+n+7)\ \Gamma(m+4)\ \Gamma(n+4)}$,

$$a_0 = 3(2m+2n+9)(2m+2n+11)(2m+2n+13)(m+n+5)(m+n+6),$$

$$a_1 = 15(2m+2n+9)(2m+2n+11)(2m+2n+13)(m+3)(m+n+5),$$

$$a_2 = 15(2m+2n+9)(2m+2n+11)(m+3)(4m^2+4mn+34m+11n+61),$$

$$a_3 = \frac{(m+2)(m+3)(m+4)(2m+7)(2m+2n+7)(2m+2n+9)(2m+2n+11)}{(2n+3)(2n+4)(2n+5)} \times$$

$$\times \Big\{(m+n)\left[4m^2(m+n)+60m^2-2mn+30n+105\right] + 269m^2-mn+225(m+1)\Big\},$$

$$a_4 = \frac{3(2m+2n+5)(2m+2n+7)(2m+2n+9)(m+2)(m+3)^2(m+4)(2m+5)(2m+7)}{2(2n+3)(2n+5)},$$

$$b_0 = 3(m+1)(m+2)(m+3)(2m+3)(2m+5)(2m+7),$$

$$b_1 = \frac{3(2m+2n+9)(m+1)(m+2)(m+3)(2m+5)(2m+7)}{2(2n+3)(2n+5)} \text{ X}$$

$$\text{X} \left\{ 4m^4+(8n+56)m^3+(4n^2+88n+299)m^2 + (32n^2+336n+736)m+24n^2+288n+570 \right\},$$

$$b_2 = \frac{(2m+2n+9)(m+n+5)(2m+2n+11)(m+1)(m+2)(m+3)(2m+3)(2m+7)}{(2n+3)(2n+4)(2n+5)} \text{ X}$$

$$\text{X} \left\{ 4m^4+(8n+72)m^3+(4n^2+94n+461)m^2+(22n^2+305n+1173)m+60n^2+390n+1080 \right\},$$

$$b_3 = \frac{(2m+2n+9)(2m+2n+11)^2(2m+2n+13)(m+n+5)(m+n+6)}{(2n+3)(2n+4)(2n+5)} \text{ X}$$

$$\text{X } (m+1)(m+2)^2(m+3)(2m+3)(2m+5),$$

$$I_0(k) = x^k(1-x)^{n+1},$$

and

$$J(i) = \frac{x^i(1-x)^{n+1}}{n+1} + \frac{i\, x^{i-1}(1-x)^{n+2}}{(n+1)\ (n+2)} + \frac{i(i-1)x^{i-2}(1-x)^{n+3}}{(n+1)(n+2)(n+3)}$$

$$+ \dots + \frac{i(i-1)\dots 2.1.x^0(1-x)^{n+1+i}}{(n+1)(n+2)\dots\dots(n+1+i)} .$$

Now, since $b_0\beta(m+1,n+1)+b_1\beta(m+2,n+1)-b_2\beta(m+3,n+1)+b_3\beta(m+4,n+1) = 1$ (5)

the c.d.f. also can be expressed in the alternate form

$$P_r(\theta_7 \leq x) = K \left\{ -a_0I_0(m+6)+a_1I_0(m+5)-a_2I_0(m+4)+a_3I_0(m+3)-a_4I_0(m+2) \right.$$

$$\left. +b_0I(m) \quad +b_1I(m+1) \ -b_2I(m+2) \ +b_3I(m+3) \right\}, \tag{6}$$

where

$$I(i) = \int_0^x \theta^i(1-\theta)^n \, d\theta .$$

It may be pointed out that although the expression on the right side of (4) is comparatively neat, the algebraic computation behind it was extremely heavy.

4. UPPER PERCENTAGE POINTS

Pillai (1954, 1956, 1957, 1960) gave the upper 5°/o and 1°/o points of the largest root for s=2 and s=5 using his approximation formulae. Sen (1957) computed similar upper percentage points for three roots, Ventura (1957) for four roots, both following Pillai's method. Pillai and Bantegui (1959) gave such percentage points for s=6. All these percentage points were given for values of m=0(1)4 and n varying from 5 to 1000. Further, Jacildo (1959) extended the tables for s=2 and s=3 for values of m=5,7,10 and 15 and the same range of values of n as before. Pillai (1960) has published all these percentage points for s=2,3,4,5 and 6.

Foster and Rees (1957) have tabulated the upper percentage points (80, 85, 90, 95 and 99) of the largest root for s=2, m=-0.5,0(1)9 and n=1(1)19(5)49, 59,79. Foster (1957, 1958) has further extended these tables for values of s=3 and 4. The arguments they have used for tabulation are $\nu_1 = 2n+s+1$ and $\nu_2 = 2m+s+1$. Heck (1960) has given some charts of upper 5°/o, 2.5°/o and 1°/o points for s=2(1)5, $m = -\frac{1}{2}$, 0(1)10 and $n \geq 5$.

Upper 5°/o and 1°/o points were computed for θ_7 using (4) for values of m = 0(1)5,7 and 10 and n ranging from 5 and 1000. These are presented in Tables 1 and 2. The computation was carried out on IBM 7090 but a trial value was extrapolated from Pillai's tables (1960) of percentage points for s=2,3,4,5 and 6, for each of 8 x 16 = 128 combinations of m and n in order to be fed into the machine. However, the trial values for m = 5,7 and 10 were computed using tables for s=2 and 3 only.

Table 1. Upper 5% Points of the Largest Root for s=7

n \ m	0	1	2	3	4	5	7	10
5	.85229	.87214	.88715	.89893	.90846	.91632	.9288	.9445
10	.69490	.72561	.75028	.77064	.78778	.80243	.8264	.8540
15	.57912	.61295	.64111	.66505	.68575	.70387	.7342	.7695
20	.49436	.52818	.55698	.58197	.60396	.62353	.6570	.6970
25	.43049	.46310	.49132	.51617	.53832	.55827	.5929	.6352
30	.38090	.41189	.43903	.46319	.48495	.50472	.5395	.5825
40	.30923	.33684	.36144	.38367	.40398	.42267	.4561	.4987
60	.22433	.24644	.26650	.28496	.30209	.31811	.3474	.3858
80	.17590	.19414	.21088	.22642	.24098	.25471	.2801	.3141
100	.14463	.16011	.17441	.18776	.20036	.21230	.2345	.2647
130	.11417	.12676	.13845	.14945	.15987	.16981	.1885	.2141
160	.094297	.104892	.114774	.124101	.132975	.14147	.15750	.17965
200	.076532	.085273	.093455	.101205	.108603	.11570	.12917	.14792
300	.052023	.058098	.063813	.069251	.074465	.079493	.08909	.10258
500	.031710	.035480	.039040	.042442	.045714	.048883	.05496	.06357
1000	.016046	.017979	.019811	.021566	.023261	.024905	.02807	.03259

Table 2. Upper 1% Points of the Largest Root for s=7

n \ m	0	1	2	3	4	5	7	10
5	.89470	.90908	.91991	.92839	.93522	.9408	.9498	.9614
10	.75082	.77656	.79714	.81405	.82824	.8403	.8600	.8818
15	.63628	.66646	.69144	.71260	.73083	.7467	.7732	.8038
20	.54905	.58029	.60677	.62966	.64973	.6675	.6978	.7336
25	.48171	.51253	.53909	.56238	.58308	.6016	.6338	.6727
30	.42859	.45835	.48430	.50732	.52798	.5467	.5795	.6198
40	.35059	.37767	.40170	.42335	.44306	.4612	.4934	.5342
60	.25649	.27866	.29872	.31712	.33415	.3500	.3789	.4168
80	.20204	.22055	.23748	.25317	.26783	.2816	.3071	.3409
100	.16660	.18243	.19700	.21058	.22336	.2354	.2579	.2882
130	.13187	.14483	.15684	.16810	.17875	.1889	.2079	.2338
160	.109113	.120064	.130255	.139852	.148966	.15767	.17406	.19664
200	.088695	.097764	.106235	.114242	.121870	.12918	.14300	.16220
300	.060419	.066754	.072702	.078351	.083759	.08896	.09888	.11277
500	.036892	.040839	.044561	.048111	.051521	.05482	.06113	.07005
1000	.018692	.020724	.022645	.024483	.026254	.02798	.03127	.03597

For $m = 0(1)4$, values of x were first computed such that: $1 - P_r(\theta_7 \leq x)$ was within a unit difference in the sixth decimal. This actually gave six places accuracy for the percentage points. However, only five places were retained in general. For $m=5$ and 7, $1 - P_r(\theta_7 \leq x)$ was allowed to be within five units difference in the fifth decimal and for $m=10$, one unit in the fourth decimal.

5. ERROR OF APPROXIMATION

A detailed examination of the error of approximation for previous studies has been made by Pillai and Bantegui (1959). The practice followed previously of checking the difference between the approximate and exact probabilities for a percentage point has not been adhered to in the present study for two reasons; firstly, the approximation has been observed to retain practically the same nature of accuracy for values studied already, i.e. for $s=2$ to $s=6$. For instance, for $m=0$ and $n=500$ the difference between the approximate and exact probabilities for the upper 5°/o point was observed to be 0.0000326 for $s=3$ while it still remained 0.0000410 for $s=6$. Secondly, even the rough extrapolations made (from simple adjustments on first differences of percentage points for successive values of s) from Pillai's tables to obtain trial values, gave results very close to the final computed values considering four significant digits. In fact the following table shows the agreement between the trial values and the final values for $m = 0(1)4$.

Table 3. Frequency of the Numbers of Units of Difference (in The Fourth Significant Digit) Between the Trial and Final Percentage Points

Number of Units	Frequency	Cumulative Frequency	Percentage Cumulative Frequency
0	55	55	34.3
1	32	87	54.4
2	21	108	67.5
3	19	127	79.4
4	10	137	85.6
above 4	23	160	100.0
Total	160		

It may be observed that differences of the number of units more than four in the fourth significant digit has arisen only in 14.4 percent of cases and most of these have been due to the fact that the percentage points for $s=2$ (and $s=4$ in a few cases) used carried only three significant digits.

That the approximation gives sufficient accuracy for m up to 10 or even more has been checked by Foster (1957), Jacildo (1959) and Heck (1960) and the accuracy for the retention of five decimal places in the percentage points has been ascertained previously by comparison with Foster's tables (1957). In fact, the approximate formula is adequate even for percentage points slightly below the $95^{o}/o$ level. Foster and Rees (1957) have also discussed this point for $s=2$.

6. SUMMARY

In this paper the cumulative distribution function of the largest of seven roots of a matrix in multivariate analysis has been studied. An approximation for obtaining the upper percentage points (5°/o or less) to the c.d.f. of the largest of seven roots has been obtained following Pillai's method. Upper 5°/o and 1°/o points were computed for small integral values of one parameter and values of a second parameter ranging up to 1000.

The author wishes to thank Mr. Peter E. Dress, Instructor in Mathematics, Department of Statistics, Miss Barbara A. McCollough and Mrs. Shirley Wolfe, Statistical Laboratory, Purdue University, for the excellent programming of this material for the IBM 7090 computer, Purdue University's Computer Sciences Center. But for Professor Virgil L. Anderson's interest and encouragement, these computations might not have been carried out on the IBM. The author wishes to express, therefore, his sincere thanks to Professor Anderson.

REFERENCES

Fisher, R. A. (1939). Ann. Eugenics, London, 9, 238.

Foster, F. G. (1957). Biometrika, 44, 441.

Foster, F. G. (1958). Biometrika, 45, 492.

Foster, F. G. and Rees, D. H. (1957). Biometrika, 44, 237.

Girshick, M. A. (1939). Ann. Math. Statist., 10, 203.

Heck, D. L. (1960). Ann. Math. Statist. 31, 625.

Hsu, P. L. (1939). Ann. Eugenics., London, 9, 250.

Jacildo, L. (1959). Further studies on the distributions of the largest of two and three roots. Unpublished Thesis. The Statistical Center, University of the Philippines.

Pillai, K. C. S. (1954). On some distribution problems in multivariate analysis, Mimeo. Series No. 88, Institute of Statistics, University of North Carolina.

Pillai, K. C. S. (1955). Ann. Math. Statist., 26, 117.

Pillai, K. C. S. (1956a). Biometrika, 43, 122.

Pillai, K. C. S. (1956b). Ann. Math. Statist., 27, 1106.

Pillai, K. C. S. (1957). Concise Tables for Statisticians, The Statistical Center, University of the Philippines.

Pillai, K. C. S. (1960). Statistical Tables for Tests of Multivariate Hypotheses, The Statistical Center, University of the Philippines.

Pillai, K. C S. and Bantegui, C. G. (1959). Biometrika, 46, 237.

Roy, S. N. (1939). Sankhyā, 4, 381.

Roy, S. N. (1945). Sankhyā, 7, 133.

Roy, S. N. (1953). Ann. Math. Statist., 24, 220.

Roy, S. N. (1957). Some Aspects of Multivariate Analysis, New York, John Wiley and Sons,

Roy, S N. and Mikhail, W. F (1961). Ann. Math. Statist., 32, 1145.

Sen, P. (1957). On a multivariate test criterion and its applications, Unpublished Thesis. The Statistical Center, University of the Philippines.

Ventura, S R. (1957). On the extreme roots of a matrix in multivariate analysis and associated tests, Unpublished Thesis. The Statistical Center, University of the Philippines.