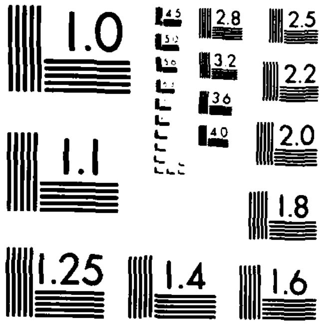
AD-A149 457 A NEW METHOD OF APPROXIMATING NORMAL AND T-TAIL AREAS 1/1
(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER
A P SOMS SEP 84 MRC-TSR-2750 DAAG29-80-C-0041

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AD-A149 457

MRC Technical Summary Report #2750

A NEW METHOD OF APPROXIMATING NORMAL AND t-TAIL AREAS

Andrew P. Soms

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1984

(Received July 20, 1984)

Approved for public release
Distribution unlimited

DTIC
ELECTE
JAN 22 1985
S D
D

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

DTIC FILE COPY

85 01 15

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

A NEW METHOD OF APPROXIMATING NORMAL AND t-TAIL AREAS

Andrew P. Soms*

Technical Summary Report #2750
September 1984

ABSTRACT

The bounds of Boyd (1959) and Soms (1980, 1984) for the tail areas of the normal and t-distributions are used to obtain a new method of approximating the tail areas. The absolute and relative errors are given.

AMS (MOS) Subject Classifications: 62Q05

Key Words: absolute error, normal percentile, normal tail area, relative error, t-percentile, t-tail area

Work Unit Number 4 (Statistics and Probability)

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED

*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the University of Wisconsin-Milwaukee.

SIGNIFICANCE AND EXPLANATION

A reasonably accurate and simple method of evaluating normal and t-tail areas and percentiles is described. This approach controls both absolute and relative errors, unlike the usual methods which only control the absolute error. The control of both errors is of importance in calculating Bonferroni percentiles and Bonferroni descriptive probability levels, which are useful in simultaneous inference. For example, a nonspecialist in statistics, unaware of the limitations of standard routines, might report a Bonferroni descriptive probability level that had a relative error of 100% or more.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

A NEW METHOD OF APPROXIMATING NORMAL AND t-TAIL AREAS

Andrew P. Soms*

1. THE METHOD

We begin by introducing notation and stating the main results of Boyd (1959) and Soms (1980, 1984). Boyd (1959) showed that if

$$\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2), \quad \bar{F}(x) = \int_x^\infty \phi(t)dt ,$$

and $R_x = \bar{F}(x)/\phi(x)$, $x > 0$, then

$$p(x,\gamma_{min}) < R_x < p(x,\gamma_{max}) ,$$

where $p(x,\gamma) = (\gamma + 1)/[(x^2 + (2/\pi)(\gamma + 1)^2)^{1/2} + \gamma x]$, $\gamma_{max} = 2/(\pi - 2)$, $\gamma_{min} = \pi - 1$, and the bounds are the best possible in the class $\{p(x,\gamma), \gamma > -1\}$. This is also discussed in Johnson and Kotz (1970, Ch. 33). Soms (1980), 1984) extended the above results and showed that if for arbitrary real $k > 0$ and $x > 0$,

$$f_k(t) = c_k(1 + t^2/k)^{-(k+1)/2}, \quad c_k = \frac{\Gamma((k + 1)/2)}{\Gamma(k/2)(\pi k)^{1/2}} ,$$

$$\bar{F}_k(x) = 1 - F_k(x) = \int_x^\infty f_k(t)dt ,$$

$$R_k(x) = \bar{F}_k(x)/[(1 + x^2/k)f_k(x)] ,$$

for $k > 2$, $\gamma_{max} = 4c_k^2/(1 - 4c_k^2)$ and $\gamma_{min} = \frac{k}{2(k + 2)c_k^2} - 1$,

and for $k < 2$, γ_{min} and γ_{max} are interchanged, and

*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the University of Wisconsin-Milwaukee.

$$p(x,\gamma) = \frac{1+\gamma}{(x^2 + 4c_k^2(1+\gamma)^2)^{1/2} + \gamma x} ,$$

then

$$p(x,\gamma_{min}) < R_k(x) < p(x,\gamma_{max}) ,$$

or equivalently,

$$\left(1 + \frac{x^2}{k}\right) f_k(x) P(x,\gamma_{min}) < \bar{F}_k(x) < \left(1 + \frac{x^2}{k}\right) f_k(x) p(x,\gamma_{max}) ,$$

and the bounds again are best in the same sense as for the normal. It was also shown there that if $k = 2$, $\gamma_{max} = \gamma_{min} = \gamma_2$ and $R_k(x) = p(x,\gamma_2)$.

The numerical properties of these bounds are discussed in the above references. The important fact to be noted here is that the bounds control both absolute and relative error. Using the bounds as a starting point we now develop a simple method of evaluating normal and t-tail areas that controls both absolute and relative error, as opposed to the usual methods, which generally only control absolute error.

We consider estimates of the tail area of the form

$$\left(\frac{a+bx}{c+dx}\right) p(x,\gamma_{min}) \phi(x) + \left(1 - \frac{a+bx}{c+dx}\right) p(x,\gamma_{max}) \phi(x) \qquad (1.1)$$

for the tail area of the normal and

$$\left(\frac{a+bx}{c+dx}\right) p(x,\gamma_{min}) g_k(x) + \left(1 - \frac{a+bx}{c+dx}\right) p(x,\gamma_{max}) g_k(x) \qquad (1.2)$$

for the tail area of the t, where $g_k(x) = (1 + x^2/k) f_k(x)$. We want the estimates to lie between the upper and lower bounds for the tail area and be strictly decreasing functions of x and therefore impose the added restrictions that

$$bc > ad$$

and

$$0 < \frac{a + bx}{c + dx} < 1 ,$$

all $x > 0$. Since $f(0) = a/c$, we may, without loss of generality, assume that $c = 1$ and so our weight functions f are of the type

$$f(x) = \frac{a + bx}{1 + dx} , \tag{1.3}$$

where $0 < a < 1$, $d > 0$, $b > ad$ and $b/d < 1$. We then seek that particular choice of f which minimizes the absolute error. A direct computer search led to

$$f(x) = \frac{.71x}{1 + .71x} \tag{1.4}$$

for the normal and

$$f(x) = \frac{b_k x}{1 + b_k x} , \tag{1.5}$$

$$b_k = .71 + 1.8/k - .19/k^2 , \tag{1.6}$$

for the t, where, as noted before, k is the degrees of freedom. (1.6) was obtained by finding the optimal constants for $k = 25, 10, 5, 3, 1.5, 1, .5$ and fitting a constrained (constant equal to .71) regression line. However, in the interests of simplicity, for $k < 2$, we did not interchange γ_{min} and γ_{max} and so (1.5) and (1.6) are understood to apply for all k with γ_{min} and γ_{max} defined as for $k > 2$. Numerical evidence indicates that, at least for $k = 1$, the above optimal estimate is still a decreasing function of x.

The maximum absolute and relative errors of the optimal estimates are remarkably constant over the range $1 < k < \infty$ and hence we only give the normal figures. For (1.4), the maximum absolute error is $.66 \times 10^{-4}$ and the maximum relative error is $.97 \times 10^{-3}$. We emphasize once more, that, unlike the usual methods, which generally control only absolute error and give misleading results for large x or small tail areas, the above controls both absolute and relative error and hence can be used to calculate ordinary and Bonferroni descriptive levels and ordinary and Bonferroni percentiles.

As a check, we calculated the standard textbook table of the normal, given, e.g., in Brown and Hollander (1977) and found at most a difference of 1 in the fourth decimal place. We also compared the small normal percentiles given in Abramowitz and Stegun (1965, p. 977) to the ones obtained from (1.4) and after rounding both to three decimal places found that there was at most a difference of 1 in the third decimal place. Similar results apply to the t.

2. CONCLUDING REMARKS

We have given a method of calculating normal and t-tail areas and percentiles which controls both absolute and relative errors and is simple to implement even on small computers. Four short FORTRAN programs that do this are given in the Appendix. The programs are completely self-contained, using a short FORTRAN subroutine to calculate c_k based on Stirling's formula for large k and an approximation to the gamma function given in Abramowitz and Stegun (1965, p. 257) for small k.

APPENDIX

The listings of the four programs referred to in the paper are given below. If $k < 2$ and α is small, the t-percentile program may be slow.

```
      SUBROUTINE NORT(X,TP)
C NORMAL TAIL AREA FROM X TO INFINITY, X NON-NEGATIVE
      DATA PIE/3.1415927/
      PHI(X)=1./SQRT(2*PIE)*EXP(-X**2/2.)
      P(X,G)  =(1.+G)/((X**2+2./PIE*(1.+G)**2)**.5+G*X)
      PAMIN(X)=.71*X/(1.+.71*X)
      TPL(X,GL,GU)=(PAMIN(X)*P(X,GL)+(1.-PAMIN(X))*P(X,GU))*PHI(X)
      GMIN=PIE-1.
      GMAX=2./(PIE-2.)
      TP=TPL(X,GMIN,GMAX)
      RETURN
      END

      SUBROUTINE NORTI(ALPHA,XALP)
C UPPER NORMAL PERCENTILE XALP CORRESPONDING TO TAIL
C PROBABILITY ALPHA, ALPHA BETWEEN 0 AND .5
      DATA EPS/.000005/
      DATA PIE/3.1415927/
      PHI(X)=1./SQRT(2*PIE)*EXP(-X**2/2.)
      P(X,G)  =(1.+G)/((X**2+2./PIE*(1.+G)**2)**.5+G*X)
      PAMIN(X)=.71*X/*1.+.71*X)
      TPL(X,GL,GU)=(PAMIN(X)*P(X,GL)+(1.-PAMIN(X))*P(X,GU))*PHI(X)
      GMIN=PIE-1.
      GMAX=2./(PIE-2.)
      X=1.
102   TP=TPL(X,GMIN,GMAX)
      IF (TP.LT. ALPHA) GO TO 103
      X=X+1
      GO TO 102
103   BL=X-1.
      BU=X
104   X=(BL+BU)/2.
      TP=TPL(X,GMIN,GMAX)
      IF (TP.LT. ALPHA) BU=X
      IF (TP.GE. ALPHA) BL=X
      IF (ABS(BU-BL).LE. EPS) GO TO 105
      GO TO 104
105   XALP=(BL+BU)/2.
      RETURN
      END
```

```
      SUBROUTINE TT(DF,X,TP)
C T TAIL AREA FROM X TO INFINITY, X NON-NEGATIVE
      P(X,G,D)=(1.+G)/((X**2+4.*CK(D)**2*(1.+G)**2)**.5+G*X)
      FKX(X,D)=CK(D)*(1.+X**2/D)**((-D+1.)/2.)
      GMAX(X)=4.*CK(X)**2/(1.-4.*CK(X)**2)
      GMIN(X)=X/(2.*(X+2.)*CK(X)**2)-1.
      TCON(DF)=.71+1.8/DF+1.9/DF**2
      PAMIN(DF,X)=TCON(DF)*X/(1.+TCON(DF)*X)
      TPL(DF,X,GL,GU)=(PAMIN(DF,X)*P(X,GL,DF)+(1.- PAMIN(DF,X))*
     1P(X,GU,DF))*FKX(X,DF)
      TP=TPL(DF,X,GMIN(DF),GMAX(DF))
      RETURN
      END

      SUBROUTINE TTI(DF,ALPHA,XALP)
C UPPER T PERCENTILE XALP CORRESPONDING TO TAIL PROBABILITY ALPHA,
C ALPHA BETWEEN 0 AND .5
      P(X,G,D)=(1.+G)/((X**2+4.*CK(D)**2*(1.+G)**2)**.5+G*X)
      FKX(X,D)=CK(D)*(1.+X**2/D)**((-D+1.)/2.)
      GMAX(X)=4.*CK(X)**2/(1.-4.*CK(X)**2)
      GMIN(X)=X/(2.*(X+2.)*CK(X)**2)-1.
      TCON(DF)=.71+1.8/DF+1.9/DF**2
      PAMIN(DF,X)=TCON(DF)*X/(1.+TCON(DF)*X)
      TPL(DF,X,GL,GU)=(PAMIN(DF,X)*P(X,GL,DF)+(1.- PAMIN(DF,X))*
     1P(X,GU,DF))*FKX(X,DF)
      DATA EPS/.000005/
      X=1.
102   TP=TPL(DF,X,GMIN(DF),GMAX(DF))
      IF (TP.LT. ALPHA) GO TO 103
      X=X+1.
      GO TO 102
103   BL=X-1.
      BU=X
104   X=(BL+BU)/2.
      TP=TPL(DF,X,GMIN(DF),GMAX(DF))
      IF (TP.LT. ALPHA) BU=X
      IF (TP.GE. ALPHA) BL=X
      IF(ABS(BU-BL).LE. EPS) GO TO 105
      GO TO 104
105   XALP=(BL+BU)/2.
      RETURN
      END

      FUNCTION CK(X)
      ALNG(X)=(X-.5)*ALOG(X)-X+1/(12.*X)-1/(360.*X**3)
      GAMMA(T)=1./T-.57719165+.98820589*T-.89705694*T**2+.91820686*
     1T**3-.75670408*T**4+.48219939*T**5-.19352782*T**6+.03586834*
     2T**7
      DATA PIE/3.1415927/
      Z1=(X+1)/2.
      Z2=X/2.
```

```
      IF (Z2 .GT. 10.) GO TO 11
      TEMP1=1.
201   IF (Z1 .LE. 1.) GO TO 202
      TEMP1=TEMP1*(Z1-1.)
      Z1=Z1-1.
      GO TO 201
202   TEMP1=TEMP1*GAMMA(Z1)
      TEMP2=1.
203   IF (Z2 .LE. 1.) GO TO 204
      TEMP2=TEMP2*(Z2-1.)
      Z2=Z2-1.
      GO TO 203
204   TEMP2=TEMP2*GAMMA(Z2)
      CK=TEMP1/TEMP2/SQRT(PIE*X)
      RETURN
 11   CK=EXP(ALNG((X+1.)/2.)-ALNG(X/2.)-.5*ALOG(PIE*X))
      RETURN
      END
```

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1965), Hanc . of Mathematical Functions, National Bureau of Standards, Washington, D.C.

Brown, W. and Hollander, M. (1977), Statistics A Biomedical Introduction, John Wiley & Sons, New York.

Johnson, N. L., and Kotz, S. (1970), Continuous Univariate Distributions-2, Houghton Mifflin Co., Boston.

Boyd, A. V. (1959), "Inequalities for Mills' Ratio," Reports of Statistical Applications Research, Japanese Union of Scientists and Engineers, 6, 44-6.

Soms, A. P. (1980), "Rational Bounds for the t-Tail Area," Journal of the American Statistical Association, 75, 438-40.

Soms, A. P. (1984), "A Note on an Extension of Rational Bounds for the t-Tail Area to Arbitrary Degrees of Freedom," Communications in Statistics - Theory and Methods, 13, 887-891.

APS/ed

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2750	2. GOVT ACCESSION NO. AD-A149457	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) A NEW METHOD OF APPROXIMATING NORMAL AND t-TAIL AREAS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Andrew P. Soms		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE September 1984
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

absolute error, normal percentile, normal tail area, relative error, t-percentile, t-tail area

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

The bounds of Boyd (1959) and Soms (1980, 1984) for the tail areas of the normal and t-distributions are used to obtain a new method of approximating the tail areas. The absolute and relative errors are given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

END

FILMED

2-85

DTIC