Advanced algorithms for the numerical integration of t

D. LOUIS WOOD

University of Arkansas, Little Rock, Arkansas

Wood and Wood (1984) introduced algorithms that resulted in a t integration program, TINT, which provided five-decimal-place two-tailed probabilities for t s ranging from $df = 1$ to $df = 113$. It produced computer overflow at $df = 114$ but then again functioned at its maximum of 115 degrees of freedom. The prominent aspects of that program were two subroutines that exactly resolved the gamma functions involved and that found the area under $f(t)$ using Simpson's rule. The article focused on these subroutines rather than providing a probability computational tool, and the program is thereby deficient in four ways for the scientist solely seeking t probabilities. First, five-place accuracy is sometimes insufficient, especially when the probability being sought is very small. Second, TINT does not present probabilities in a directly usable format, in that six-place figures are presented although the routine is accurate to only five places because of the limitation imposed by single-precision arithmetic. Third, a gap in the programming at $df = 114$ and a maximum $df = 115$ are restrictive. Finally, the program runs extremely slowly where $df < 4$, especially where $df = 1$.

This paper presents an advanced algorithm for the numerical integration of t that remedies all four limitations of TINT. First, the accuracy is to 8 decimal places for all t curves having $df > 3$ and 12 places for $df < 3$. The accuracies are explicated in Table 1, where probabilities displayed are two-tailed.

Second, probabilities are displayed in a rounded-off decimal form that is exactly accurate to the number of decimal points printed; for example, a probability of .1273814297 is presented on the computer screen with eight-place accuracy as .12738143. No rounding off or interpretation by the user is required. Third, probabilities are computed for all $f(t)$ from $df = 1$ through $df = 1234$. Finally, probabilities for $f(t)$ having $df < 4$ are provided quickly.

Improvements. Numerical integration of t is prefaced by the resolution of the following expression where $n = df$:

$$C = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}.$$

Highly accurate probability determination requires the exact solution of C , which poses a problem; that is, because $\Gamma[(n+1)/2]$ and $\Gamma(n/2)$ are factorial processes, they

both become huge numbers when n is at all large, and that leads to computer overflow.

The technique of Wood and Wood (1984) identifies factors that are common to both gammas and divides them out before multiplication begins, thereby doubling the usable number of degrees of freedom to the maximum $df = 115$ for that algorithm.

The present algorithm improves the Wood and Wood (1984) technique by dividing each remaining term in each gamma function by 176 before multiplication, thus reducing the magnitude of both products and greatly enlarging the smallest value of df that produces overflow.

The numerator symbolized as $G1$ in the program, thus reduced, is next divided by the reduced denominator G , and then that quotient is divided by $\sqrt{n\pi}$, as shown in the following equation:

$$C = \frac{G1}{G} \cdot \frac{1}{\sqrt{n\pi}}$$

The result is C when n is odd, because then the number of divisions in the numerator equals the number done in the denominator, so the process of dividing by 176 has not altered the value of C .

When n is even, one more term exists in $G1$ than in G ; therefore, the numerator has been divided by 176 one more time than has the denominator, resulting in a C value that is too small. This is corrected by multiplying the computed C value by 176 for a precise determination. The division operation described here expands the usable number of degrees of freedom more than tenfold to 1234.

Parenthetically, the method of dividing factors by 176 to avoid computer overflow is more accurate than is the factorial finding procedure described by Dunlap, Myers, and Silver (1984), which computes factorials using Napierian logarithms, as follows:

```
80 INPUT "Enter N ",N
90 DEFDBL F
100 FA=LOG(2)
110 FOR I=3 TO N
120 FA=FA+LOG(I)
130 NEXT
140 FL=EXP(FA)
150 PRINT "N factorial = " FL
```

Table 1
Accuracy of t Integration as a Function of t and
of Degrees of Freedom

df	t	Accuracy
$df > 3$	all	8 places
$df = 3$	$t < 151$	8 places
$df = 3$	$151 < t < 287$	7 places
$df = 2$	all	12 places
$df = 1$	all	12 places

The author's mailing address is: Department of Psychology, University of Arkansas at Little Rock, 2801 S. University, Little Rock, AR 72204.

This algorithm should return precisely correct answers, but sometimes it does not. For example, when run in double-precision arithmetic and where $N = 13$, the IBM PC, using BASIC/D in DOS 3.1, prints the answer of 6,227,017,745, whereas the correct answer is 6,227,020,800. Error has therefore been incurred in the sixth decimal place, which cannot be allowed in a high-precision algorithm such as the one presented here.

The second improvement contained by the program over the Wood and Wood (1984) algorithm is enhanced accuracy. Results are ($df > 3$) displayed as 8-decimal-place probabilities that are always fully correct. This improvement over the 5-place probabilities of Wood and Wood (1984) was achieved primarily by writing the algorithm specifically for the IBM PC and compatible computers that use a version of Microsoft BASIC offering the BASIC/D option. This variety of BASIC functions almost identically to Microsoft BASIC, except that variables defined in double precision are processed to 16 decimal places during the execution of mathematical functions such as SQR, LOG, and EXP. Because this routine extracts square roots hundreds and sometimes thousands of times per run, high-precision square roots are necessary for accurate computations of probabilities.

The final improvements in the program apply to $f(t)$ when $df < 4$. In these cases, the Wood and Wood (1984) program was stressed and ran very slowly. The present routine computes probabilities to 12-place accuracy where $df = 1$ or $df = 2$ and completes each calculation in less than 1 sec. The technique employed is not numerical integration, but instead is an exact technique introduced by Student (1908) and well explicated in a paper by Brophy (1987). Brophy's version is used in this paper. For $df = 1$ the exact area between the absolute value of any t entered by the user of the program and $-t$ is given by the following simple formula:

$$\text{area} = \frac{2}{\pi} \arctan(t).$$

The two-tailed probability is, of course, given by $P = 1 - \text{area}$. Where $df = 2$, Brophy cites another of Student's (1908) formulas for the exact areas between $-t$ and t , which follows:

$$\text{area} = \frac{t}{\sqrt{2+t^2}}$$

Simpson's rule is used to obtain probabilities for all $f(t)$ having $df > 2$, and running times are always less than 01:10 when $df > 4$.

The Wood and Wood (1984) program accepted a maximum t of 2000 when $df = 1$, which has $p = .0003183$, whereas the routine presented here accepts a $t = 999,999,999$ ($df = 1$), which has $p = .000000000001$. The 1984 program accepted a maximum t of 300 for $df = 2$, which has $p = .00001$; the present routine accepts a maximum $t = 999,999$ when $df = 2$, which has $p = .00000000001$.

Discussion. For all $f(t)$ having $df > 3$, probabilities are calculated for all t s ranging from $t = 0$ to a t having a two-tailed probability of .00000001. Larger t s elicit a screen message that the limits of accuracy have been exceeded. The limiting t value when $df = 3$ is the t having $P(t) = .0000001$; only when $df = 3$ and $t > 150$ is seven-place accuracy returned.

Accuracy is the prepotent focus of this program; therefore, all probabilities are calculated to be accurate to at least one more decimal place than is displayed on the screen as protection against round-off error. Eight-place probabilities, for example, are always computed using enough curve-area segments to assure a minimum 9-place accuracy. This holds, regardless of the number of degrees of freedom and the size of t ; for example, when $df = 1000$ and $t = 5.8$, 10 accurate digits are computed and 8 are displayed. The probabilities provided are always accurate beyond the number of digits displayed, provided that the BASIC/D language is used.

Despite the complete commitment to accuracy under which the program was developed, including being run in the slower BASIC/D, it operates surprisingly quickly. All runs where $df > 5$ are carried out in 11 sec to 26 sec on the unmodified IBM PC. Even large t s with few degrees of freedom typically require 1 min or less; for example, $P(t = 120)$, $df = 3$, requires 1 min 2 sec. Only when $df = 3$ and t is enormous do runs exceed 1 min.

The program is extremely simple to use. It begins by informing the user that it determines two-tailed t probabilities. It next requests that the number of degrees of freedom be entered, after which it resolves the gamma functions and computes C . When a large number of degrees of freedom is involved, this can take several seconds. It then requests that t be entered, which is accepted in positive or negative form, and it displays a message that computation has begun. When the Simpson's rule algorithm reaches its halfway point, the user is so informed. Finally, the number of segments into which the curve has been partitioned (W) is displayed with the time elapsed in the computations. The time includes the amount of time required to enter t , and so the user making time checks should enter t quickly while the gammas are being resolved. The program actually calculates the area between $t = 0$ and the user-entered value, which is doubled and displayed as the area between $-t$ and t . That area is subtracted from 1 and is displayed as the two-tailed probability.

Availability. A listing of the program can be obtained without charge from the author. Persons wanting a disk copy should send a formatted IBM disk (DOS 2.0 or later) in a returnable box.

REFERENCES

- BROPHY, A. L. (1987). Efficient estimation of probabilities in the t distribution. *Behavior Research Methods, Instruments, & Computers*, 19, 462-466.
- DUNLAP, W. P., MYERS, L., & SILVER, N. C. (1984). Exact multinomial

- probabilities for one-way contingency tables. *Behavior Research Methods, Instruments, & Computers*, **16**, 54-56.
- STUDENT. (1908). The probable error of a mean. *Biometrika*, **6**, 1-25.
- WOOD, D. L., & WOOD, D. (1984). TINT: A Microsoft BASIC t integration program. *Behavior Research Methods, Instruments, & Computers*, **16**, 479-480.
- (Revision accepted for publication October 14, 1987.)