AN APPROXIMATION TO THE DISTRIBUTION OF Q (A VARIATE RELATED TO THE NON-CENTRAL t)[1]

BY D. HOGBEN, R. S. PINKHAM[2] AND M. B. WILK[3]

Rutgers—The State University

1. Introduction. If W is a random normal variate with mean θ and variance 1 and Z^2 is independently distributed as chi-squared with n degrees of freedom, then the random variable Q, with non-centrality θ and n degrees of freedom, is defined by

$$Q = W/(W^2 + Z^2)^{\frac{1}{2}}, \qquad \text{(all square roots positive)}. \tag{1}$$

Properties of Q, including the probability density function and the moments, have been studied elsewhere by Hogben et al. (1964).

It is apparent that $n^{\frac{1}{2}}Q/(1 - Q^2)^{\frac{1}{2}}$ is distributed as non-central t with non-centrality θ and n degrees of freedom. For some discussion of the non-central t variate see for example Fisher (1931), Johnson and Welch (1940) and Resnikoff and Lieberman (1957). The probability integral of Q is of course implicitly defined by that of the non-central t, but the tables of the non-central t of Resnikoff and Liberman provide no information for small values of θ. The direct definition of the distribution function of Q in terms of the density is not conveniently manageable except with extensive tabulation.

Miss van Eeden (1961) has compared six types of approximation for the non-central t-distribution by tabling exact and approximate percentage points for several probability levels, degrees of freedom and non-centrality values. Because percentage points are used, further numerical study would be necessary to compare her results with the results in Table 1. None of the approximations considered by Miss van Eeden has the correct limiting distribution when $\theta = 0$.

The purpose here is to propose as an approximation to the distribution of Q that of a linearly transformed beta variate and to present some results on the adequacy of the approximation. The approximation, which is especially good for small θ, implicitly and easily yields approximate values for the non-central t. The present interest in the random variate Q arose from its relevance in another study, Hogben et al. (1962).

2. An approximation. Let X be distributed as beta (see for example Mood (1950)) with parameters a and b. Then, the suggested approximation to the

Received 6 November 1961; revised 22 July 1963.

[1] This research was supported by the Office of Naval Research under contract Nonr 404(16). Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2, 3] Now at Bell Telephone Laboratories.

TABLE 1

Comparison of true with approximate probability levels of Q

$\hat{\epsilon}$		Quantity in parentheses is $(\epsilon - \hat{\epsilon})10^4$			
ϵ	θ	$n = 4$	$n = 9$	$n = 16$	$n = 36$
.975	0.1	.9762 (−12)	.9755 (−5)	.9753 (−3)	.9751 (−1)
	0.8	.9818 (−68)	.9787 (−37)	.9773 (−23)	.9761 (−11)
	2.0	.9801 (−51)	.9800 (−50)	.9788 (−38)	.9773 (−23)
	5.0	.9758 (−8)	.9771 (−21)	.9778 (−28)	.9779 (−29)
.8	0.1	.7999 (1)	.7998 (2)	.7999 (1)	.7999 (1)
	0.8	.7921 (79)	.7965 (35)	.7982 (18)	.7993 (7)
	2.0	.7769 (231)	.7888 (112)	.7940 (60)	.7977 (23)
	5.0	.7905 (95)	.7908 (92)	.7923 (77)	.7955 (45)
.2	0.1	.2003 (−3)	.1999 (1)	.1999 (1)	.1999 (1)
	0.8	.2102 (−102)	.2018 (−18)	.2002 (−2)	.1997 (3)
	2.0	.2330 (−330)	.2090 (−90)	.2029 (−29)	.2001 (−1)
	5.0	.2149 (−149)	.2091 (−91)	.2054 (−54)	.2016 (−16)
.025	0.1	.0264 (−14)	.0257 (−7)	.0254 (−4)	.0252 (−2)
	0.8	.0380 (−130)	.0306 (−56)	.0280 (−30)	.0263 (−13)
	2.0	.0524 (−274)	.0373 (−123)	.0321 (−71)	.0281 (−31)
	5.0	.0347 (−97)	.0338 (−88)	.0324 (−74)	.0300 (−50)

distribution of Q is the distribution of

$$Y = 2X - 1$$

where a and b are chosen so that the first two moments of Q and Y are equal. Thus, if $E(\ )$ and $V(\ )$ denote expectation and variance, respectively, and if

$$E(Q) = \mu(\theta, n) = \mu \quad \text{and} \quad V(Q) = \sigma^2(\theta, n) = \sigma^2,$$

then

$$E(Y) = 2E(X) - 1 = (a - b)/(a + b) = \mu \tag{2}$$

and

$$V(Y) = 4V(X) = 4ab/(a + b)^2 (a + b + 1) = \sigma^2. \tag{3}$$

Solving Equations (2) and (3) simultaneously yields

$$a = (1 + \mu)(1 - \mu^2 - \sigma^2)/2\sigma^2 \quad \text{and} \quad b = (1 - \mu)(1 - \mu^2 - \sigma^2)/2\sigma^2. \tag{4}$$

Values of μ and σ^2 are given in Table 1 of Hogben et al. (1964).

The range of Q is $(-1, 1)$ and since X has range $(0, 1)$, the range of Y is the same as that of Q. When $\theta = 0$ the approximation is exact with $a = b = \frac{1}{2}n$. Hence, good approximations are anticipated for small θ. On the other hand, as $\theta \to \infty$, $E(Q) \to 1$ and $V(Q) \to 0$; while as $b \to 0$, $E(Y) \to 1$ and $V(Y) \to 0$.

Hence, it would seem that the approximation might also do well for large θ. Furthermore, it would seem reasonable to expect the approximation to be good for large n, since as $n \to \infty$, $E(Q) \to 0$, $V(Q) \to 0$ while the mean and variance of Y approach zero as $a = b \to 0$.

3. Accuracy of approximation. If

$$\epsilon = \Pr(Q > \lambda), \tag{5}$$

then the above approximation to ϵ is

$$\hat{\epsilon} = \Pr\{X_{a,b} > \tfrac{1}{2}(1 + \lambda)\}, \tag{6}$$

where $X_{a,b}$ denotes a beta variate with parameters a and b. The exact distribution of Q can be obtained from that of the non-central t by the relationship

$$\epsilon = \Pr\{t > n^{\frac{1}{2}}\lambda/(1 - \lambda^2)^{\frac{1}{2}}\}. \tag{7}$$

A check of the accuracy of the approximation was made for all combinations of $\theta = 0.1, 0.8, 2.0, 5.0$; $n = 4, 9, 16, 36$; $\epsilon = 0.975, 0.8, 0.2, 0.025$. The results of this study are given in Table 1.

The table was constructed by fixing ϵ, finding $t_0 = n^{\frac{1}{2}}\lambda/(1 - \lambda^2)^{\frac{1}{2}}$ such that $\epsilon = \Pr(t > t_0)$ and then finding $\hat{\epsilon}$ from

$$\hat{\epsilon} = \Pr(X_{a,b} > x) = 1 - I_x(a, b),$$

where

$$I_x(a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1 - t)^{b-1}\, dt, \tag{8}$$

and

$$x = \tfrac{1}{2}[1 + t_0/(n + t_0^2)^{\frac{1}{2}}]. \tag{9}$$

The values of t_0 were found from Table IV of Johnson and Welch (1940) using their iterative procedure. An attempt was made to use the tables of Pearson (1932) to find $\hat{\epsilon}$, but some cases (e.g. $n = 36$ and $\theta = 5.0$) were outside the range of those tables and in other cases (a close to 50 and x close to 1) Everett's trivariate interpolation formula, as given by Pearson, gave insufficient accuracy.

Hence, $\hat{\epsilon}$ was computed on an IBM "650" as follows. The integrand of the incomplete beta function was expanded and the infinite series integrated term by term to give

$$I_x(a, b) = [\Gamma(a + b)/\Gamma(a)\,\Gamma(b)] \sum_{i=0}^{\infty} s_i,$$

where

$$s_0 = x^a/a, \qquad s_i = -(b - i)\,[x(a + i - 1)/i(a + i)]\, s_{i-1}, \qquad i > 0.$$

To gain speed and accuracy in computing, $I_x(a, b)$ was computed as shown if

x was less than 0.5, but if x was greater than 0.5 $I_x(a, b)$ was computed from $I_x(a, b) = 1 - I_{1-x}(b, a)$. For large values of θ and n it was necessary to use double precision.

The complete beta function was evaluated by computing each complete gamma function, using Hastings' (1955) approximation for small values of the argument and Stirling's approximation for large values of the argument with the first five terms of the remainder as given by Huntington (1940).

The computations are exact to the number of digits shown, except for possible errors introduced by the use of Table IV of Johnson and Welch (1940), either directly or through the use of linear interpolation. From their comments on the accuracy of their table, it is estimated that $\hat{\epsilon}$ has a maximum numerical error of 5 units in the fourth decimal place, but in most cases the results will be correct or at most in error by one unit in the fourth decimal place.

4. Discussion. The results of Table 1 indicate that the approximation is good for small values of θ as contended in Section 2. Further, the absolute error decreases with n almost everywhere in the table and the quality of the approximation is poorest for intermediate values of θ, improving for larger values of θ.

This procedure provides an easily available approximation to the distribution function of Q, or any monotone function of it, employing values of the beta distribution (see Pearson (1932)) and values of the mean and variance of Q for given θ and n (see Hogben et al. (1964)).

In particular, this approximation provides values for the probability integral of the non-central t (or any monotone function of it) for small non-centrality values—a range not covered by Resnikoff and Lieberman (1957).

5. Acknowledgments. The authors thank M. Tanne and R. Thayer for their helpful computational assistance and are grateful to the referee for various helpful suggestions and references.

REFERENCES

Fisher, R. A. (1931). Properties of *Hh* functions. *British Association Mathematical Tables*, **1** XXVI–XXXV, Cambridge University Press, London.

Hastings, C., Jr. (1955). *Approximations for Digital Computers*. Princeton University Press, Princeton.

Hogben, D., Pinkham, R. S., and Wilk, M. B. (1962). Some properties of Tukey's test for non-additivity (Abstract). *Ann. Math. Statist.* **33** 1492–1493.

Hogben, D., Pinkham, R. S., and Wilk, M. B. (1964). The moments of a variate related to the non-central *t*. *Ann. Math. Statist.* **35** 298–314.

Huntington, E. V. (1940). Stirling's formula with remainder. *Biometrika*, **31** 390.

Johnson, N. L. and Welch, B. L. (1940). Applications of the non-central *t*-distribution. *Biometrika*, **31** 362–389.

Mood, A. M. (1950). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Pearson, K. (1932). *Tables of the Incomplete Beta Function*. Cambridge University Press, London.

Resnikoff, G. J. and Lieberman, G. J. (1957). *Tables of the Non-central t-Distribution*. Stanford University Press, Stanford.

van Eeden, Constance (1961). Some approximations to the percentage points of the non-central *t*-distribution. *Rev. Inst. Internat. Statist.*, **29** 4–31.