Certification of Algorithm 708: Significant-Digit Computation of the Incomplete Beta

BARRY W. BROWN and LAWRENCE B. LEVY
The University of Texas M. D. Anderson Cancer Center

Algorithm 708 (BRATIO) was run on 2730 test cases. Comparison of these results with the results from an algorithm using a continued fraction of Tretter and Walster were performed using a high-precision version of the latter algorithm implemented in Maple. Accuracy of BRATIO ranged from 9.64 significant digits to full machine double-precision, 15.65 significant digits, with the lower value occurring when a was nearly equal to b, and a was large.

Categories and Subject Descriptors: G.1.2 [**Numerical Analysis**]: Approximation

General Terms: Algorithms

Additional Key Words and Phrases: Continued Fractions, F-distribution

1. INTRODUCTION

The incomplete beta function is

$$I_x(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x u^{a-1}(1-u)^{b-1}\, du.$$

Our interest in the calculation of this function arose in the course of programming the accelerated-failure model [Kalbfleish and Prentice 1980]. According to this model, the likelihood function for censored observations is the tail probability of the F-distribution, having, say, $2a$ and $2b$ degrees of freedom. We evaluate this probability by using Eq. 26.6.2 of Abramowitz and Stegun [1972] which converts the problem to that of calculating the incomplete beta function with parameters a and b. Important special cases of the accelerated failure model send either or both degrees of freedom of the F-distribution to infinity. Calculations are thus required for a wide range of a and b; we arbitrarily set this range at 10^{-3} to 10^{10}.

This work was supported in part by grant CA16672 from the National Cancer Institute, a cooperative-study agreement with IBM, and by the personal generosity of Larry and Pat McNeil.

Authors' address: Department of Biomathematics, Box 237, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030.

Prior to the publication of the work of DiDonato and Morris [1992], we tested several available algorithms for calculating the incomplete beta function but found them inadequate. In this certification, we examine the accuracy of the results obtained from BRATIO, using as a basis for comparison the continued-fraction expansion favored by Tretter and Walster [1979; 1980]. This algorithm was implemented in the Maple language [Waterloo Maple Software 1990] and was run using high-precision arithmetic. This continued fraction is not used in BRATIO and so is appropriate as an independent test of accuracy of that code.

2. TEST CASES

The parameters a and b were systematically varied over all combinations of integral powers of ten from -3 through 10. For each set of $\{a, b\}$, values of $z = \min(x, 1 - x)$ were generated that produced the following set of values of $I_x(a, b)$:

$$\{0.0001, 0.001, 0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99, 0.999, 0.9999\}.$$

In generating the cases, $I_{0.5}(a, b)$ was first evaluated. If this value was greater than the desired value, then x was determined iteratively; otherwise, $1 - x$ was found. Computation of $I_x(a, b)$ for the generation of test cases used BRATIO; finding the value of x or $1 - x$ that yields a specified I_x used the zero-finding Algorithm R of Bus and Dekker [1975]. Great accuracy in these generated values was not necessary, since the sought value of I_x was not used in the evaluation of the accuracy of BRATIO. A relative tolerance of 10^{-10} and an absolute tolerance of 10^{-300} were provided to Algorithm R. The combinations of 14 values of a, 14 of b, and 15 values for I_x provided 2940 test cases. In 210 of these cases, a value of 0.25×10^{-300} was returned for z; this is the lowest nonzero value allowed by the generation method used. These 210 cases were eliminated from the test suite leaving 2730 test cases.

3. HIGH-PRECISION CALCULATIONS

The associated continued fraction for Mueller's expansion of the incomplete beta function after analytic subtraction, Tretter and Walster [1980, p. 323], was coded in the Maple language. The reversal formula,

$$I_x(a, b) = 1 - I_{1-x}(b, a),$$

had to be used in the code whenever $x > a/(a + b)$; otherwise the continued fraction could converge to a negative result. Early testing showed an inordinate number of iterations for fractional values of a and b. Consequently, cases in which the parameters had values less than 1 were converted to equivalent problems with the small parameters incremented by 1. This transformation used the same formula as BUP of BRATIO. All calculations utilized the Maple logarithm-of-the-Γ function. In the test cases with $x > 0.5$, the Maple routine was passed $(1 - x, b, a)$ instead of (x, a, b).

Initially, we calculated the results of the test problems using 100 decimal digits of floating-point accuracy. The evaluation of successive convergents of the continued fraction persisted until the relative difference between successive convergents was less than 10^{-50}. All answers were recorded to 30 decimal digits. Running the test cases required approximately 3.75 hours on an IBM RS6000 model 550. Cases in which a and b were large required a great number of convergent evaluations; the largest amount was 25,698.

To assess the effect of a change in accuracy of the calculation, the test problems were rerun using only 50 decimal digits and a convergence criterion of 10^{-40}. The two runs produced identical results to the 30 digits recorded.

4. TESTS OF BRATIO

The as-distributed BRATIO was converted to double-precision using **apt** from Toolpack 1.2. Routine SPMPAR was manually modified to use the double-precision constants from IPMPAR. The test cases were run with the modified BRATIO and the results analyzed using Splus [Statistical Sciences 1991].

The number of significant digits of accuracy obtained by BRATIO was calculated for each test case as

$$-\log_{10}(2\,\mathit{RelativeError}),$$

which is the usual definition: n significant digits of accuracy in a value allows an error of 5 in the $(n + 1)$st decimal place. To avoid taking the logarithm of 0, the relative error was bounded below by the double-precision machine epsilon (1.11×10^{-16} = 15.65 significant digits of accuracy). Accuracy ranged from 9.64 significant digits to 15.65 with a median of 14.65 and a lower quartile of 13.81.

As Figure 1 shows, the overall accuracy of BRATIO increases slightly as a/b moves away from 1. Linear regression indicates that (1) an average of 13.71 significant digits are obtained in cases in which $a = b$ and (2) the number increases 0.14 significant digits for each unit change in $\log_{10}(a/b)$.

5. CONCLUSION

Obtaining the accuracy found in these tests depends on supplying BRATIO with an accurate value of the minimum of x and $1 - x$. On machines with IEEE double precision if $x = 1$, any value of $1 - x$ from about 10^{-17} to 0 will meet the condition that $x + (1 - x)$ equals 1 to machine accuracy, but the value of I_x can vary considerably. For example, if $a = 10^{10}$, $b = 10^{-3}$, and $1 - x$ is 0, then $I_x = 1$; if $1 - x$ is changed to 0.6912×10^{-107}, I_x drops to 0.2. (This example was supplied by an anonymous referee.) In statistical models involving the F or log-F distribution, x and $1 - x$ are available without subtraction.

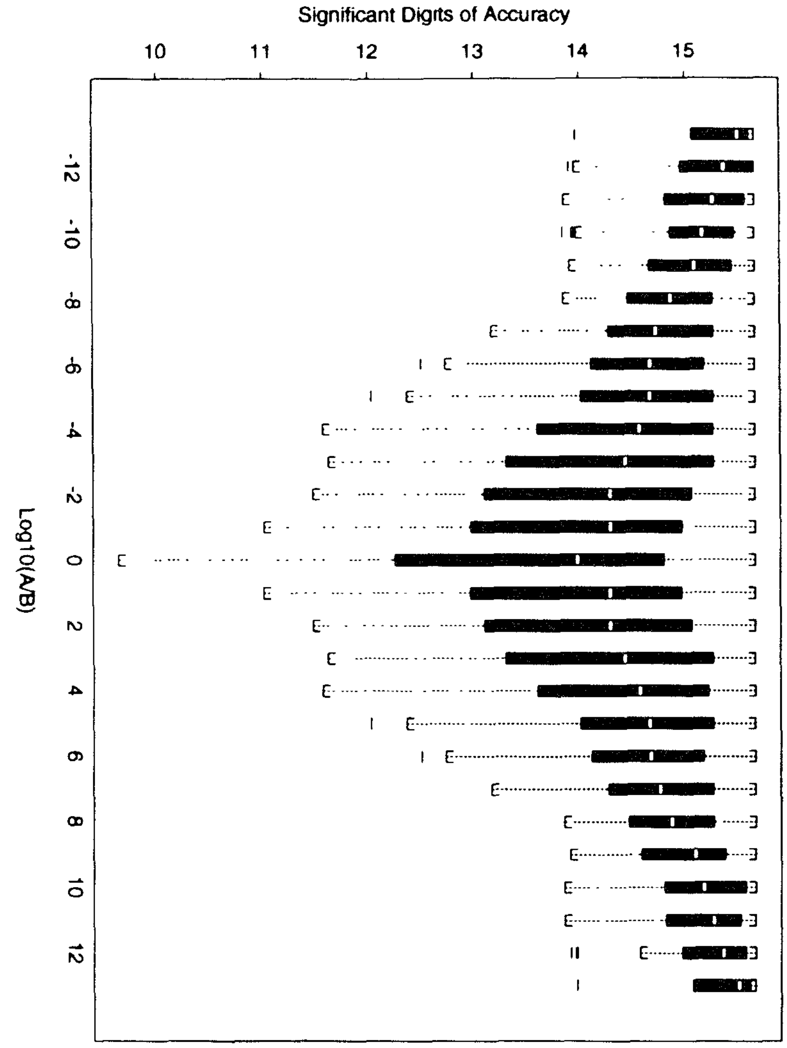

Fig. 1. Boxplot of the significant digits of accuracy calculated by BRATIO versus the ratio of the parameters of the test problems. The center of each box indicates the median value; the extremes of the box show the quartiles. The whiskers extend a length of 1.5 times the distance between the lower and upper quartile. For a Gaussian distribution, 99.3% of the data would lie within the whiskers. The accuracy of the computation exceeds 9.6 significant digits for all test cases and is greatest in cases for which a and b differ in magnitude.

The accuracy of this routine exceeds that of any other that we have examined, including a method that we devised. Because BRATIO provides a minimum of almost ten significant places of accuracy over the wide range covered by our test cases, we are adopting it.

ACKNOWLEDGMENTS

The authors acknowledge the Free Software Foundation whose developments *gawk* and *Emacs* were instrumental in solving the routine problems in this work. Without these packages, the problems would not have been routine. The authors also gratefully acknowledge the editorial assistance of Kimberly Herrick.

REFERENCES

ABRAMOWITZ, M., AND STEGUN, I. 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* U.S. Government Printing Office, Washington, D.C.

BUS, J. C. P., AND DEKKER, T. J. 1975. Two efficient algorithms with a guaranteed convergence for finding a zero of a function. *ACM Trans. Math. Softw. 1*, 4, 330–345.

DIDONATO, A. R., AND MORRIS, A. H. 1992. Algorithm 708: Significant-digit computation of the incomplete beta function ratios. *ACM Trans. Math. Softw. 18*, 3, 360–373.

KALBFLEISH, J. D., AND PRENTICE, R. L. 1980. *The Statistical Analysis of Failure Time Data.* John Wiley and Sons Inc., New York.

STATISTICAL SCIENCES, INC. 1991. *S-Plus Version 3.0.* Statistical Sciences, Inc., Seattle, Wash. (Computer Program).

TRETTER, M. J., AND WALSTER, G. W. 1980. Analytic subtraction applied to the incomplete gamma and beta functions. *SIAM J. Sci. Stat. Comput. 1*, 321–326.

TRETTER, M. J., AND WALSTER, G. W. 1979. Continued fractions for the incomplete beta function: Additions and corrections. *Ann. Stat. 7*, 2, 462–465.

WATERLOO MAPLE SOFTWARE. 1990. *Maple V.* Waterloo Maple Software, Waterloo, Ont. (Computer Program).

Received May 1993; revised September 1993; accepted October 1993