PERCENTILE POINTS OF THE LARGEST LATENT ROOT OF A MATRIX AND POWER CALCULATIONS FOR TESTING HYPOTHESIS $\Sigma = \mathbf{I}$.

by

T. Sugiyama

University of North Carolina

Institute of Statistics Mimeo Series No. 590

August 1968

This research was supported by the National Institute of Health Grant No. 5401-GM12864-04 and Contract No. 325-FED-282.

DEPARTMENT OF STATISTICS

University of North Carolina

Chapel Hill, N. C.

SUMMARY

Percentile points of the largest latent root of a sample covariance matrix are given for the cases $p = 2$, 3, and 4 when the population covariance matrix is the identity matrix; and also the power of the largest root test of the hypothesis $\Sigma = I$ for the case $p = 2$.

1. Percentile Points of the Largest Latent Root

The largest latent root of the sample covariance matrix is used as a test criterion of the hypothesis $\Sigma = I$, where Σ is the population covariance matrix of a multinormal population. From the discussion to the sensitivities of the several test criteria of the general linear hypothesis due to Schatzoff [2], we can expect that the largest root criterion has larger power than the likelihood ratio criterion when there is just one root of the equation $|\Sigma - \lambda I| = 0$ greater than 1.

Starting from the probability density function of the largest latent root of the Wishart matrix corresponding to sample size (n + 1) obtained by Sugiyama [3], and applying Kummer's transformation we obtain the cumulative distribution function expressed as a series of positive terms:

$$(1)\quad F(\lambda_1 < x) = \text{Const. } \exp\left(-\frac{nx}{2}\operatorname{tr}\Sigma^{-1}\right)(nx)^{pn/2}$$

$$\qquad {}_1F_1\left((p+1)/2;\ (n+p+1)/2;\ \frac{nx}{2}\Sigma^{-1}\right)$$

where λ_1 is the largest latent root of the sample covariance matrix, and

$$\text{Const.} = \frac{\Sigma^{-n/2}\,\Gamma_p((p+1)/2)}{2^{np/2}\,\Gamma_p((n+p+1)/2)}\ .$$

(${}_1F_1$ is the confluent hypergeometric function of matrix argument defined by

$$ {}_1F_1\ (a;\ b;\ S) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a)_\kappa}{(b)_\kappa} \frac{C_\kappa(S)}{k!} \Big) $$

To calculate the zonal polynomial $C_\kappa(I_p)$, where I_p is the p x p identity matrix, we used a formula given by James [1] which is convenient for computational programming. The computation was carried out on IBM 360 model 75 machine using double precision arithmatic (16 significant figures) in the cases p = 2 and 3. In the case p = 2, values of n from 10 to 18, we needed 95 terms of the hypergeometric series in the formula (1) to find the percentile points with sufficient accuracy; for values of n from 20 to 50, from 100 to 120 terms were needed. For example, using 100 and 120 terms for the values n = 22 and 42 respectively, we have the values of the cumulative distribution function:

Table 1 (p = 2)

n = 22(100 terms)		n = 42 (120 terms)
x	cdf	cdf
1.0	0.19168434	0.17864268
1.4	0.70195384	0.84265137
1.8	0.94929632	0.99390859
2.2	0.99488365	0.99991621
2.6	0.99963706	0.99999849
3.0	0.99997998	
3.4	0.99999888	

We do not observe a cumulative value larger than 0.99999888 in the case n = 22 and 100 terms (even for $x > 3.5$). It is obvious that the value will be closer to one if we include even higher terms, but it is quite sufficient to get the answer correct to 7 decimal places in the cases we consider.

In the case $p = 3$, we take 100 terms for the value $n = 10$ to find the percentile points. Table 2 shows some values of the cumulative distribution functions calculated in this way. For values of n larger than 10, $(n + 90)$ terms were taken.

Table 2 ($p = 3$)

n = 22 (112 terms)		n = 42 (132 terms)
x	cdf	cdf
1.1	0.10616941	0.12525754
1.4	0.45839761	0.65933736
1.8	0.86801094	0.97952710
2.2	0.98295889	0.99962802
2.3	0.99049542	0.99984993
2.6	0.99855649	
3.0	0.99990282	

It may be interesting to compare Table 1 and Table 2. We notice that the convergence of the hypergeometric function in the formula (1) when $\Sigma = I$ becomes slow when the sample size $(n + 1)$, and the number of variables p, increase. The values shown to 6 decimal places in Table 3, corresponding to the different sample sizes $n + 1$, were based on original calculations which are believed to be correct to at least 10 decimal places for $Pr(x < \theta < \infty) = 0.05$ and 7 decimal places for $Pr(x < \theta < \infty) = 0.01$ in the case $p = 2$; and also to at least 8 decimal places for the upper 5% points and about 6 decimal places for the upper 1% point in the case $p = 3$.

Table 3

Upper α percentile points of the largest latent root of a sample covariance matrix having the population covariance matrix $\Sigma = I$ and sample size n + 1.

	p = 2		p = 3	
n \ α	0.05	0.01	0.05	0.01
2	4.297438	6.080079	5.370173	7.284052
4	3.169279	4.181691	3.810174	4.875353
6	2.702481	3.439849	3.181457	3.948462
8	2.436461	3.029240	2.828000	3.439790
10	2.260743	2.763132	2.596608	3.112055
12	2.134252	2.574213	2.431132	2.880413
14	2.037911	2.431863	2.305741	2.706485
16	1.961542	2.320002	2.206759	2.570206
18	1.899176	2.229311	2.126207	2.459990
20	1.847059	2.153989	2.059093	2.368648
22	1.802700	2.090220	2.002116	2.291460
24	1.764374	2.035382	1.953000	2.225192
26	1.730847	1.987609	1.910120	2.167547
28	1.701207	1.945532	1.872281	2.116844
30	1.674767	1.908124	1.838584	2.071823
32	1.650998	1.874598	1.808336	2.031519
34	1.629485	1.844339	1.780995	1.995179
36	1.609896	1.816858	1.756133	1.962207
38	1.591965	1.791762	1.733401	1.932124
40	1.575473	1.768731	1.712516	1.904540
42	1.560240	1.747502	1.693245	1.879133
44	1.546115	1.727856	1.675393	1.855638
46	1.532972	1.709609	1.658798	1.833831
48	1.520704	1.692604	1.643319	1.813522
50	1.509219	1.676711	1.628840	1.794552

Table 4 gives values in the case p = 4 which are believed to be correct to about 4 and 2 decimal places for upper 5% and 1% points respectively. Because of the large amount of computation time needed, we give here only a few percentile points:

Table 4 (p = 4)

n \ α	0.95	0.99
2	6.3391	8.36
4	4.3805	5.49
6	3.6043	4.40
8	3.1717	3.80
10	2.8902	3.42
20	2.2423	2.56
30	1.9794	2.22

Using the same program, we could get values for the upper 5% and 1% points when the number of dimensions is higher than 4. However this would take considerably more computation time.

2. Power tabulation for test of hypothesis $\Sigma = I$ based on the largest latent root.

If p = 2, then the cumulative distribution function of the largest latent root is

$$(1)' \qquad F(\lambda_1 < x) = \frac{|\Sigma^{-1}|^{n/2}\, n^n}{(n+1)!} \exp\left(-\frac{n}{2}\operatorname{tr}\Sigma^{-1}\right) x^n \sum_{k=0}^{\infty} \sum_{\kappa} \frac{((p+1)/2)_\kappa}{((n+p+1)/2)_\kappa} \left(\frac{n}{2}\right)^k \frac{C_\kappa(\Sigma^{-1})}{k!} .$$

where the partition κ of the integer k runs over all partitions such that $\kappa = (k_1, k_2)$, $k_1 \geq k_2 \geq 0$, and $k_1 + k_2 = k$. Since the series in formula (1)' converges more slowly when the sample size (n + 1) increases and the alternative deviates more from the null-hypothesis, it becomes necessary to calculate zonal polynomials $C\ (\Sigma^{-1})$ for high-order partition. In the case P = 2, using an explicit formula given by James [1], it is possible to obtain values, by double precision arithmatic calculations, up to about $k_1 - k_2 = 70$ in the case $\sigma_1 = 1.4$ and $\sigma_2 = 0.6$; to about $k_1 - k_2 = 80$ when $\sigma_1 = 1.5$ and $\sigma_2 = 0.5$, and to about $k_1 - k_2 = 105$ when $\sigma_1 = 1.6$ and $\sigma_2 = 0.4$. Table 5 gives values of the power with respect to the alternatives σ_1 ($\sigma_1 + \sigma_2 = 2$) and using tests at 5% and 1% significance levels. Naturally, the changes in numbers of decimal places in the power given in Tables 4 and 5 arise from the restriction of the computation of the zonal polynomials. The values of the power given by Tables 4 and 5 were calculated using 5% and 1% percentile points which are correct up to 7 decimal places.

Figure 1 shows the power curves, for different sample sizes n + 1, with respect to the fixed alternatives σ_1 = 1.2, 1.3, 1.4, 1.5, 1.6, and 1.7 using a test with 5% significance level, and figure 2 for different alternatives with fixed sample sizes n + 1 = 3, 11, 21, 31, and 41. Because the power curves for 1% significance level show similar tendencies as for the 5% cases, they are omitted.

Power curves for the different sample size n + 1. ($\sigma_1 + \sigma_2 = 2$)

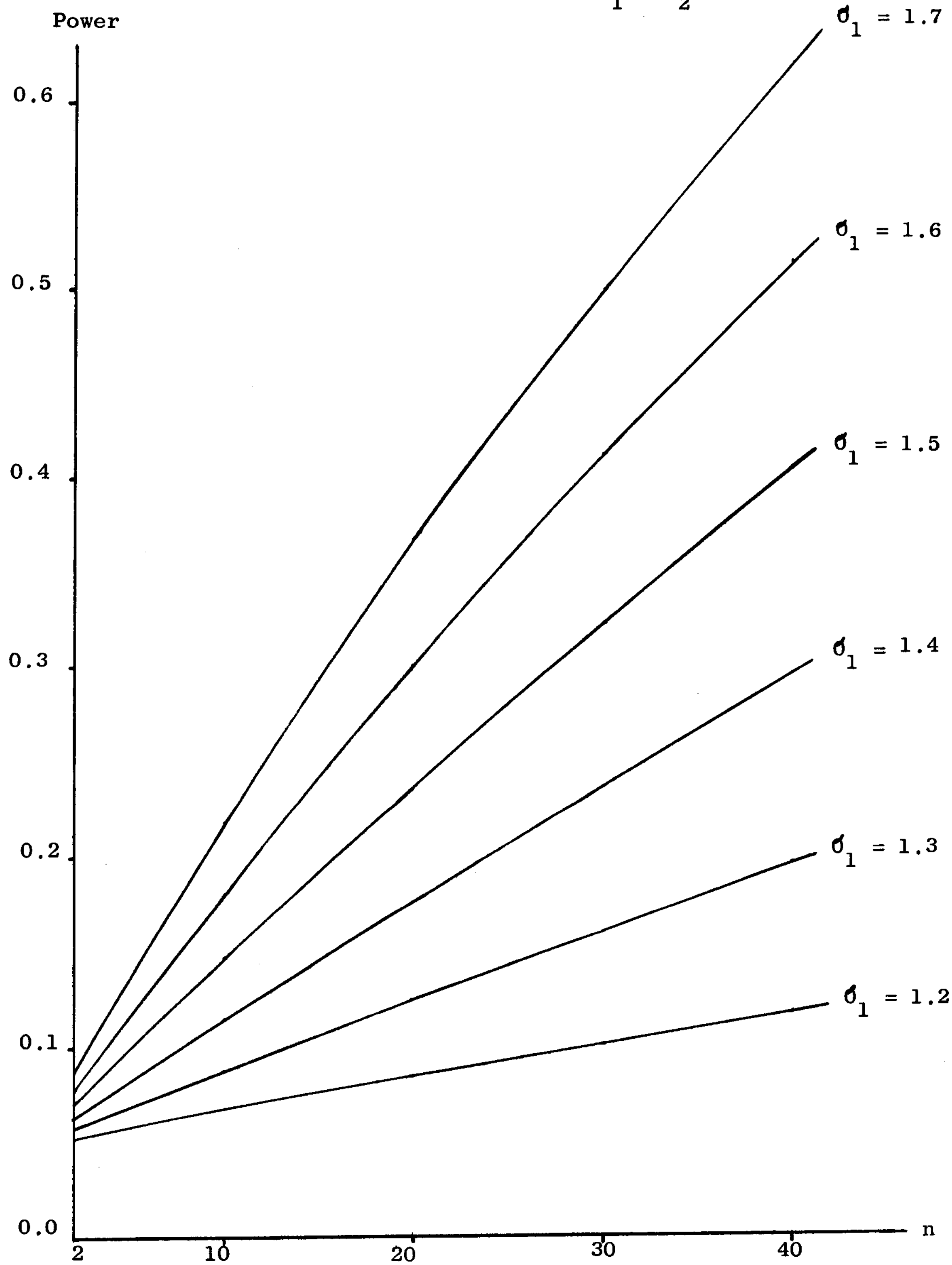

Figure 1.

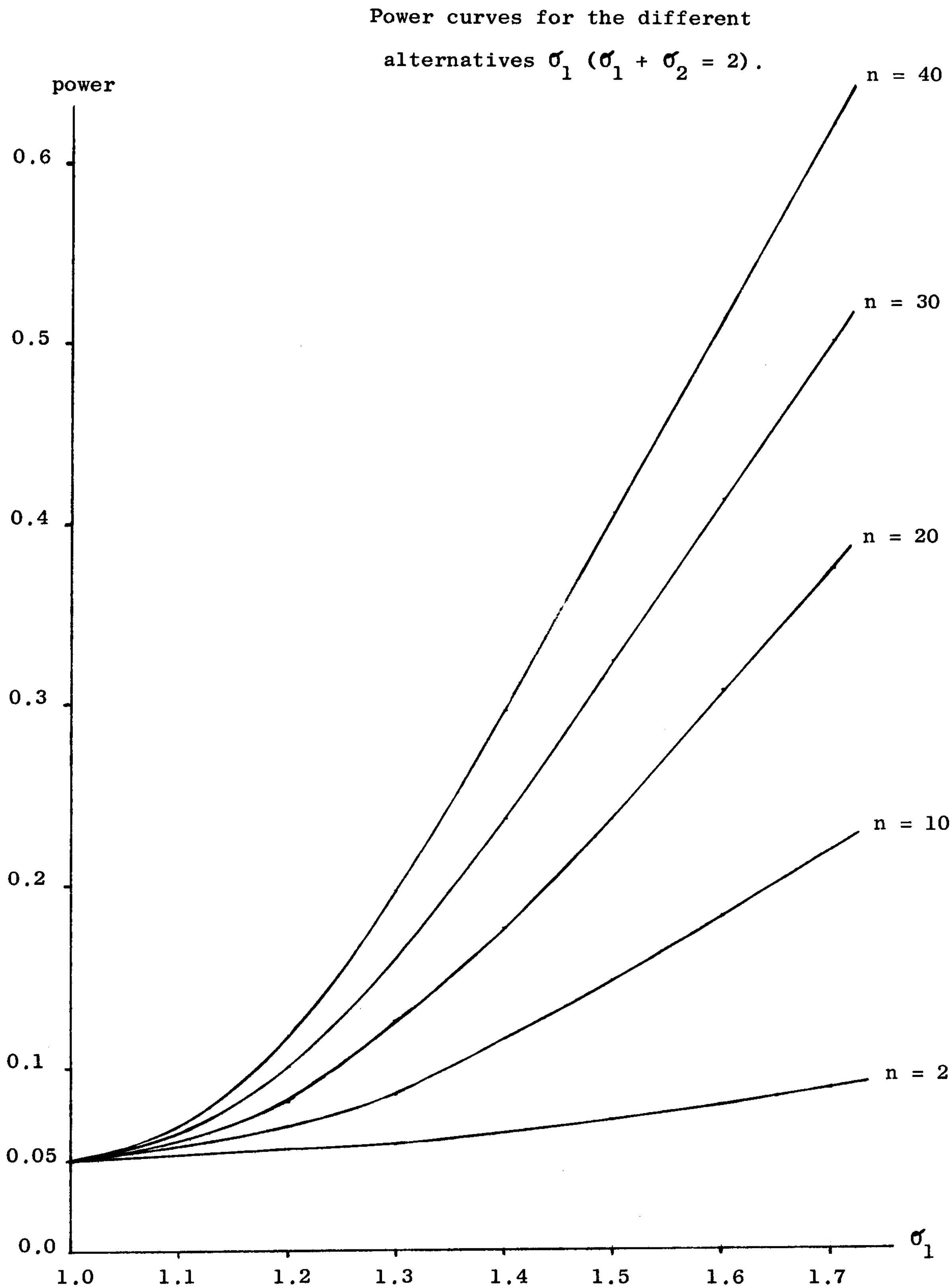
Power curves for the different
alternatives σ_1 ($\sigma_1 + \sigma_2 = 2$).
power
0.6
0.5
0.4
0.3
0.2
0.1
0.05
0.0
n = 40
n = 30
n = 20
n = 10
n = 2
σ_1
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7

Figure 2.

Table 5

Values of the power for the alternatives σ_1 ($\sigma_1 + \sigma_2 = 2$) and the sample size n + 1 based on 5% significance level.

n \ σ_1	1.2	1.3	1.4	1.5	1.6	1.7
2	.05362	.05804	.06400	.07126	.07948	.08835
4	.05744	.06629	.07791	.09164	.10689	.12316
6	.06103	.07402	.09086	.11062	.13247	.15579
8	.06452	.08150	.10342	.12904	.15735	.18753
10	.06795	.08885	.11576	.14718	.18185	.21873
12	.07133	.09612	.12798	.16515	.20609	.24946
14	.07469	.10334	.14014	.18302	.23010	.27972
16	.07803	.11053	.15225	.20079	.25389	.30947
18	.08135	.11770	.16434	.21848	.27742	.33867
20	.08467	.12486	.17641	.23608	.30069	.36727
22	.08798	.13202	.18845	.25358	.32365	.39522
24	.09128	.13918	.20048	.27097	.34629	.42248
26	.09458	.14634	.21249	.28823	.36857	.44903
28	.09789	.15350	.22448	.30535	.39047	.47484
30	.10119	.16066	.23643	.32233	.41197	.49984
32	.1045	.1678	.2484	.3391	.4330	.5241
34	.1078	.1750	.2602	.3558	.4537	.5479
36	.1111	.1822	.2721	.3722	.4739	.5713
38	.1144	.1894	.2839	.3884	.4936	.5950
40	.1177	.1965	.2957	.4045	.5128	.6199

Table 6

Values of the power for the alternatives σ_1 ($\sigma_1 + \sigma_2 = 2$) and the sample size n + 1 based on 1% significance level.

n \ σ_1	1.2	1.3	1.4	1.5	1.6	1.7
2	.0120	.0144	.0176	.0215	.0260	.0309
4	.0134	.0176	.0231	.0300	.0380	.0470
6	.0147	.0205	.0284	.0383	.0500	.0634
8	.0160	.0235	.0337	.0468	.0624	.0805
10	.0173	.0264	.0391	.0555	.0753	.0983
12	.0185	.0293	.0447	.0646	.0887	.1168
14	.0198	.0323	.0504	.0739	.1026	.1359
16	.0210	.0354	.0562	.0835	.1169	.1557
18	.0223	.0385	.0622	.0934	.1317	.1759
20	.0236	.0416	.0683	.1036	.1468	.1966
22	.0248	.0448	.0746	.1140	.1622	.2175
24	.0261	.0481	.0810	.1247	.1780	.2391
26	.0274	.0515	.0876	.1356	.1940	.2618
28	.0287	.0548	.0943	.1467	.2102	.2873
30	.0301	.0583	.1012	.1580	.2266	.3076
32	.031	.062	.108	.169	.243	
34	.033	.065	.116	.181	.260	
36	.034	.069	.123	.193	.277	
38	.036	.073	.131	.205	.295	
40	.037	.077	.140	.217	.311	

ACKNOWLEDGEMENT: The author wishes to express his sincere thanks to Mr. James O. Kitchen for his valuable suggestions in the computation of the zonal polynomial, and Professor Norman L. Johnson for his kind encouragement.

REFERENCE

[1] James, A. T. (1966). Inference on latent roots by calculation of hypergeometric functions of matrix argument. *Mult. Ann.*, Edited by Krishnaiah.

[2] Schatzoff, M. (1966). Sensitivity comparisons among tests of the general linear hypothesis. *Journal of the American Statistical Association*, *61*, 415-435.

[3] Sugiyama, T. (1967). On the distribution of the largest latent root of the covariance matrix. *Ann. Math. Statist.* *38*, 1148-1151.