

ON THE DISTRIBUTION OF THE RATIO OF THE  $i$ TH OBSERVATION<sup>1</sup>  
IN AN ORDERED SAMPLE FROM A NORMAL POPULATION TO AN  
INDEPENDENT ESTIMATE OF THE STANDARD DEVIATION

BY K. C. S. PILLAI AND K. V. RAMACHANDRAN

*University of North Carolina*

**1. Summary.** This paper deals with the distribution of any observation,  $x_i$ , in an ordered sample of size  $n$  from a normal population with zero mean and unit standard deviation. The distribution has been developed as a series of Gamma functions, and has been used to obtain the distribution of  $q_i = (x_i/s)$ , where  $s$  is an independent estimate of the standard deviation with  $\nu$  degrees of freedom. In a similar manner the distribution of the Studentized maximum modulus  $u_n = |x_n/s|$  has been obtained and upper 5 per cent points of  $q_n$  and upper and lower 5 per cent points of  $u_n$  have been given. The method of obtaining the different distributions essentially depends on appropriate expansions of the normal probability integral and its powers in the intervals  $-\infty$  to  $x$  and 0 to  $x$ .

**2. Introduction.** The study of ordered samples from a normal population has led many authors to the construction of different Studentized tests based on outlying observations. One of the important tests in this group is that based on the Studentized range, for which tables of significance levels have been given by May [4] and Pillai [8]. Nair [5] has considered the distribution of the Studentized extreme deviate from the sample mean.

In the present paper the Studentized extreme deviate from the population mean as well as the Studentized maximum modulus are discussed and their distributions given for small sample sizes. Roy and Bose [1] and Tukey [9] have suggested the use of Studentized maximum modulus for simultaneous confidence interval statements. These authors have illustrated the use of the upper percentage points of the Studentized maximum modulus.

Box [2], [3] has suggested the criterion  $u_n$  as a possible test for platykurtosis. He points out that if the mean is assumed known, then  $u_n$  is the likelihood criterion for testing the null hypothesis of normality against the alternative that the distribution is rectangular. Significance is attained if  $u_n$  is too small; the test uses the lower tail area of the Studentized maximum modulus. The Studentized extreme deviate from the population mean can be used in different situations, including the problem of simultaneous confidence interval statements.

---

Received 6/29/53, revised 2/8/54.

<sup>1</sup> Presented to the meetings of the Institute of Mathematical Statistics at Chicago, Illinois, December 30, 1952.

TABLE I  
Coefficients  $a_i^{(k)}$  for determining coefficients in expansion of normal probability integral (cf. Eqns. (3.7) and (3.8))

$i$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
0	.50000000	.25000000	.12500000	.06250000	.03125000	.01562500	.00781250
1	.39894228	.39894228	.29920671	.19947114	.12466946	.074801677	.043634312
2	.08333333	.24248827	.30123241	.28039908	.22498535	.16483276	.11356002
3	.00000000	.066490381	.16322920	.22672284	.24184706	.22106882	.18252664
4	.0*69444444	.013888889	.055413735	.11879665	.17364827	.20227295	.20317475
5	.0*44326920	.099735572	.019947114	.048314782	.092106224	.13648695	.16794983
6	.0*38580247	.050799863	.011225059	.021654495	.042857313	.074771485	.11017451
7	-.0*14072038	.090588746	.0*49660421	.011190082	.021145714	.038066474	.062529157
8	.0*16075103	.0416322188	.0*12255041	.0*47138211	.010563732	.019623411	.033666909
9	.0*17590048	.0*68527894	.0*24943180	.0*14392049	.0*44470780	.0*96627902	.017740216
10	.0*53583677	.042256972	.0*13636594	.0*44476986	.0*15758833	.0*41848164	.0*87577403
11	-.0*71070901	.0*41143388	.0*60698505	.0*19767291	.0*58046643	.0*16453177	.0*39238553
12	.0*14884355	-.0*13383047	.0*10421860	.0*76549643	.0*24399876	.0*66245098	.0*16581736
13	.0*40243139	.0*21190527	.0*38170362	.0*18085787	.0*91007838	.0*27544134	.0*70311663
14	.0*35438940	.0*23053505	.0*80135071	.0*37096286	.0*26274003	.0*10376136	.0*29463704
15	-.0*81580690	.0*93991620	.0*46425639	.0*17964457	.0*76383520	.0*34166748	.0*11424039
16	.0*173831125	-.0*11517901	.0*47000596	.0*74037123	.0*30334492	.0*11514181	.0*41122684
17	.0*058932946	.0*36894111	-.0*13802971	.0*12599930	.0*10884773	.0*43525045	.0*14989845
18	.0*13672431	.0*94822617	.0*27623121	.0*69957761	.0*25212504	.0*15096316	.0*56433342
19	-.0*119067827	.0*012939560	.0*28220259	.0*99240268	.0*52692237	.0*42168894	.0*19806847
20	.0*422787385	-.0*050145804	.0*13135284	.0*53937376	.0*23621441	.0*11881132	.0*62274205
21	.0*358406921	.0*241086623	-.0*14925379	.0*55745257	.0*92015238	.0*44059326	.0*19924791
22	.0*34526341	.0*129526830	.0*162487064	-.0*14669723	.0*15685712	.0*15015557	.0*70504086
23	-.0*16064956	.0*11907969	.0*14601423	.0*36846548	.0*11547196	.0*34471983	.0*23286637
24	.0*847953251	-.0*14440848	.0*24978183	.0*34521877	.0*12647069	.0*73620775	.0*63698971
25	.0*41326546	.0*31756106	-.0*188085611	.0*17015699	.0*64692264	.0*31172628	.0*17828625
26	.0*261478527	.0*469636292	.0*99813888	-.0*18472470	.0*167974001	.0*11601965	.0*63081504
27	-.0*897882639	.0*178408304	.0*362775844	.0*98391492	-.0*115741833	.0*019918567	.0*20448365
28	.0*273188723	-.0*430387828	.0*34409747	.0*20151299	.0*246685732	.0*18079118	.0*47412257
29	.0*21740932	.0*18024897	-.0*36804948	.0*37879276	.0*242511479	.0*116394551	.0*10212511
30	.0*81320803	.0*786164817	.0*611854144	-.0*15211642	.0*21457246	.0*79226637	.0*41797437

3. Power series expansion for the normal probability integral. In this section we develop a power series expansion for the normal probability integral over the range  $-\infty$  to  $x$ . Let

$$(3.1) \quad I(k, x) = \left( \int_{-\infty}^x e^{-t^2/2} dt / \sqrt{2\pi} \right)^k.$$

An appropriate expansion (cf. Section 5) for  $I(k, x)$  is given by

$$(3.2) \quad I(k, x) = e^{-kx^2/6} (a_0^{(k)} + a_1^{(k)} x + a_2^{(k)} x^2 + \dots),$$

where the  $a$ 's are given by the recurrence relations

$$(3.3) \quad (2j + 1)a_{2j+1}^{(k)} = (k/\sqrt{2\pi}) [a_{2j}^{(k-1)} - (1/3)a_{2j-2}^{(k-1)} + \dots + \{(-1)^j/(j!3^j)\} a_0^{(k-1)}] + (k/3)a_{2j-1}^{(k)},$$

$$(3.4) \quad (2j + 2)a_{2j+2}^{(k)} = (k/\sqrt{2\pi}) [a_{2j+1}^{(k-1)} - (1/3)a_{2j-1}^{(k-1)} + \dots + \{(-1)^j/(j!3^j)\} a_1^{(k-1)}] + (k/3)a_{2j}^{(k)} \quad (j = 0, 1, 2, \dots),$$

and  $a_0^{(k)} = (\frac{1}{2})^k$ . Thus

$$(3.5) \quad \left( \int_x^\infty e^{-t^2/2} dt/\sqrt{2\pi} \right)^m = \left( \int_{-\infty}^{-x} e^{-t^2/2} dt/\sqrt{2\pi} \right)^m = I(m, -x)$$

and by using (3.2)

$$(3.6) \quad I(m, -x) = e^{-mx^2/6}(a_0^{(m)} - a_1^{(m)}x + a_2^{(m)}x^2 - \dots).$$

Hence

$$(3.7) \quad I(k, m, x) = e^{-(k+m)x^2/6}(b_0^{(k,m)} + b_1^{(k,m)}x + b_2^{(k,m)}x^2 + \dots),$$

where  $I(k, m, x) = I(k, x)I(m, -x)$  and

$$(3.8) \quad b_j^{(k,m)} = \sum_{i=0}^j (-1)^{j-i} a_i^{(k)} a_{j-i}^{(m)}.$$

Pillai [7] has given a similar expansion for the powers of the normal probability integral in the interval 0 to  $x$ . The  $a_i^{(k)}$  coefficients for  $i$  ranging from 0 to 30 and  $k$  from 1 to 7 are given in Table I.

**4. Distributions of the  $i$ th ranked observation, Studentized extreme deviate and Studentized maximum modulus.** Let  $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$  be an ordered sample from a normal population with zero mean and unit standard deviation. The distribution of any ranked observation,  $x_i$ , is given by

$$(4.1) \quad p(x_i) = [n!/(i-1)!(n-i)! \sqrt{2\pi}] I(i-1, x_i) I(n-i, -x_i) e^{-x_i^2/2}.$$

Using (3.7),  $p(x_i)$  takes the form

$$(4.2) \quad p(x_i) = [n!/(i-1)!(n-i)! \sqrt{2\pi}] e^{-(n+2)x_i^2/6} (b_0^{(i-1, n-i)} + b_1^{(i-1, n-i)} x_i + \dots).$$

The distribution of an independent estimate of the standard deviation  $s$  is given by

$$(4.3) \quad p(s) = [2(\nu/2)^{\nu/2}/\Gamma(\nu/2)] s^{\nu-1} e^{-\nu s^2/2}.$$

Multiplying (4.2) by (4.3), using the transformation  $q_i = x_i/s$ , and integrating with respect to  $s$  in the interval 0 to  $\infty$ , we get

$$(4.4) \quad p(q_i) = \frac{n!(\nu/2)^{\nu/2}}{(i-1)!(n-i)! \sqrt{2\pi} \Gamma(\nu/2)} \sum_{j=0}^{\infty} b_j^{(i-1, n-i)} q_i^j \left[ \frac{6}{(n+2)q_i^2 + 3\nu} \right]^{(j+\nu+1)/2} \Gamma\left(\frac{j+\nu+1}{2}\right).$$

Using (4.4), the probability integral of  $q_i$  can be evaluated with the help of Tables of the Incomplete Beta Function [6]. Putting  $i = n$  in (4.4) gives

$$(4.5) \quad p(q_n) = \frac{n(\nu/2)^{\nu/2}}{\Gamma(\nu/2) \sqrt{2\pi}} \sum_{j=0}^{\infty} q_n^j \left[ \frac{6}{(n+2)q_n^2 + 3\nu} \right]^{(j+\nu+1)/2} \Gamma\left(\frac{j+\nu+1}{2}\right) a_j^{(n-1)}.$$

TABLE II  
Upper 5% points of  $q_n = (x_n/s)$ , for ordered samples of sizes  $n$   
with  $\nu$  degrees of freedom

$\nu \backslash n$	1	2	3	4	5	6	7	8
3	2.35	3.10	3.53	3.85	4.12	4.34	4.53	4.71
4	2.13	2.72	3.06	3.31	3.51	3.67	3.82	3.98
5	2.01	2.53	2.83	3.03	3.21	3.34	3.45	3.58
6	1.94	2.42	2.68	2.86	3.02	3.14	3.25	3.36
7	1.89	2.34	2.60	2.75	2.90	3.01	3.11	3.21
8	1.86	2.29	2.52	2.67	2.82	2.92	3.02	3.11
9	1.83	2.24	2.47	2.62	2.75	2.85	2.94	3.04
10	1.81	2.22	2.44	2.59	2.70	2.79	2.88	2.97
12	1.78	2.18	2.38	2.53	2.63	2.72	2.81	2.90
14	1.76	2.14	2.34	2.49	2.58	2.67	2.75	2.83
15	1.75	2.13	2.32	2.47	2.56	2.65	2.73	2.81
16	1.75	2.12	2.31	2.45	2.54	2.63	2.71	2.78
18	1.73	2.10	2.29	2.43	2.52	2.61	2.68	2.75
20	1.72	2.09	2.27	2.41	2.50	2.58	2.65	2.72
24	1.71	2.06	2.24	2.38	2.47	2.55	2.62	2.68
30	1.70	2.04	2.22	2.35	2.44	2.52	2.59	2.65
40	1.68	2.02	2.20	2.32	2.41	2.49	2.55	2.61
60	1.67	2.00	2.17	2.29	2.38	2.45	2.51	2.57
120	1.66	1.98	2.14	2.26	2.35	2.42	2.47	2.53
$\infty$	1.64	1.96	2.12	2.23	2.32	2.39	2.44	2.49

Upper 5 per cent points of  $q_n$ , computed using (4.5), are given in Table II, for  $n$  from 1 to 8.

For obtaining the distribution of the maximum  $|x|$ , we may start with the probability law

$$(4.6) \quad \sqrt{2/\pi} e^{-t^2/2} \quad 0 < t < \infty$$

and noting [7] that

$$(4.7) \quad \left[ \int_0^x e^{-t^2/2} dt \right]^k = x^k e^{-kx^2/6} [1 + C_2^{(k)} x^4 + C_3^{(k)} x^6 + \dots]$$

TABLE III

Upper 5% points of  $u_n = |x_n/s|$  for ordered samples of sizes  $n$  with  $\nu$  degrees of freedom

$\nu \backslash n$	1	2	3	4	5	6	7	8		
5	2.57	3.09	3.40	3.62	3.78	3.92	4.04	4.14		
10	2.23	2.61	2.83	2.98	3.10	3.19	3.28	3.35		
15	2.13	2.47	2.67	2.81	2.91	2.99	3.06	3.12		
20	2.09	2.41	2.59	2.72	2.82	2.90	2.97	3.02		
24	2.06	2.38	2.56	2.68	2.78	2.84	2.91	2.96		
30	2.04	2.35	2.52	2.64	2.73	2.80	2.86	2.91		
40	2.02	2.32	2.49	2.60	2.69	2.76	2.82	2.86		
60	2.00	2.29	2.46	2.56	2.65	2.72	2.77	2.82		
120	1.98	2.26	2.43	2.53	2.61	2.68	2.73	2.77		
$\infty$	1.96	2.23	2.39	2.49	2.57	2.64	2.69	2.73		

(J. W. Tukey [9] states that some upper percentage points of the Studentized maximum modulus were computed by P. Nemenyi.)

we get

$$(4.8) \quad p(|x_n|) = n(\sqrt{2/\pi})^n e^{-(n+2)x_n^2/6} \sum_{j=0}^{\infty} C_j^{(n-1)} x_n^{2j+n-1}.$$

Since  $u_n = |x_n/s|$ , the distribution of  $u_n$  is given by

$$(4.9) \quad p(u_n) = n \left(\frac{2}{\pi}\right)^{n/2} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \sum_{j=0}^{\infty} C_j^{(n-1)} \left[ \frac{6}{(n+2)u_n^2 + 3\nu} \right]^{(n+2j+\nu)/2} \cdot u_n^{n+2j-1} \Gamma\left(\frac{n+2j+\nu}{2}\right).$$

It may be noted that in (4.8) and (4.9)  $C_0^{(n-1)} = 1$  and  $C_1^{(n-1)} = 0$ . The  $C$  coefficients are given by Pillai [7] (in his notation they are  $K$  coefficients). Using (4.9), the upper and lower 5 per cent points of  $u_n$  have been computed with the help of Tables of the Incomplete Beta Function, and are given in Tables III and IV for small values of  $n$ .

**5. Convergence of the series.** For examining the convergence of the different series developed in sections 3 and 4, let us start with series (4.7) for the case  $k = 1$ , given by

$$(5.1) \quad \int_0^x e^{-t^2/2} dt = x e^{-x^2/6} [1 + C_2^{(1)} x^4 + \dots].$$

TABLE IV

Lower 5% points of  $u_n = |x_n/s|$  for ordered samples of sizes  $n$  with  $\nu$  degrees of freedom

$\nu \backslash n$	1	2	3	4	5	6	7	8	9	10
1	.08	.29	.44	.55	.62	.68	.73	.78	.82	.86
2	.07	.29	.46	.57	.66	.73	.79	.84	.88	.92
3	.07	.29	.46	.59	.68	.76	.82	.87	.92	.96
4	.07	.29	.46	.60	.70	.78	.84	.90	.95	.99
5	.07	.29	.47	.60	.70	.79	.85	.91	.96	1.01
10	.06	.28	.47	.61	.71	.81	.89	.95	1.01	1.06
15	.06	.28	.47	.62	.73	.83	.91	.97	1.03	1.08
20	.06	.28	.47	.62	.73	.83	.91	.98	1.04	1.09
24	.06	.28	.47	.62	.74	.84	.92	.98	1.04	1.09
30	.06	.28	.47	.62	.74	.84	.92	.99	1.05	1.10
40	.06	.28	.47	.63	.74	.84	.92	.99	1.05	1.11
60	.06	.28	.47	.63	.75	.85	.93	1.00	1.06	1.11
120	.06	.28	.47	.63	.75	.85	.93	1.00	1.06	1.11
$\infty$	.06	.28	.47	.63	.75	.85	.93	1.00	1.06	1.11

If we expand  $e^{-t^2/2}$  as a power series and integrate term by term (assuming its validity, which is easily shown in this case), we get an expansion of the integral in the form

$$(5.2) \quad \int_0^x e^{-t^2/2} dt = \int_0^x \left( 1 - \frac{t^2}{2} + \frac{t^4}{8} - \dots \right) dt = x \left[ 1 - \frac{x^2}{6} + \frac{x^4}{40} - \dots \right].$$

As the first two terms in square brackets in (5.2) are contained in  $e^{-x^2/6}$ , the appropriateness of the series expansion (5.1) is immediately obvious. Since the integral

$$(5.3) \quad \int_{-\infty}^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt,$$

the expansion (3.2) follows from (5.1). An examination of the convergence of the series in (5.1) will thus be enough to show the convergence of the series (3.2). It can easily be shown [7] that the  $C$ 's follow the recurrence relation

$$(5.4) \quad 3(2i+1)C_i^{(1)} - C_{i-1}^{(1)} = (-1)^i / 3^{i-1} i!.$$

Hence

$$(5.5) \quad C_i^{(1)} = \left(\frac{2}{3}\right)^i \frac{i!}{(2i+1)!} \left[ 1 - 1 + \frac{3}{2!} - \frac{3 \cdot 5}{3!} \dots + (-1)^i \frac{3 \cdot 5 \dots (2i-1)}{i!} \right]$$

and

$$(5.6) \quad \frac{-C_i^{(1)}}{C_{i-1}^{(1)}} = \frac{1}{3(2i+1)} \left[ -1 + \frac{3 \cdot 5 \cdots (2i-1)}{i!A} \right]$$

where

$$(5.7) \quad A = \frac{3 \cdot 5 \cdots (2i-3)}{(i-1)!} - \frac{3 \cdot 5 \cdots (2i-5)}{(i-2)!} + \dots$$

It may be noticed that since the right hand side of (5.1) is an alternating series,  $-C_i^{(1)}/C_{i-1}^{(1)}$  is always positive. Moreover,  $A$  is also positive (except when  $A$  is the sum of the first two terms on the right side of (5.5) which is equal to zero). Since the second term in the square bracket of (5.6) is positive, the right side of (5.6) will be increased if we decrease  $A$ . Now if we neglect all the terms of  $A$  except the first two (where the sum of the neglected terms is positive), we decrease  $A$  and hence increase the right side of (5.6). In other words

$$(5.8) \quad \frac{-C_i^{(1)}}{C_{i-1}^{(1)}} < \frac{1}{3(2i+1)} \left[ -1 + \frac{(2i-3)(2i-1)}{i(i-2)} \right] = \frac{(i-1)^2}{i(i-2)(2i+1)}.$$

Hence when  $i$  is large

$$(5.9) \quad -C_i^{(1)}/C_{i-1}^{(1)} < 1/2i.$$

Again, if we retain the first four terms in the expression for  $A$  in (5.7), we get

$$(5.10) \quad \frac{-C_i^{(1)}}{C_{i-1}^{(1)}} < \frac{(i-1)(11i^3 - 78i^2 + 169i - 105)}{3i(i-2)(2i+1)(5i^2 - 29i + 39)}.$$

If  $i$  is large

$$(5.11) \quad -C_i^{(1)}/C_{i-1}^{(1)} < 11/30i.$$

The right side of (5.10) can be made smaller if we consider more terms in the approximation to  $A$ . Hence the series  $\sum_0^\infty |C_i^{(1)}|$  is absolutely convergent, and hence  $\sum_0^\infty C_i^{(1)}$  is convergent and the absolute value of the ratio of the  $i$ th to the  $(i-1)$ th term of the power series in (5.1) (with which we are really concerned) is less than  $11x^2/30i$  for large values of  $i$  (considering only the first four terms in (5.7) to approximate  $A$ ). Hence the series (5.1) is absolutely convergent and therefore the powers of the series are also convergent.

Now consider the series expansion in (3.2). For  $k=1$ , it can be shown that the sum of the terms involving even powers of  $x$  (which are all positive) is  $\frac{1}{2}$ . Hence from the absolute convergence of the series (5.1), the absolute convergence of (3.2) is immediate. It may be noticed that the series (5.1) is rather rapidly convergent, so that, for a relatively small  $x$ , only a few terms of the series will suffice for any degree of accuracy desired in practice.

**6. Acknowledgment.** The authors wish to express their indebtedness to Professors R. C. Bose and S. N. Roy for suggesting this problem and for their guidance in the preparation of this paper.

## REFERENCES

- [1] S. N. ROY AND R. C. BOSE, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513-536.
- [2] G. E. P. BOX, "Nonnormality and tests on variances," *Biometrika*, Vol. 40 (1953), pp. 318-335.
- [3] G. E. P. BOX, "A note on regions for tests of kurtosis," *Biometrika*, Vol. 40 (1953), pp. 465-468.
- [4] JOYCE M. MAY, "Extended and corrected tables of the upper percentage points of the 'Studentized' range," *Biometrika*, Vol. 39 (1952), pp. 192-193.
- [5] K. R. NAIR, "The distribution of the extreme deviate from the sample mean and its Studentized form," *Biometrika*, Vol. 35 (1948), pp. 118-144.
- [6] K. PEARSON, *Tables of the Incomplete Beta Function*, Cambridge University Press, (1934).
- [7] K. C. S. PILLAI, "On the distributions of midrange and semirange in samples from a normal population," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 100-105.
- [8] K. C. S. PILLAI, "On the distribution of 'Studentized' range," *Biometrika*, Vol. 39 (1952), pp. 194-195.
- [9] J. W. TUKEY, "The problem of multiple comparisons," Preliminary report (unpublished), Princeton University, p. 169.