A New Method of Evaluating Normal and t-Tail Areas

Andrew P. Soms*

Department of Mathematical Sciences
University of Wisconsin-Milwaukee, Milwaukee, WI 53201
Mathematics Research Center, 610 Walnut Street
Madison, WI 53705

Key Words and Phrases: absolute error, Bonferroni percentiles, normal distribution, relative error, t-distribution.

Abstract

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

*This research was supported by the Mathematics Research Center, University of Wisconsin-Madison, under Contract No. DAAG29-80-C-0041 and the University of Wisconsin-Milwaukee.

1. The Method

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

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

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

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

where $p(x,\gamma) = (\gamma + 1)/[(x^2 + (2/\pi)(\gamma + 1)^2)^{\frac{1}{2}} + \gamma x]$, $\gamma_{max} = 2/(\pi-2)$, $\gamma_{min} = \pi - 1$, 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 (1980a, 1980b) extended the above results and showed that if for arbitrary real $k > 0$ and $x > 0$,

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

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

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

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

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

then

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

or equivalently,

$$(1+\frac{x^2}{k})f_k(x)p(x,\gamma_{min}) < \overline{F}_k(x) < (1+\frac{x^2}{k})f_k(x)p(x,\gamma_{max}) \quad ,$$

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

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

for the tail area of the normal and

$$(\frac{a+bx}{c+dx})p(x,\gamma_{min})f_k(x) + (1 - \frac{a+bx}{c+dx})p(x,\gamma_{max})f_k(x) \tag{1.2}$$

for the tail area of the t. 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

$$0 \leq \frac{a+bx}{c+dx} \leq 1 \quad , \quad \text{all } x \geq 0.$$

Since $f(0) = \frac{a}{c}$, we may, without loss of generality, assume that $c = 1$ and so our weight functions f are of the type

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

where $0 \leq a \leq 1$, $d > 0$, $bc > ad$, and $\frac{b}{d} \leq 1$. 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

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

$$b_k = .70 + 1.82/k - .2/k^2 \quad , \tag{1.6}$$

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 regression line to them. However, in the interests of simplicity, for $k \leq 2$, 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 $1 \leq k \leq \infty$ and hence we only give the normal figures. For (1.4), the maximum absolute error is $.66 \cdot 10^{-4}$ and the maximum relative error is $.97 \cdot 10^{-3}$. We emphasize once more, that, unlike the usual methods, which generally control only absolute error, 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 which controls both absolute and relative errors. The listings of the short FORTRAN programs are available on request from the author. Preliminary results indicate that it is possible to improve on the accuracy of the approximations here described at a modest increase in complexity and these results will be reported shortly.

Bibliography

Abramowitz, M. and Stegun, I. A. (1965), Handbook 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. (1980a), "Rational Bounds for the t-Tail Area," Journal of the American Statistical Association, 75, 438-40.

Soms, A. P. (1980b), "A Note on an Extension of Rational Bounds for the t-Tail Area to Arbitrary Degrees of Freedom," Technical Summary Report No. 2106, Mathematics Research Center, University of Wisconsin-Madison and to appear in Communications in Statistics.