ON THE DISTRIBUTION OF AN ANALOGUE OF STUDENT'S t

By K. C. S. Pillai
University of Travancore, Trivandrum

1. Introduction. The independence of the sample range and the mean in random samples from a normal population has been already established [1], [2]. Using this property of independence, J. F. Daly [1] has shown that with the help of the distribution law of the sample range, w, tabulated by Pearson and Hartley [3] the probability distribution of an analogue of Student's t-test given by $G = (\bar{x} - a)/w$ can be studied, where $\bar{x}$ is the mean and a the location parameter. E. Lord [2] has prepared, by quadrature, exhaustive tables of levels of significance of G for sample size varying from 2 to 20 corresponding to the probabilities 0.10, 0.05, 0.02, 0.01, 0.002, and 0.001. An approximation to the distribution of u = const. G has been studied by P. B. Patnaik [4] and the power of the u-test has been investigated by Lord [5]. E. S. Pearson [6] has examined the effect of nonnormality on u-test involving range. The purpose of the present note is to develop the probability distribution of G as a series whose terms reduce to Beta functions and to observe the efficiency of the G-test.

2. The distribution of G. For a sample of size n from a normal population with mean a and standard deviation unity the distribution of $\bar{x}$ is given by

$$p(\bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{1}{2}n(\bar{x}-a)^2}, \tag{1}$$

and the distribution of the semi-range, W, has been shown by the author [7] to be given by

$$p(W) = ke^{-(n+4)W^2/6} \sum_{i=0}^{\infty} C_i W^{n+2i-2}, \tag{2}$$

where

$$k = \frac{n-1}{2^4\Gamma(5)n^{7/2}} (\sqrt{2/\pi})^{n-1} \tag{3}$$

and C_i are functions of n. The distribution (2) is useful for small sample sizes, and C_i coefficients have been computed for sample sizes up to 8, using an appropriate expansion [8] of the normal probability integral.

Since $\bar{x}$ and W are independently distributed,

$$p(\bar{x}, W) = \sqrt{\frac{n}{2\pi}}\, ke^{-\frac{1}{2}n(\bar{x}-a)^2} e^{-(n+4)W^2/6} \sum_{i=0}^{\infty} C_i W^{n+2i-2}. \tag{4}$$

Putting

$$G = \frac{\bar{x} - a}{2W}, \tag{5}$$

we have

$$(6) \qquad p(G, W) = k\sqrt{\frac{2n}{\pi}}\, e^{-[2nG^2+(n+4)/6]W^2} \sum_{i=0}^{\infty} C_i W^{n+2i-1},$$

and hence, integrating out W from (6),

$$(7) \qquad p(G) = k\sqrt{\frac{n}{2\pi}} \sum_{i=0}^{\infty} \frac{C_i \Gamma[\frac{1}{2}n + i]}{[2nG^2 + (n+4)/6]^{\frac{1}{2}n+i}}.$$

The terms of (7) are easily seen to reduce to Beta functions and hence the probability integrals of G can be evaluated with the help of tables of the incomplete Beta function.

3. Confidence intervals. For a given probability level α and sample size n, using distribution (7), the limit, λ, may be obtained from the relation

$$(8) \qquad \int_0^{\lambda} p(G)\, dG = \frac{1-\alpha}{2},$$

and the average length of confidence interval for the location parameter a based on G will be given by

$$(9) \qquad I(G) = 2\lambda E(w).$$

$E(w)$ for sample size ranging from 2 to 1000 has been computed by L. H. C. Tippett [9], and with the help of Tippett's table of mean range confidence intervals, $I(G)$, for different values of α and n have been obtained. Values of $\frac{1}{2\sigma} I(G)$ are given in Table I. The average length of confidence interval $I(G)$ for a particular value of α and n may be compared with the corresponding average length of confidence interval based on Student's t-test, given by

$$(10) \qquad I(t) = 2\lambda' \frac{E(S)}{\sqrt{n}},$$

where

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)$$

and λ' is given by

$$(11) \qquad \int_0^{\lambda'} p(t)\, dt = \frac{1-\alpha}{2},$$

$p(t)$ being the probability function of t.

Values of $\frac{1}{2\sigma} I(t)$ for different values of α and n are given in Table I. The comparative efficiency of the G and t-tests may be studied by observing

$$(12) \qquad E = \frac{I(t)}{I(G)},$$

and the values of E have been provided in Table I.

TABLE I[1]

Confidence interval for a based on G and t-tests

n	α	$\frac{I(G)}{2\sigma}$	$\frac{I(t)}{2\sigma}$	E
3	.100	1.4979	1.4941	.9975
	.050	2.2071	2.2017	.9976
	.010	5.0913	5.0783	.9974
	.001	16.2148	16.1675	.9971
4	.100	1.0891	1.0839	.9952
	.050	1.4761	1.4658	.9930
	.010	2.7093	2.6907	.9931
5	.100	0.9025	0.8963	.9931
	.050	1.1792	1.1670	.9897
	.010	1.9608	1.9354	.9870
6	.050	1.0112	0.9987	.9876
	.010	1.5916	1.5663	.9841
7	.050	0.9006	0.8873	.9852
	.010	1.3711	1.3442	.9804
8	.050	0.8200	0.8070	.9841
	.010	1.2214	1.1938	.9774
10	.050	0.7078	0.6957	.9829
	.010	1.0248	0.9996	.9754
12	.050	0.6321	0.6211	.9826
	.010	0.9026	0.8765	.9711
14	.050	0.5791	0.5663	.9779
	.010	0.8142	0.7897	.9699
17	.050	0.5167	0.5062	.9797
	.010	0.7212	0.6974	.9670
20	.050	0.4706	0.4619	.9815
	.010	0.6536	0.6313	.9659

[1] In computing I(G) for $n > 8$ the values of λ have been taken from Lord's Tables [2].

It is interesting to note from Table I that in small samples the average length of confidence interval for the G test compares favourably with that for the t-test. The test has the advantage that it is easier for computation.

REFERENCES

[1] J. F. DALY, "On the use of sample range in an analogue of Student's t-test," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71–74.
[2] E. LORD, "The use of range in place of standard deviation in the t-test," *Biometrika*, Vol. 34 (1947), pp. 41-67.
[3] E. S. PEARSON AND H. O. HARTLEY, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), pp. 301–310.
[4] P. B. PATNAIK, "The use of mean range as an estimator of variance in statistical tests," *Biometrika*, Vol. 37 (1950), pp. 78–87.
[5] E. LORD, "The power of the modified t-test (u-test) based on range," *Biometrika*, Vol. 37 (1950), pp. 64–77.
[6] E. S. PEARSON, "Some notes on the use of range," *Biometrika*, Vol. 37 (1950), pp. 88–92.
[7] K. C. S. PILLAI, "Some notes on ordered samples from a normal population," *Sankhyā*, Vol. 11 (1951), pp. 23–28.
[8] K. C. S. PILLAI, "On the distributions of midrange and semi-range in samples from a normal population," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 100–105.
[9] L. H. C. TIPPETT, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925), pp. 364–387.

A PROPERTY OF SOME TYPE A REGIONS

BY HERMAN CHERNOFF

University of Illinois

1. Summary. In a test of an hypothesis one may regard a sample in the critical region as evidence that the hypothesis is false. Let us assume that for some reason it is desired to increase the critical size of the test, i.e., to make rejection of the hypothesis more probable. Then one may expect that an observation which led to rejection in the first test should still lead to rejection in the new test. In other words, one should expect $W_\alpha \supset W_{\alpha'}$ if $\alpha > \alpha'$, where W_α is the critical region for the test of size α. An example is given where regions of type A[1] are uniquely specified except for sets of measure zero, but fail to have this property.

2. Example. We shall consider type A regions for the hypothesis $\theta = 0$ where our sample consists of one observation with density

$$p(x, \theta) = (2\pi)^{-\frac{1}{2}}(1 + \theta)^{\frac{1}{2}} \exp[-(x - \theta)^2(1 + \theta)/2] \qquad \text{for } \theta > -1.$$

[1] Regions of type A were introduced by Neyman and Pearson (see [1]).