APIC'95, El Paso, Extended Abstracts, A Supplement to the international journal of Reliable Computing

Self-validating Computation for Selected Probability Distribution Functions

Morgan C. Wang[1], Kurt Lin[2], and William J. Kennedy[3]

Self-validating numerical methods based upon interval analysis for the computation of selected probability distribution functions are reported. Self-validating computation can be achieved in many different ways. We employed interval analysis to achieve the goal of self-validation. And the Hausdorf metric between two intervals $A = [a, b]$ and $B = [c, d]$ defined as $\max(|a-c|, |b-d|)$ can be used to serve as a guaranteed error bound to validity the computation results.

In this approach, the input data can be taken as either thin intervals (numbers) or thick intervals (intervals which have width greater than zero). Rounded interval arithmetic [1] need to be used when implement these numerical methods on a digital computer which has finite precision. Thus, when the input data are thin intervals, this approach produces a rounded interval which is guaranteed to contain the theoretically correct value of the desired probability or percentile. Since the Hausdrof metric between midpoint of the computed interval and the true value is less than the half-width of the computed interval, the half-width of the computed interval can serve as a guaranteed absolute error bound giving validity to this midpoint approximation. When the input data are thick intervals, this approach produces a rounded interval which is guaranteed to contain the theoretically correct interval. Since the Hausdorf metric between the computed interval and the theoretically correct interval is less than the width of the computed interval, the width of the computed interval can be used as a guaranteed error bound giving validity to this interval approximation.

The major advantage of this approach is the additional information, provided by the guaranteed error bounds, about the reliability of the computed results. The conventional method, which uses scalar computation and produces scalar approximations, does not provide the error bounds for the computed values, i.e., it does not provide the reliability information directly. Although reliability information for conventional method is theoretically available, it is only obtained as the result of extensive error analysis. However, even with extensive error analysis the reliability information is still not numerically available at run time and error bounds from classical error analysis tend to be very conservative. Therefore, it is difficult to locate the region of the variable and parameter space in which these scalar algorithms produce accuracy within a specified level.

The need for guaranteed accuracy within stated limits arises frequently when computing probabilities and percentiles. For example, when comparing competing scalar algorithms to see which yields greater accuracy, or when evaluating a new algorithm, a reliable source of essential true values is needed. Existing tables usually do not provide sufficiently accurate entries, or cover a sufficiently large region of the variable and parameter spaces, to be satisfactory for this application. Another example occurs whenever a probability function enters as a factor in an algebraic expression which must be evaluated, possibly for purposes of tabling. Accuracy in the end result will depend in part on the level of accuracy of the computed probability. For example, the distribution of the sample correlation coefficient, r, from a bivariate normal population with correlation ρ, has central F cumulative distribution function (CDF) as a factor. Accurately approximating the CDF values of the sample correlation coefficient will, of course, depend in part on the level of accuracy achieved in the approximation of the central F CDF [6].

The probability distribution functions which included univariate normal, central and non-central chi- square, central and non-central F, bivariate normal [9], multivariate normal [8], and distribution for the sample correlation coefficients are considered in this report. These methods suggested in this report have been extensively tested and implemented on both IBM PC and SUN workstation. Excellent results were obtained over very large regions

The authors are with [1]Department of Statistics, University of Central Florida, P. O. Box 162370, Orlando, Florida 32816-2370, e-mail: cwang@pegasus.cc.ucf.edu, with the [2]Institute for Simulation and Training, University of Central Florida, 3280 Progress Drive, Orlando, FL 32826, e-mail: klin@pegasus.cc.ucf.edu, and with the [3]Department of Statistics, Iowa State University, Ames, Iowa 50011.

of the variable and parameter space in every distribution functions considered. When a failure occurred, an excessively large interval resulted. This served to notify of failure. The methods are not completely fail safe, because the results are not valid if floating-point underflow or overflow occurs. However, underflows and overflows can be detected, so there is a large measure of dependability provided by this methodology. The authors believe that self-validating computations should be done more frequently in statistical computing.

References

[1] G. Alefeld and J. Herzberger, **Introduction to Interval Computations**, Academic Press, N.Y., 1983.

[2] G. F. Corliss, "Application of Differentiation Arithmetic", In: **Reliability in Computing: The Role of Interval Methods in Scientific Computing**, Academic Press, N.Y., 1988.

[3] G. F. Corliss, "Industrial Applications of Interval Techniques", In: C. Ullrich (ed.), **Computer Arithmetic and Self-Validating Numerical Method**, Acadmeic Press, N.Y., 1990.

[4] G. F. Corliss and L. B. Rall, "Adaptive, Self-Validating Numerical Quadrature," *SIAM Journal on Scientific and Statistical Computing*, 1987, Vol. 8, pp. 831–847.

[5] L. B. Rall, **Automatic Differentiation: Techniques and Application**, Lecture Notes in Computer Science No. 120, Springer Verlag, New York, 1981.

[6] M. C. Wang and W. J. Kennedy, "A Self-Validating Numerical Method for Computation of Central and Non-Central F Probabilities and Percentiles," *Statistical and Computation*, 1995 (to appear).

[7] M. C. Wang and W. J. Kennedy, "Self-Validating Computations of Probabilities for Selected Central and Non-central Univariate Probability Functions," *Journal of the American Statistical Association*, 1994, Vol. 89, pp. 878–887.

[8] M. C. Wang and W. J. Kennedy, "A Numerical Method for Accurately Approximating Multivariate Normal Probabilities," *Computational Statistics and Data Analysis*, 1992, vol. 13, pp. 197–210.

[9] M. C. Wang and W. J. Kennedy, "Comparison of Algorithms for Bivariate Normal Probability Over a Rectangle Based on Self-Validating Results From Interval Analysis," *Journal of Statistical Computation and Simulation*, 1990, Vol. 37, pp. 13–25.