Order Number 9202406

Numerical methods for self-validating computation of probabilities and percentiles in selected distributions using interval analysis

Wang, Chung-Ching Morgan, Ph.D.

Iowa State University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Numerical methods for self-validating computations of probabilities and percentiles in selected distributions using interval analysis

by

Chung-Ching Morgan Wang

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Members of the Committee:

Signature was redacted for privacy.

Iowa State University
Ames, Iowa
1991

TABLE OF CONTENTS

GENERAL INTRODUCTION

Most scientific computations are carried out on computers which employ fixed-precision floating-point number systems. Under these types of number systems, real numbers are approximated by a subset of real numbers called machine representable numbers (or floating-point numbers). Because of this kind of representation at least two types of errors are generated. The first type is due to initially approximating a non-machine representable number and the second type is caused by both the intermediated and final results being approximated by a machine number.

Are these two sorts of errors a problem in scientific computing? Before we answer this question, let us study the following examples which were written in FORTRAN and executed on an IBM 386 personal computer equipped with an INTEL 80387 numeric coprocessor.

Example 1:
```
X=4.D0/7.D0
Y=7.D0*(X-0.5D0)-0.5D0
IF (Y .EQ. 0.D0) Z=0.D0
IF (Y .NE. 0.D0) Z=(DEXP(Y)-1.D0)/Y
```

In this example, it is very easy to obtain the correct value of Z=0 by hand. However, on most powerful computers which employ floating-point number systems the value of Z isn't zero. The reason is that Y is a very small number instead of zero, i.e. , the computer executes the second IF statement instead of the first one. This example tells us that comparison of real numbers is not reliable even after as few as three floating-point arithmetic operations. Therefore, we should avoid using real number comparisons in scientific computing (Kahan 1980).

```
Example 2:   Y=2.D0
             X=Y
             DO 1 I=1,54
           1   X=DSQRT(X)
             DO 2 I=1,54
           2   X=X*X
```

In the above example, this simple FORTRAN algorithm only executes 108 computation operations which are all free of overflow, underflow, and cancellation. The only place rounding error may occur is in the square root or multiplication operations. Even though this program is simple the output value of X is one instead of two. In fact, for all values of x between 0.5 and 2.0 this program will produce the same answer (Kahan 1980). If we use rounded interval arithmetic the resulting interval is [1.99 , 2.01] instead of [1 , 1]. Although, this resulting interval is too wide to be useful, it is a warning of something going wrong.

```
Example 3:   X=1.D0
             Y=1.D25
             Z=(X+Y)-Y
```

This algorithm, which includes only two floating-point number arithmetic operations, produced the value Z=0, which is incorrect. The correct value of Z, i.e., Z=1, was not generated since double precision floating-numbers have at most 17 significant decimal digits. Note, however, that 25 decimal digits are needed to avoid rounding error in the "X+Y" operation on a digital computer (Bohlender 1990).

The above three examples indicate that these two kinds of errors indeed

pose a problem in scientific computing; therefore, nobody is surprised by the following three facts.

Fact 1: An algorithm may work very well in one part of the parameter space but fail in the rest of it, for example, Drezner's algorithm (Drezner 1990) for computing the bivariate normal probabilities. Some of the computed results of this algorithm are presented in Table 1. The "true" probabilities associated with these computed values are presented in Table 2.

Fact 2: An iterative algorithm with a given error tolerance may underestimate the absolute error associated with the computed result. Rust and Voit's algorithm (Rust and Voit 1990), which computes the noncentral chi-square probabilities, is such an algorithm. The actual absolute error is as much as one hundred times larger than the specified local error tolerance.

Fact 3: An iterative algorithm may fail to terminate for a given error tolerance in one part of its parameter space. An algorithm which demonstrates such a failure is Schervish's algorithm (Schervish 1984) for computing multivariate normal probabilities. In cases such as these, an infinite loop is usually the result.

These facts point out that the accuracy of the values produced by an algorithm still pose a problem in today's software. Thus, an algorithm which not only produces an answer but also produces a guaranteed error bound would be of interest, especially for the following situations : (1) an essentially true answer is required for an accuracy comparison study among several competing algorithms or an accuracy study of a newly developed algorithm, and

(2) the computed result needs to satisfy the given accuracy requirements because it is to be used in subsequent computations.

The computational method which produces an error bound associated with the computed value is sometimes called an automatic error analysis method or self-validating computation method. Self-validating results can be achieved in many different ways. We will use interval analysis to obtain self-validation.

Although, correctly applied interval computation will always yield an interval which contains the unknown true answers, the interval may be too wide to be useful. Therefore, a "good" algorithm which employs interval arithmetic is still needed. Several important tools which make interval arithmetic more attractive along with interval arithmetic will be discussed in the first section. Then these developed methods will be applied to selected probabilities and percentiles computation.

Table 1 Interval inclusion of selected bivariate normal probabilities

I.D.	HH	KK	RHO	Inclusion of Probability Lowerbound	Upperbound
1	0.1190	-1.3580	-0.9000	0.0187881584996960	0.0187881584996962
2	2.2770	2.4000	-0.9000	0.9804093512655087	0.9804093512655089
3	-0.726	0.5530	-0.7000	0.0762817182590762	0.0762817182590763
4	1.2010	2.8710	-0.5000	0.8830800190626548	0.8830800190626553
5	-1.527	-0.5890	-0.3000	0.0069816661176445	0.0069816661176446
6	3.2360	4.1010	-0.1000	0.9993733469951126	0.9993733469951128
7	-1.887	-0.2400	0.1000	0.0146400388318886	0.0146400388318887
8	2.9240	1.2120	0.3000	0.8862649576796010	0.8862649576796012
9	2.3540	-1.4710	0.5000	0.0706394390095590	0.0706394390095592
10	-0.884	4.5910	0.7000	0.1883481064180262	0.1883481064180264
11	4.9000	-1.3820	0.9000	0.0834858505529643	0.0834858505529644
12	-1.841	-0.1480	0.9000	0.0328088113489369	0.0328088113489371

Table 2 Drezner's bivariate normal probabilities

I.D.	HH	KK	RHO	Probabilities
1	0.1190	-1.3580	-0.9000	0.0098486821009678
2	2.2770	2.4000	-0.9000	219.69074470413275
3	-0.726	0.5530	-0.7000	0.0655221953599864
4	1.2010	2.8710	-0.5000	0.8830800190630133
5	-1.527	-0.5890	-0.3000	0.0069816661187741
6	3.2360	4.1010	-0.1000	0.9993733469951126
7	-1.887	-0.2400	0.1000	0.0146400388318886
8	2.9240	1.2120	0.3000	0.8862649576797093
9	2.3540	-1.4710	0.5000	0.0706394390085940
10	-0.884	4.5910	0.7000	-149.63841029481098
11	4.9000	-1.3820	0.9000	-1236.427400978697
12	-1.841	-0.1480	0.9000	0.7881471139366692

1. TOOLS EMPLOYED IN SELF-VALIDATING COMPUTATION

Although we obtain the goal of self-validation by applying interval arithmetic to our computational process, there are three other numerical tools - automatic differentiation, continued fractions , and Taylor series expansions - which can be used to make the computational process more efficient and attractive. Therefore, a review of these important tools is in order.

1.1 Interval arithmetic

Interval arithmetic was first introduced by Ramon E. Moore in 1966. The basic idea of interval arithmetic is that each number x in a calculation will be represented by an interval $[x_1,x_2]$, where x_1 and x_2 are chosen in such a way that we can guarantee that $x_1 < x < x_2$. Throughout the calculation, we deal with such intervals instead of numbers. In light of these facts, an interval can be thought of as a numerical tool in applied mathematics - a tool which can be especially useful in analyzing computational error, constructing upper and lower bounds of integral functions, and providing a natural stopping criteria for iterative numerical methods.

In the next subsection, we will review the basic elements of interval arithmetic. Then we will discuss the algebraic properties of interval arithmetic and interval inclusion of functions. Finally, we will explain how to implement interval arithmetic on a digital computer.

1.1.1 Interval arithmetic operations The field of real numbers is denoted by $\mathbb{R}$, and the elements of $\mathbb{R}$ are denoted by lower case letters a,b,c,...,x,y,z. The set of all real closed and bounded intervals is denoted

by $I\mathbb{R}$, and their elements are expressed by upper case letters. It should be realized that the real number a in $\mathbb{R}$ is considered as a special element [a , a] in $I\mathbb{R}$. We call this special element in $I\mathbb{R}$ a "point interval" or a "degenerated interval".

Let $A = [a_1 , a_2]$ and $B = [b_1 , b_2]$ be elements in $I\mathbb{R}$. Then interval arithmetic operations are defined as

$$A \circ B = \{ a \circ b \mid a \in A \text{ and } b \in B \}$$

where o is one of the arithmetic operators +, -, *, /; and / is defined only when $0 \notin B$. This definition of the interval arithmetic operations is equivalent to the following rules:

(1). Addition : $A+B=[a_1+b_1 , a_2+b_2]$,

(2). Subtraction: $A-B=[a_1-b_2 , a_2-b_1]$,

(3). Multiplication: $A*B=[\min(a_1b_1,a_1b_2,a_2b_1,a_2b_2), \max(a_1b_1,a_1b_2,a_2b_1,a_2b_2)]$,

(4). Reciprocal: $1/A=[1/a_2 , 1/a_1]$ if $0 \notin A$,

(5). Division: $A/B=A*(1/B)$ if 1/B is defined.

Another important fact which was proved by Ratchek and Rokne (1984) is presented here:

Theorem 1: If A and B are in $I\mathbb{R}$, then AoB is in $I\mathbb{R}$, where o represents any interval arithmetic operator for which AoB is defined.

Now, several definitions which will be used in later discussion are given:

Definition 1: Equality

Two intervals A and B are equal if $a_1=b_1$ and $a_2=b_2$.

Definition 2: Inclusion

$A \subseteq B$ if and only if $b_1 \leq a_1$ and $a_2 \leq b_2$

Definition 3: Order Relation

$A < B$ if and only if $a_2 < b_1$,

$A \leq B$ if and only if $a_2 \leq b_1$,

$A > B$ if and only if $a_1 > b_2$,

$A \geq B$ if and only if $a_1 \geq b_2$.

Definition 4: Width of an Interval

The width of an interval $A=[a_1,a_2]$ is a_2-a_1 and is denoted by $w(A)$.

Definition 5: Absolute Value of an Interval

The absolute value of an interval $A=[a_1,a_2]$ is $\max(a_1,a_2)$ and is denoted by $|A|$.

Definition 6: Midpoint of an Interval

The midpoint of an interval $A=[a_1,a_2]$ is $0.5*(a_1+a_2)$ and is denoted by $m(A)$.

Definition 7: Symmetric Interval

The interval $A=[a_1,a_2]$ is symmetric if $a_1=-a_2$.

1.1.2 Algebraic properties of interval arithmetic The algebraic properties of interval arithmetic and real number arithmetic are different since the real numbers form a complete ordered field, whereas arithmetic operations which use interval numbers do not follow all the rules of a field. For example under $\mathbf{R}$, a*(b+c) = a*b+a*c; however under $I\mathbf{R}$, A*(B+C) isn't necessarily equal to A*B+A*C. The same discrepancy holds for the inverse property of multiplication. Under $\mathbf{R}$, a*(1/a)=1; but under $I\mathbf{R}$, A*(1/A)

isn't necessarily equal to [1,1].

Thus, it is necessary for us to present the following algebraic properties of interval arithmetic. Since the proofs of these properties can be found in Moore (1979), Alefeld and Herzberger (1983), and Ratschek and Rokne (1984), we will present these properties without giving proofs. Suppose A, B, C, and D are in $I\mathbb{R}$. Then the following properties all hold true in $I\mathbb{R}$:

(1). Commutativity in Multiplication and Addition:

$A+B=B+A$ and $A*B=B*A$.

(2). Associativity in Addition and Multiplication:

$A+(B+C)=(A+B)+C$ and $A*(B*C)=(A*B)*C$.

(3). Subdistributivity:

$A*(B+C)\subseteq A*B+A*C$.

(4). Inclusion Monotonicity in Arithmetic Operations:

If $A\subseteq C$ and $B\subseteq D$ and o is any arithmetic operator then $AoB\subseteq CoD$.

(5). Cancellation Law in Addition:

$A+B=A+C$ implies $B=C$.

(6). Existence of the Identity in Addition and Multiplication:

$\exists\ 0=[0,0]$ such that $A+0=0+A=A$,

$\exists\ 1=[1,1]$ such that $A*1=1*A=A$.

(7). Distributivity for several special cases:

$A*(B+C)=A*B+A*C$

if (i). $A = [a,a]$,

(ii). $B*C > 0$,

(iii). B and C are symmetric intervals.

Some properties involving the width and midpoint operations are also of interest. These properties are as follows:

(8). $A \subseteq B$ implies that $w(A) \leq w(B)$.

(9). $w(A+B)=w(A)+w(B)$ and $w(A-B)=w(A)-w(B)$.

(10). $w(a*B)=|a|*w(B)$.

(11). $m(A+B)=m(A)+m(B)$ and $m(A-B)=m(A)-m(B)$.

Additional properties of interval arithmetic and a complete list of research work in this area can be found in Moore (1979), Nickel (1975, 1980, 1985, and 1990), and Alefeld and Herberger (1983).

1.1.3 Interval Inclusion of Functions In this subsection, we consider the problem of finding interval inclusion of functions. Let f be a real-valued function of n real variables $x_1, x_2, \ldots, x_n$, defined on real intervals $X_1, X_2, \ldots, X_n$, respectively. The "united extension" of the function f over $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n) = \bigcup_{(x_1, x_2, \ldots, x_n) \in (X_1, X_2, \ldots, X_n)} \{f(x_1, x_2, \ldots, x_n)\}.$$

Although the united extension of a function is unique, in general, it need not be an interval and it may not be obtained in finite many steps. Therefore, instead of considering the united extension of a general class of functions we only consider interval extensions of rational functions.

By a "rational function," we mean a function which can be expressed in terms of the trigonometric functions and their inverses, the exponential function e^x and its inverse $\ln(x)$, the identity function x, and the constant functions, using finite numbers of binary operations of addition, subtraction, multiplication, and division, and the unary operations of

composition and exponentiation.

An "interval extension" of a function is an interval-valued function F of n interval variables $X_1,\ldots,X_n$ with the property

$$F(x_1,x_2,\ldots,x_n) = f(x_1,x_2,\ldots,x_n),$$

for all real arguments. Clearly, interval inclusion of a function is not unique. Therefore, a form of interval inclusion which has short width is desired. The standard centred form, first suggested by Moore (1966), is a good form of an interval extension of a rational function. Explicit formulas for these forms were found by Ratschek (1980) for rational functions. Further discussion can be found in Ratchek and Rokne (1984).

We will be interested in finding the inclusion monotonic interval extensions of certain functions. The inclusion monotonic interval extension will be simply called interval inclusion. The definitions of inclusion monotonic and natural interval extensions are as follows:

Definition 8: Inclusion Monotonic

An interval-valued function F of $X_1,X_2,\ldots,X_n$ is inclusion monotonic if for any subsets $Y_1,Y_2,\ldots,Y_n$ of $X_1,X_2,\ldots,X_n$ respectively

$$F(Y_1,\ldots,Y_n) \subseteq F(X_1,\ldots,X_n).$$

Definition 9: Natural Interval Extension

The natural interval extension of a real-valued function f can be obtained by substituting interval operations for the corresponding scalar operations. The resulting interval-valued function is called the natural interval extension of f.

Clearly, the natural interval extension is unique. Moreover, we can

show that the natural interval extensions of rational functions are inclusion monotonic. Therefore, a general procedure for obtaining an interval inclusion of a given function is as follows:

1. Obtain an interval inclusion of non-rational parts of this function.
2. Substitute the resulting intervals to the non-rational parts of the function. The resulting function is a rational function.
3. Obtaining the standard centred form of the resulting function.
4. The natural interval extension of this resulting form is a good interval inclusion of the original function.

Finally, we want to point out that the general rule for obtaining interval inclusion of a non-rational function is still not available. The step 1 of the previous procedure is different in different cases. Thus, we leave the details of this discussion until part I through part IV of this dissertation, but we assume for now that interval inclusion of non-rational function is available.

1.1.4 Rounded Interval Arithmetic A digital computer can only approximate the set of all real numbers using a finite subset. This subset constitutes the elements in the floating-point number system, which will be denoted by $\mathbb{RM}$. Since the properties of floating-point numbers are different from those of real number system, implementation of real interval arithmetic on a digital computer should be done very carefully if the properties of real interval arithmetic are to remain valid. Although there are many ways in which this task can be successfully accomplished, "rounded interval arithmetic" is used. Rounded interval arithmetic can be implemented on almost any kind of digital computer by using different programming languages

such as ALGOL, FORTRAN, ASSEMBLER, and PASCAL-SC. We used ASSEMBLER language to write these FORTRAN callable subroutines to perform the rounded interval arithmetic operations on an IBM personal computer equipped with an INTEL 80387 numerical coprocessor.

We know that the floating-point number is bounded, i.e. , there exists y_l and y_u in $\mathbf{RM}$ such that for all y in $\mathbf{RM}$, we have that $y_l \le y \le y_u$. Only those real numbers lying in the interval $[y_l, y_u]$ can be approximated effectively by a machine number. This approximation process can be achieved by a mapping from $\mathbf{R}$ to $\mathbf{RM}$ defined as

$$\forall\, x \in \mathbf{R},\ \exists\, xm \in \mathbf{RM} \text{ s.t. } fl(x)=xm.$$

The mapping is called "rounding" whenever monotonicity is satisfied by this mapping, i.e., $\forall\, x \le y$ in $\mathbf{R}$, we have $fl(x) \le fl(y)$. The mapping is called "downward directed rounding" whenever it is a rounding and $fl(x)=\nabla x$, such that for all $xm \le x$ in $\mathbf{RM}$ we have $xm \le \nabla x$ and for all $xn \ge x$ in $\mathbf{RM}$ we have $xn \ge \nabla x$. This mapping is called "upward directed rounding" whenever it is a rounding and $fl(x)=\Delta x$, such that for all $xm \ge x$ in $\mathbf{RM}$ we have $xm \ge \Delta x$ and for all $xn \le x$ in $\mathbf{RM}$ we have $xn \le \Delta x$.

For any real interval $A=[a_1, a_2]$ with $y_l \le a_1 \le a_2 \le y_u$, the interval $AM=[\nabla a_1, \Delta a_2]$ is the corresponding rounded interval in $\mathbf{RM}$. The set of all rounded intervals is denoted by $I\mathbf{RM}$. The mapping from a real interval to another real interval is called directed rounding if it performs downward directed rounding on the lower endpoint and upward directed rounding on the upper endpoint, i.e. , for any real interval $A=[a_1, a_2]$, the corresponding rounded interval is $fl(A)=\ddagger(A)=[\nabla a_1, \Delta a_2]$. Now, we give a formal definition of rounded interval arithmetic.

Definition 10: Rounded Interval Arithmetic

Let AM, BM $\in$ I $\mathbb{R}$M and let o $\in$ (+,-,*,/) be an arithmetic operator. Then the result of the arithmetic operation o on A and B using directed rounding is given by C=$\ddagger$(AoB).

Clearly, if neither overflow nor underflow occurs, the resulting interval C is in I $\mathbb{R}$M. Also, for any interval A$\subseteq$B in I $\mathbb{R}$, let AM=$\ddagger$A and BM=$\ddagger$B then AM$\subseteq$BM. Using this fact, we can prove the following two important theorems in rounded interval arithmetic. Assume that overflow or underflow doesn't occurred for simplicity of discussion.

Theorem 2:

Let AM and BM $\in$ I $\mathbb{R}$M and let o$\in$(+,-,*,/) be an arithmetic operator. Then $\forall$ am$\in$AM and bm$\in$BM there exists a CM in I $\mathbb{R}$M such that

$$(1).\ AM\ o\ BM \subseteq CM = \ddagger(AM\ o\ BM),$$

$$(2).\ am\ o\ bm \in CM = \ddagger(AM\ o\ BM).$$

Theorem 3:

Let AM$\subseteq$BM and CM$\subseteq$DM $\in$ I $\mathbb{R}$M and o be an arithmetic operator. Then there exist EM and FM in I $\mathbb{R}$M such that

$$EM=\ddagger(AM\ o\ CM)\subseteq FM=\ddagger(BM\ o\ DM).$$

Using the above two theorems along with the definition of rounded interval arithmetic, we can prove all of the facts about interval arithmetic which were presented in the previous two subsections assuming no overflow or underflow occur in the rounded arithmetic operations.

1.2 Automatic differentiation

Automatic differentiation is also called differentiation arithmetic, formula differentiation, or algorithm differentiation. It is a technique for

computing derivatives of a function using recurrence relations derived from the differentiation rules of calculus without using algebraic expressions of these derivatives. This technique is much easier and more efficient and accurate than using traditional numerical software along with the algebraic expression. It makes scalar or interval algorithms involving derivatives easier to use, since the difficulty of obtaining derivatives no longer exists.

In our application, we need to apply automatic differentiation to compute the Taylor series coefficients of an integrand function f(x) at a given point c, where f is an analytic function of x for which the Taylor series coefficients exist. Before we discuss how to apply the recurrence relationships of elementary functions to compute the Taylor series coefficients of the integrand function in this study, we list these formulas using the following notation,

$$\left.\frac{f^{(j)}(x)}{j!}\right|_{x=c} = (f)_j$$

to denote the jth order Taylor series coefficient of f evaluated at the point c.

Let f and g be any two analytic functions of x. We can derive recurrence relations from the chain rule and basic differentiation rules. These rules are as follows:

$$(1).\ (e^f)_k = \left(\frac{1}{k}\right) \sum_{j=1}^{k} j(e^f)_{k-j}\,(f)_j \quad \text{if } k>0,$$

$$(2).\ (\ln f)_k = \left(\frac{1}{f}\right)\left[(f)_k - \frac{1}{k}\sum_{j=1}^{k-1} j(\ln f)_j\,(f)_{k-j}\right] \quad \text{if } k>0,$$

(3). $(\sin f)_k = (\frac{1}{k}) \sum_{j=1}^{k} j\,(\cos f)_{k-j}\,(f)_j$ if $k>0$,

(3). $(\cos f)_k = -(\frac{1}{k}) \sum_{j=1}^{k} j\,(\sin f)_{k-j}(f)_j$ if $k>0$,

(5). $(\sinh f)_k = (\frac{1}{k}) \sum_{j=1}^{k} j\,(\cosh f)_{k-j}\,(f)_j$ if $k>0$,

(6). $(\cosh f)_k = -(\frac{1}{k}) \sum_{j=1}^{k} j(\sinh f)_{k-j}\,(f)_j$ if $k>0$,

(7). $(f^a)_k = (\frac{1}{f}) \sum_{j=1}^{k} \left[\frac{(a+1)j}{n} - 1\right](f^a)_{k-j}\,(f)_j$ if $k>0$,

(8). $(\tan^{-1} f)_k = \frac{1}{(1+f^2)}\left((f)_k - \frac{1}{k}\sum_{j=1}^{k-1} j(\tan^{-1} f)_j\,(1+f^2)_{k-j}\right)$ if $k>0$,

(9). $(\sin^{-1} f) = \frac{1}{\sqrt{1-f^2}}\left[(f)_k - \frac{1}{k}\sum_{j=1}^{k} j(\sin^{-1} f)_j(\sqrt{1-f^2})_{k-j}\right]$ if $k>0$,

(10). $(\cos^{-1} f)_k = -(\sin^{-1} f)_k$ if $k>0$,

(11). $(f+g)_k = (f)_k + (g)_k$ if $k\geq 0$,

(12). $(f-g)_k = (f)_k - (g)_k$ if $k\geq 0$,

(13). $(f*g)_k = \sum_{j=0}^{k} (f)_j\,(g)_{k-j}$ if $k\geq 0$,

(14). $(f/g)_k = (1/g)((g)_k - \sum_{j=1}^{k} (f)_j(f/g)_{k-j})$ if $k\geq 0$.

If f(x) only has derivatives up to a finite order n, the above rules can be simplified. For example, if $f(x)=-x^2/2$, the formula (1) simplifies to

$$(e^f)_k = (1/k) \sum_{j=1}^{\min(2,k)} j(e^f)_{k-j}\,(f)_j$$

and if f(x)=x, the formula (1) simplifies to

$$(e^f)_k = (1/k)(e^f)_{k-1}.$$

In this study, we are interested in finding the Taylor coefficients of an integrand function which can be expressed in terms of the trigonometric functions and their inverses, the exponential function e^x and its inverse ln(x), the identity function x, and the constant functions, using finite numbers of the binary operations of addition, subtraction, multiplication, and division, and the unary operations of composition and exponentiation. A general procedure for computing the required coefficients was given by Moore (1979). An outline of this procedure is:

1. Represent the integrand function by a finite code list of binary or unary operations of these elementary functions.
2. On a line-by-line basis, generate subroutines of Taylor series coefficients for each item in the list using the appropriate rules.
3. Organize the subroutines and the data handling so that the derived program will evaluate and store the Taylor coefficients in order, i.e., the first order Taylor coefficients of each item in the list then the second order coefficients, until all the required coefficients has been computed.

This procedure can be carried out on a digital computer by using either rounded interval arithmetic or scalar floating point arithmetic. In this study, we applied rounded interval arithmetic to implement the desired software. To make this procedure more understandable, we give following two examples.

Example 1: Integrand function is h(x)=tan(x).

We have the code list $T_1 = \sin x$,

$T_2 = \cos x$,

$h(x) = T_1 / T_2$.

Applying the recursion relation's rules, we obtain

$$(T_1)_k = (1/k)(\cos x)_{k-1},$$

$$(T_2)_k = -(1/k)(\sin x)_{k-1},$$

$$(T_3)_k = (1/T_2)\left[(T_1)_k - \sum_{j=1}^{k} (T_2)_j (T_1/T_2)_{k-j}\right].$$

Example 2: The integrand function is

$$h(x) = e^{-\sin x} \ln(1+x^2).$$

We can represent h(x) by the list

$$T_1 = 1+x^2,$$
$$T_2 = \sin x,$$
$$T_3 = e^{-T_2},$$
$$T_4 = \cos x,$$
$$T_5 = \ln T_1,$$
$$T_6 = T_3 * T_5.$$

Applying these rules, we obtain

$$(T_1)_k = 2x \quad \text{if } k=1$$
$$= 1 \quad \text{if } k=2,$$
$$= 0 \quad \text{if } k>2$$

$$(T_2)_k = -\frac{1}{k}(T_4)_{k-1},$$

$$(T_3)_k = (1/k) \sum_{j=1}^{k} j(e^{-T_2})_{k-j}(-T_2)_j,$$

$$(T_4)_k = -(T_2)_{k-1}/k,$$

$$(T_5)_k = \left(\frac{1}{T_1}\right)\left[(T_1)_k - \frac{1}{k}\sum_{j=1}^{k-1} j(T_5)_j(T_1)_{k-j}\right],$$

$$(T_6)_k = \sum_{j=0}^{k} (T_3)_j(T_5)_{k-j}.$$

The similar idea of computing the derivatives of an analytic function with one variable can be extended to compute the partial derivatives of an analytic function with several variables. The technique for computing the partial derivatives is very useful in the computation of the information matrix associated with a maximum likelihood estimator. The details of this technique, can be found in Moore (1979), Lawson (1988), and Rall (1981), will not be included in our discussion.

1.3 Self-validating numerical integration

Traditional numerical integration usually computes a scalar approximation of an integral

$$If = \int_a^b f(x)\,dx \qquad (1.1)$$

with or without an estimated error bound. On the contrary, the self-validation integration method using interval arithmetic always obtains an interval J=[c , d] which is guaranteed to contain the unknown integration result. Whenever the resulting interval is computed , the mid-point I^*=0.5*(c+d) can be considered as an point estimator of the unknown value If. The absolute and relative error bound of I^* are guaranteed less than 0.5*(d-c) and (d-c)/(d+c), respectively. Although the width of this computed interval depends on the actual computation error in evaluating the

integration rule and the error term, and the error term of this integration rule, we can choose an integration rule and corresponding error term such that both absolute error and relative error are satisfied within given error tolerances.

In this study, two different types of integration method will be considered. One type of integration method is based on Taylor series expansion which is suitable for integration over a finite range. Although other quadrature methods such as Newton-Cotes and Gauss-Cotes can accomplish the same task, the Taylor series method has some remarkable advantages which were identified by Corliss and Rall (1987). The details of the Taylor series method will be given in section 1.3.1. Another type of integration method is based on a continued fraction. This method will be used to deal with integrating over an infinite range. The reason for not considering any standard quadrature method for this range of integration is the difficulty of evaluating $f^{(n)}(\epsilon)$ where $\epsilon \in (-\infty, +\infty)$. The continued fraction method will be discussed in section 1.3.2.

1.3.1 Taylor Series Methods In this section, we use the Taylor series method to find an interval inclusion J of the integral (1.1) where both a and b are finite numbers. This technique was first suggested by Moore (1979) and then developed by Corliss and Rall (1987). To describe this technique, first consider the general extrapolatory form of expansion

$$If = \sum_{i=1}^{n} w_i f(x_i) + d_n (f(\xi))_p \, h^{p+1} \qquad (1.2)$$

where $h = b-a$ and $a < \xi < b$. A formula of type (1.2) will be called a quadrature formula of order p on n points. The ordinary Gauss, Newton-Cotes,

and Taylor Polynomial integration formulas have this form. Formula (1.2) gives the exact value of If when the function f is a polynomial of order p or less. But for other types of functions, we need to evaluate the error term with the unknown value ξ.

Let $r_n = \sum_{i=1}^{n} w_i f(x_i)$ and $e_n = d_n(f(\xi))_p h^{p+1}$ denote the integration rule and error term, respectively. Furthermore, let F(X) be a natural interval inclusion of f on an interval X, $(F)_p$ be a natural interval inclusion of $(f)_p$, $W_i = [\nabla w_i, \Delta w_i]$, and $X_i = [\nabla x_i, \Delta x_i]$. Then we have an interval inclusion of the integration rule

$$r_n \in R_n = \sum_{i=1}^{n} W_i F(X_i) \tag{1.3}$$

and similarly for the error term we have

$$e_n \in E_n = D_n(F(X))_p H^{p+1} \tag{1.4}$$

where $X = [\nabla a, \Delta b]$, $H = [\nabla h, \Delta h]$, and $D_n = [\nabla d_n, \Delta d_n]$. A desired interval inclusion of If is

$$If \in R_n + E_n. \tag{1.5}$$

For Taylor series method, we need to expand f at the midpoint c=(a+b)/2 of X. For $n \leq p$, and ϵ between c and x,

$$f(x) = f(c) + f'(c)(x-c) + \ldots + f^{(n-1)}(c)\,\frac{(x-c)^{n-1}}{(n-1)!} + f^{(n)}(\epsilon)\,\frac{(x-c)^n}{n!}$$

$$\in f(c)+f'(c)(x-c) + \ldots + f^{(n-1)}(c)\,\frac{(x-c)^{n-1}}{(n-1)!} + F^{(n)}(X)\,\frac{(x-c)^n}{n!} \tag{1.6}$$

where $F^{(n)}$ is an interval inclusion of $f^{(n)}$. Integrating the right member of (1.6) term by term, we have a form of (1.5)

$$\text{If} \in \sum_{i=0}^{n-1} f^{(i)}(c)\frac{(x-c)^{i+1}}{(i+1)!}\Bigg|_a^b + F^{(n)}(X)\frac{(x-c)^{n+1}}{(n+1)!}\Bigg|_a^b \subseteq J_n \tag{1.7}$$

where

$$J_n = 2\sum_{\substack{i=0 \\ i \text{ even}}}^{n-1} F^{(i)}(C)\,\frac{H^{i+1}}{(i+1)!} + \begin{cases} F^{(n)}(X)\,\dfrac{H^{n+1}}{(n+1)!} - F^{n}(X)\dfrac{H^{n+1}}{(n+1)!}, & n \text{ odd} \\ 2F^{(n)}(X)\dfrac{H^{n+1}}{(n+1)!}, & n \text{ even} \end{cases} \tag{1.8}$$

H is the interval hull $(\nabla(c-a),\nabla(b-c),\Delta(c-a),\Delta(b-c))$, and $C = [\nabla c,\Delta c]$.

Application of (1.7) will be for increasing n. The intersection of interval inclusions of If for each n will yield a sequence of approximations

$$I_0 = J_0$$

$$I_n = I_{n-1} \cap J_n \quad n=1,2,\ldots,p.$$

Computations will terminate when the successive interval I_n ceases to decrease in width or the midpoint of I_n satisfies the required absolute and relative error tolerance. The resulting interval is self-validating for If.

1.3.2 Continued Fractions Method In this section, we describe the continued fractions methods to construct an interval inclusion J of (1.1) when either |a| or |b| is infinite. Clearly, there is a close connection between certain continued fractions and infinite series. We can transform an infinite series to a continued fraction. Moreover, there is more than one continued fraction associated with an infinite series. Therefore, we can convert one type of continued fraction to another type until the final type satisfies our need. Thus, any integral which can be evaluated using an infinite series can be evaluated using continued fractions. In this study,

we only present some useful results which we need to use later. The details about continued fractions can be found in books such as Wall (1948) and Bowman and Shenton (1989).

A continued fraction is given by two sequences of real numbers $(a_n; n\geq 1)$ and $(b_n; n\geq 0)$, and can be written as

$$v = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \ldots}}}$$

or more compactly as $v = b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \ldots$. The mth convergent of a continued fraction is defined as $v_m = b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \ldots \frac{a_m}{b_m}$ for $m\geq 0$. The continued fraction is said to be converge to a finite value v if $\mathrm{Lim}_{m\to\infty} = v < \infty$. One way to evaluate a convergent continued fraction is to use the forward recurrence formula (Kennedy and Gentle 1980) which is defined through the following equations

$$\begin{aligned} c_m &= b_m c_{m-1} + a_m c_{m-2} \\ d_m &= b_m d_{m-1} + a_m d_{m-2}\ , \quad m = 1,2,\ldots \end{aligned} \tag{1.9}$$

with $c_{-1} = 1$, $d_{-1} = 0$, $c_0 = b_0$, $d_0 = 1$ and the mth convergent is given by $v_m = c_m / d_m$.

A continued fraction is called an alternating continued fraction if the even order convergents are monotone increasing and odd order convergents are monotone decreasing. Theorem 4 was stated and proved by Dudley (1987), and will be given without proof. This result will be used to obtain an interval inclusion of alternating continued fractions.

Theorem 4 For a continued fraction $v = b_0 + \frac{a_1}{b_1 +} \frac{a_2}{b_2 +} \frac{a_3}{b_3 +} \ldots$, such that $a_m \geq 0$ and $b_m > 0$ for all $m \geq 1$, we have

$$v_0 \leq v_2 \leq \ldots \leq v \leq \ldots \leq v_3 \leq v_1 \tag{1.10}$$

when it converges to $v < \infty$; if it does not converge, (1.10) remains true without the stipulation "$\leq v \leq$".

Now, let $Ia_n = [\nabla a_n , \Delta a_n]$, $Ib_n = [\nabla a_n , \Delta b_n]$, and define an interval extension of the forward recurrence formulas

$$\begin{aligned} Ic_m &= Ib_m \cdot Ic_{m-1} + Ia_m \cdot Ic_{m-2} \\ Id_m &= Ib_m \cdot Id_{m-1} + Ia_m \cdot Id_{m-2}, \quad m = 1,2,\ldots, \end{aligned} \tag{1.11}$$

with $Ic_{-1} = [1,1]$, $Id_{-1} = [0,0]$, $Ic_0 = [\nabla b_0 , \Delta b_0]$, $Id_0 = [1,1]$, and an interval inclusion of mth convergent $Iv_m = Ic_m/Id_m = [IVL_m , IVU_m]$. Applying the result of Theorem 4, we have $IVL_0 \leq IVL_2 \leq \ldots \leq v \leq \ldots \leq IVU_3 \leq IVU_1$. Suppose that we now define $JV_m = [IVL_{2m} , IVU_{2m+1}]$, $m = 0,1,2,\ldots$. Clearly, (JV_m) is a nested sequence. Therefore, the widths of successive intervals can be used as a basis for termination of computations. The resulting interval is self-validating for v.

2. EXPLANATION OF DISSERTATION FORMAT

The remainder of this dissertation consists of four parts, each of which applies some part of the self-validating computation method given in the previous chapter to compute the probabilities of one or several continuous probability functions. These methods are organized such that the resulting intervals are not only guaranteed to contain the unknown probabilities but are also narrow enough to use the midpoints of these resulting intervals as a standard to make accuracy comparison studies among competing algorithms. Each of the four parts is written as a manuscript for submission to a scientific journal. Therefore, the format of these four parts may be different in order to satisfy the requirements of the four different scientific journals.

Part I deals with the problem of two dimensional integration over a rectangle. We use this method to compute the probabilities of bivariate normal random variable over a rectangle as an example. Since the width of this resulting interval is short enough, an accuracy comparison study based on these self-validating results is given.

Part II extends the result of Part I to the computation of multidimensional integration over a rectangular region. We apply this method to the computation of probabilities of multivariate normal random variable over a rectangular region as an example. An accuracy comparison study is constructed as well.

Part III extends the result of Part I to the computation of two dimensional integration over an infinite interval. We use the computation of the probabilities of bivariate normal random variable as an example.

Part IV is an extensive study of the computation of the cumulative probabilities and percentiles of the commonly used continuous univariate random variables.

PART I.

COMPARISON OF ALGORITHMS FOR BIVARIATE NORMAL PROBABILITY OVER A RECTANGLE BASED ON SELF-VALIDATED RESULTS FROM INTERVAL ANALYSIS

Comparison of Algorithm for Bivariate Normal Probability Over a Rectangle Based on Self-Validated Result from Interval Analysis

Morgan C. Wang and William J. Kennedy

Department of Statistics

Iowa State University

Ames, Iowa 50011

ABSTRACT

Comparison of algorithms for computing probabilities and percentiles is often carried out in an effort to identify the best algorithm for various applications. One requirement when conducting comparative studies is some useable source of "satisfactory approximations to correct answers" to use as a basis when making accuracy comparisons. This paper reports success in applying elements of interval analysis to obtain a self-validating computational method for Bivariate Normal Probabilities. Results from applying this method can be used to provide a basis for accuracy studies of algorithms for Bivariate Normal probabilities. A study to compare several methods for computing probabilities over rectangles for this probability distribution, using the self-validated base values, was carried out. The paper reports a choice of best method.

Key Words and Phrases: *Interval Analysis, Normal Probabilities, Bivariate Normal, Algorithm, Error*

1. INTRODUCTION

A recently conducted study by Terza and Welland (1988) compared eight different algorithms for evaluating the Bivariate Normal CDF. Accuracy and efficiency were the criteria on which the comparisons were made. The algorithm by Divgi (1979) was judged best for each criterion.

The basis for comparing results with regard to accuracy were numbers obtained using one of the algorithms, namely that of Drezner (1978). Although it appears that the accuracy comparisons were successfully made in this study, the possibility exists that variation in levels of accuracy of the basis algorithm over different regions might have caused erroneous conclusions to be made when comparing the algorithms for achieved accuracy. What is needed in studies of this type is a base algorithm which provides a computed value along with a useful bound for the error in that value. In other words, a *self-validating* computational method and associated algorithm is needed to provide numbers for use in comparing accuracy of competing algorithms.

This paper has two major objectives. The first is to describe a method for self-validating computation of Bivariate Normal probabilities over rectangular regions. The second major objective is to give the results of a comparative study of methods for computing such probabilities. Surprisingly, we will see in later sections that even the most accurate of the algorithms currently in use for the Bivariate Normal is substantially less accurate and no more efficient than a Taylor approximation for computing probabilities over rectangles.

Self-validation of numerical computations can be achieved in many ways. We will employ interval analysis to obtain self-validation. For our application, interval analysis will be implemented in the computer so that the endpoints of the computed interval $J = [c,d]$ will be so as to guarantee that J contains the true value of the desired probability. Furthermore, we will arrange that the width $w(J) = d-c$ of the interval J can be made arbitrarily small within computing machine tolerances. Basic results which we employ are given by Moore (1979) and Corliss and Rall (1987).

In the next section elements of interval arithmetic will be reviewed, and implementation in computers will be considered. Then the problem of self-validating evaluation of integrals will be dealt with in Section 3. Application of the methodology to the Bivariate Normal distribution will be considered in Section 4, and the results of comparisons of algorithms utilizing self-validated base values will be given in Section 5 along with recommendation of best method for computing probability over rectangular regions.

2. ELEMENTS OF INTERVAL ANALYSIS

By an interval we mean a closed bounded set of real numbers $[a,b] = \{x: a \le x \le b\}$. A convenient notation for an interval is $X = [\underline{X}, \overline{X}]$ where $\underline{X}$ and $\overline{X}$ are the lower and upper limits, respectively.

Interval arithmetic employs the basic operations, defined as follows, for intervals X and Y.

$$X+Y = [\underline{X}+\underline{Y}, \overline{X}+\overline{Y}]$$

$$X-Y = [\underline{X}-\overline{Y}, \overline{X}-\underline{Y}]$$

$$X \cdot Y = [\min(\underline{X}\underline{Y}, \underline{X}\overline{Y}, \overline{X}\underline{Y}, \overline{X}\overline{Y}), \max(\underline{X}\underline{Y}, \underline{X}\overline{Y}, \overline{X}\underline{Y}, \overline{X}\overline{Y})]$$

$$1/Y = [1/\overline{Y}, 1/\underline{Y}] \text{ if } 0 \notin Y$$

$$X/Y = X \cdot (1/Y)$$

Algebraic properties of interval arithmetic are given by Moore (1979).

Let f be a real valued function of n real variables $x_1, x_2, \ldots, x_n$. The *united extension* of the function f is the set

$$\overline{f}(X_1, X_2, \ldots, X_n) = \bigcup_{(x_1, x_2, \ldots, x_n) \in (X_1, X_2, \ldots, X_n)} \{f(x_1, x_2, \ldots, x_n)\}$$

The *interval extension* of f is an interval valued function F of n interval variables $X_1, X_2, \ldots, X_n$ such that

$$F(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n)$$

for all real arguments. In other words when all of the arguments of F are degenerate intervals, F coincides with f. An interval valued function F of $X_1, X_2, \ldots, X_n$ is called *inclusion monotonic* if for subsets $Y_1, Y_2, \ldots, Y_n$ $Y_i \subseteq X_i$, $i=1,2,\ldots,n$ implies $F(Y_1, Y_2, \ldots, Y_n) \subseteq F(X_1, X_2, \ldots, X_n)$. It can be verified that interval arithmetic is inclusion monotonic.

We will be interested in finding inclusion monotonic interval extensions of certain functions. These will simply be called *interval inclusions.*

Assistance in this regard is provided by the Fundamental Theorem of Interval Analysis. This theorem states that if a real valued function f can be evaluated by a finite sequence of +, -, *, and / operations, then the same sequence of operations performed using interval arithmetic yields an interval inclusion of $\overline{f}$. The theorem can be extended to include expressions that involve monotone functions such as the exponential function, for example.

Digital computers can only approximate real arithmetic, so implementation of interval operations must be done carefully if the properties of real interval analysis are to remain valid. A recommended procedure is to utilize *rounded interval arithmetic*. By this we mean that directed roundings are used in all arithmetic operations. For computing lower endpoints, rounding is toward $-\infty$, and upper endpoints are computed with directed rounding toward $+\infty$. Thus if no underflow or overflow occurs, the interval which results at each stage of a series of arithmetic operations using intervals has floating-point number endpoints and it contains the theoretical interval of real numbers. Directed rounding can be programmed in any computer. In certain computers the floating-point hardware includes such rounding support. One example is IBM and compatible microcomputers equipped with the Numeric Processor Extension (NPX).

In subsequent sections we will use the notation ∇ and Δ to denote the downward and upward roundings, respectively from the real numbers to floating-point numbers. A computed interval result will be understood to have endpoints obtained under directed roundings.

3. SELF-VALIDATING EVALUATION OF INTEGRALS BASED ON INTERVAL ARITHMETIC AND TAYLOR SERIES

In this section we consider the problems of finding an interval inclusion J of the integral

$$IF = \int_b^a f(x)\,dx.$$

The technique suggested by Moore (1979), based on a Taylor expansion will be used. To describe this technique, first consider the general extrapolatory form of expansion

$$IF = \sum_{i=1}^{n} w_i f(x_i) + d_n \frac{f^{(p)}(\epsilon)\, h^{p+1}}{p!} \tag{3.1}$$

where $h = b-a$ and $a < \xi < b$. Formula (3.1) requires that f is differentiable p-times, and the value of ξ is unknown. The following development is essentially that of Corliss and Rall (1987).

Let $r_n = \sum_{i=1}^{n} w_i f(x_i)$ denote the *integration rule* in formula (3.1) and $e_n = d_n h^{p+1} f_p(\xi)$ denote the *error term*, where

$$f_p(\epsilon) = \frac{f^{(p)}(\epsilon)}{p!} \tag{3.2}$$

is the Taylor coefficient of order p in the expansion of $f(\xi+h)$. Furthermore, let F(X) be an interval inclusion of f on X, F_p denote an interval inclusion of f_p, $W_i = [\nabla w_i, \Delta w_i]$, and $X_i = [\nabla x_i, \Delta x_i]$. Then we have the interval inclusion of the integration rule

$$r_n \in R_n = \sum_{i=1}^{n} W_i F(X_i), \tag{3.3}$$

and similarly for the error term we have

$$e_n \in E_n = D_n H^{p+1} \cdot F_p(X), \tag{3.4}$$

where $X = [\nabla a, \Delta b]$, $H = [\nabla h, \Delta h]$, and $D_n = [\nabla d_n, \Delta d_n]$. The desired interval inclusion of If is

$$If \in R_n + E_n \tag{3.5}$$

Realization of (3.5) inevitably requires that we compute interval inclusions $F_i(C)$ of the Taylor coefficients $f_i(c)$ in the expansion of f about a point c. Moore (1979) gives methods for carrying out these computations. We will show how to obtain the inclusions for the Bivariate Normal distribution in the next section.

Next consider the expansion of f at the midpoint $c = (a+b)/2$ of X. For $n \leq p$, and ξ between c and x,

$$f(x) = f(c) + f'(c)(x-c) + \ldots + f^{n-1}(c)\frac{(x-c)^{n-1}}{(n-1)!} + f^{(n)}(\epsilon)\frac{(x-c)^n}{n!} \tag{3.6}$$

$$\epsilon\, f(c) + f'(c)(x-c) + \ldots + f^{(n-1)}(c)\frac{(x-c)^{(n-1)}}{(n-1)!} + F^{(n)}(X)\frac{(x-c)^n}{n!}$$

where $F^{(n)}$ is an interval inclusion of $f^{(n)}$. Integrating the right member of (3.6) term by term, we have a form of (3.5)

$$IF \in \sum_{i=0}^{n-1} f^{(i)}(c)\frac{(x-c)^{i+1}}{(i+1)!}\Bigg|_a^b + F^{(n)}(X)\frac{(x-c)^{(n+1)}}{(n+1)!}\Bigg|_a^b \subseteq J_n \tag{3.7}$$

where

$$J_n = 2\sum_{\substack{i=0 \\ i \text{ even}}}^{n-1} F^{(i)}(C)\frac{H^{i+1}}{(i+1)!} + \begin{cases} F^{(n)}(X)\dfrac{H^{n+1}}{(n+1)!} - F^{(n)}(X)\dfrac{H^{n+1}}{(n+1)!}, & n \text{ odd} \\ 2F^{(n)}(X)\dfrac{H^{n+1}}{(n+1)!}, & n \text{ even} \end{cases} \tag{3.8}$$

H is the interval hull $(\nabla(c-a),\nabla(b-c),\Delta(c-a),\Delta(b-c))$, and $C = [\nabla c,\Delta c]$.

Application of (3.7) will be for increasing n. The intersection of interval inclusions of If for each n will yield a sequence of approximations

$$I_0 = J_0$$

$$I_n = I_{n-1} \cap J_n \quad n=1,2,\ldots,p.$$

Computations will terminate when the successive intervals I_n cease to decrease in width. The resulting interval is self-validating for If.

4. APPLICATION TO THE NORMAL DISTRIBUTION

The Bivariate Normal probability over a rectangle can be expressed as

$$P = \int_{A_1}^{B_1} \int_{A_2}^{B_2} k \exp(-\tfrac{1}{2}\underline{Z}' V^{-1} \underline{Z}) d\underline{Z}$$

where $\underline{Z}' = (z_1, z_2)$ and V is a 2×2 real matrix. Under the transformation $\underline{Y} = T^{-1}\underline{Z}$, where T is the triangular matrix

$$T = \begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix}$$

such that $V = TT'$, we have

$$P = \int_{a_1}^{b_1} \int_{a_2(y_1)}^{b_2(y_1)} (2\pi)^{-1} \exp(-\tfrac{1}{2}(y_1^2+y_2^2)) dy_2 dy_1 \qquad (4.1)$$

where

$$a_1 = A_1/t_{11},\ b_1 = B_1/t_{11},\ a_2(y_1) = (A_2 - t_{21}y_1)/t_{22},\ b_2(y_1) = (B_2 - t_{21}y_1) / -t_{22}.$$

Let us express P in (4.1) as

$$P = \int_{a_1}^{b_1} v(y)g(y)dy \qquad (4.2)$$

where

$$u(y) = \exp(-y^2/2)$$

$$v(y) = (2\pi)^{-\frac{1}{2}} u(y)$$

$$g(y) = \int_{a_2(y)}^{b_2(y)} (2\pi)^{-\frac{1}{2}} \exp(-y_2^2/2) dy_2.$$

Now we apply (3.7) to the integral (4.2) using $c = (a_1+b_1)/2$, $X = [a_1, b_1]$, and $f(y) = v(y)g(y)$.

The main difficulty in doing this is computing the Taylor coefficients at the point c. We will use the so-called automatic differentiation technique suggested by Moore (1979) to overcome this difficulty. Using Moore's notation, the Taylor coefficients of a function f(x) at x = c are denoted as

$$(f(c))_0 = f(c)$$

$$(f(c))_k = \frac{1}{k!} \frac{d^k f(x)}{dx^k} \bigg|_{x=c} \qquad k=1,2,\ldots$$

Expressions for sum and product coefficients using differentiable functions s(x) and t(x), say, yield the forms

$$(s(c) + t(c))_k = (s(c))_k + (t(c))_k$$

$$(s(c)t(c))_k = \sum_{j=0}^{k} (s(c))_j (t(c))_{k-j}.$$

For selected elementary functions $(t(x))^a$ and $e^{s(x)} = \exp(s(x))$, we have

$$\left[(t(c))^a\right]_k = \frac{1}{t(c)} \sum_{j=0}^{k-1} (a-j(a+1)/k)(t(c))_{k-j} \left[(t(c))^a\right]_j$$

$$(e^{s(c)})_k = \sum_{j=0}^{k-1} (1-j/k)(e^{s(c)})_j (s(c))_{k-j}.$$

These expressions are utilized in a straight-forward manner to find, with respect to v(y) and g(y) in (4.2),

$$(g(c))_0 = [\Phi(b_2(c)) - \Phi(a_2(c))]$$

$$(g(c))_1 = (t_{21}/(t_{22}\sqrt{2\pi}))[u(a_2(c)) - u(b_2(c))]$$

$$(g(c))_k = \frac{1}{k}(t_{21}/(t_{22}\sqrt{2\pi}))[(u(a_2(c)))_{k-1} - (u(b_2(c)))_{k-1}] \qquad (k > 1)$$

$$(v(c))_0 = (2\pi)^{-\frac{1}{2}} (u(c))$$

$$(v(c))_k = \sum_{i=0}^{k-1} (1-i/k)(v(c))_i (u(c))_{k-i} \quad (k \geq 1)$$

$$(v(c)g(c))_k = \sum_{j=0}^{k} (v(c))_j (g(c))_{k-j} \quad (k \geq 0) \tag{4.3}$$

where

$$u(c) = -c^2/2$$

$$a_2(c) = (A_2 - t_{21}c)/t_{22}$$

$$b_2(c) = (B_2 - t_{21}c)/t_{22}$$

and Φ is the cumulative distribution function of the univariate standard normal distribution. Values of $(v(c)g(c))_k$ can be computed beginning with k=0 and progressing through as many successive coefficients as are needed. Note that the only integral involved in $(v(c)g(c))_0$ is Φ.

The right member of (4.3) involves only rational and monotone nondecreasing functions. Interval extensions of these functions are readily obtained so that the interval extension of (4.3), for use in (3.8) applied to the integral (4.2), is readily available. Interval computation of univariate standard normal probabilities are easily obtained using the Taylor series form (3.8) and automatic differentiation similar to that shown above. Details of this simpler application will not be given.

5. COMPARISON OF ALGORITHMS FOR BIVARIATE NORMAL PROBABILITY OVER A RECTANGLE

Software to support rounded interval analysis of desired probabilities was written for an IBM microcomputer equipped with an Intel 80287 Numeric Processor Extension. This software utilized the interval Taylor expansion with interval error term and automatic differentiation described in Section 4. Using appropriate numbers n of terms in the expansion, interval inclusions of desired probabilities whose midpoints were guaranteed to have relative error of less than 10^{-15} were easily obtained over a very large region of the variables and parameter space. This software gave the needed ability to compute base values for use in a comparative study of scalar algorithms which we will next consider.

Probabilities over rectangles are currently usually computed using appropriate linear combinations of values of the CDF at the corners of the rectangles. The reason for this is that software which directly gives the desired probability over a rectangle is not generally available. Unfortunately, it is not unusual for severe cancellation to occur when forming the linear combination so that computed probabilities are frequently rather inaccurate. Of course this is disastrous in situations that need good approximations to the probabilities. One possible solution to this problem is to use interval analysis software to provide answers having guaranteed accuracy. This is not an especially desirable solution because the nature of the software is specialized and it requires larger amounts of computer time in execution than traditional scalar based software. What we desire is, of course, sufficiently accurate scalar algorithms.

The success of interval computation of the Taylor expansion in the form

(3.7) for the Bivariate Normal integral suggested the possibility that simply evaluating the finite sum (our integration rule) using scalar arithmetic might yield useful approximations to the desired probabilities. It was decided to try what we will call the "scalar implementation." Thus in an effort to successfully compute probabilities over rectangular regions, we used double precision FORTRAN and coded a scalar implementation of the Taylor integration using automatic differentiation as given in (4.3) and expanding n terms. The value of n used differed in various regions. It was assigned a maximum value 200 based on experience obtained with the interval version. Each expansion was carried to the point where either terms became sufficiently small or n=200 was reached. No attempt was made to adjust the n-term approximation for the error term, other than judicious choice of n. The scalar implementation was compared to a FORTRAN double precision version of the Divgi(1979) algorithm, judged best by Terza and Welland(1988), in an effort to determine which method was preferable. The interval algorithm was used to provide accurate values on which to base the comparisons. Over the region in which each variables' range is in the interval [-7,7], and the correlation ρ is allowed to vary from 0.1 through 0.9, the scalar implementation of the Taylor integration maintained excellent accuracy as judged by comparison with self-validated computed intervals. We will next give a summary of some of the data obtained in this study.

An absolute relative error measure, obtained using the midpoint of the computed interval inclusion of the probability as the "*true value*", was defined as

$$RE = |\text{true value} - \text{computed value}|/(\text{true value}).$$

Given that the width of the interval inclusion in each case was always very small, the "*true value*" used was necessarily sufficiently close to the desired probability.

Figure 1 summarizes relative error data obtained by integrating separately over each of the 196 unit squares in the square region $[-7,7]\times[-7,7]$, for each indicated correlation ρ. Divgi's algorithm does not reliably compute very small probabilities, so only those unit squares which yielded a probability of 10^{-15} or greater were actually used to prepare Figure 1. The mean of relative errors for each method are plotted separately for each ρ in the figure. Table 1 gives some of the means and standard deviations actually computed.

Figure 2 shows average relative errors for all ρ values for those unit square areas in each of four nested regions. The regions are Region 1 = $([-2,2]\times[-2,2])$, Region 2 = $([-4,4]\times[-4,4])$, Region 3 = $([-6,6]\times[-6,6])$, and Region 4 = $([-7,7]\times[-7,7])$. Predictably, the Divgi method performs best in the first region and its performance degrades as one moves further into tail regions. In all of these integrations the scalar implementation method was simply applied to the given integral with no attempt (other than adjustable n) made to increase accuracy. However, if one uses composite quadrature, which involves summing results from applying the scalar implementation over subsets of the overall rectangle of integration, then accuracy to essentially machine precision in computed probability can be achieved in most cases. Composite quadrature was used to investigate the behavior of the scalar implementation versus the Divgi method for larger ρ values. The results of this investigation will now be given. Table 2 defines twelve integrals, each

with large ρ, and gives an interval inclusion for each. The first six integrals are defined over a square region $[a_1,b_1] = [a_2,b_2]$. Table 3 shows the results obtained for each integration using the scalar implementation with composite quadrature. The subregions are those obtained by dividing the interval $[a_1,b_1]$ into subintervals having length 0.25. Thus the value of the first integral was obtained as the sum of integrations over $[-0.5,-0.25] \times [-0.5,0.5]$; $[-0.25,0.0] \times [-0.5,0.5]$; $[0.0,0.25] \times [-0.5,0.5]$; and $[0.25,0.5] \times [-0.5,0.5]$. Table 4 gives results obtained using Divgi's method in each of the twelve cases. Note that the Divgi algorithm yields excellent results when the true probability is fairly large, but cancellation degrades the result as the true value becomes small relative to the CDF values involved. Computing times in Tables 3 and 4 are to the nearest second.

A new algorithm for CDF approximation was brought to the attention of the authors during preparation of this paper. This algorithm, given by Drezner and Wesolowsky, appears to be a competitor for the Divgi algorithm. Since the results of the studies conducted in this paper are not strongly influenced by the accuracy of CDF approximation, given the reasonably accurate approximations are obtained, it was decided not to incorporate the as yet unpublished Drezner and Wesolowsky work in this paper.

The results of the studies given above show that the algorithm based on Taylor expansion of the integrand and automatic differentiation is clearly the best choice based solely on accuracy comparisons. Fortunately, the scalar implementation also required approximately the same amount of computer time as the Divgi method for performing most tasks. Over very large regions

of integration, the scalar implementation does require several seconds more computer time than that used by the Divgi method, but the actual amount of time needed is not large. Thus the scalar implementation seems to be a clearly preferable method for this application.

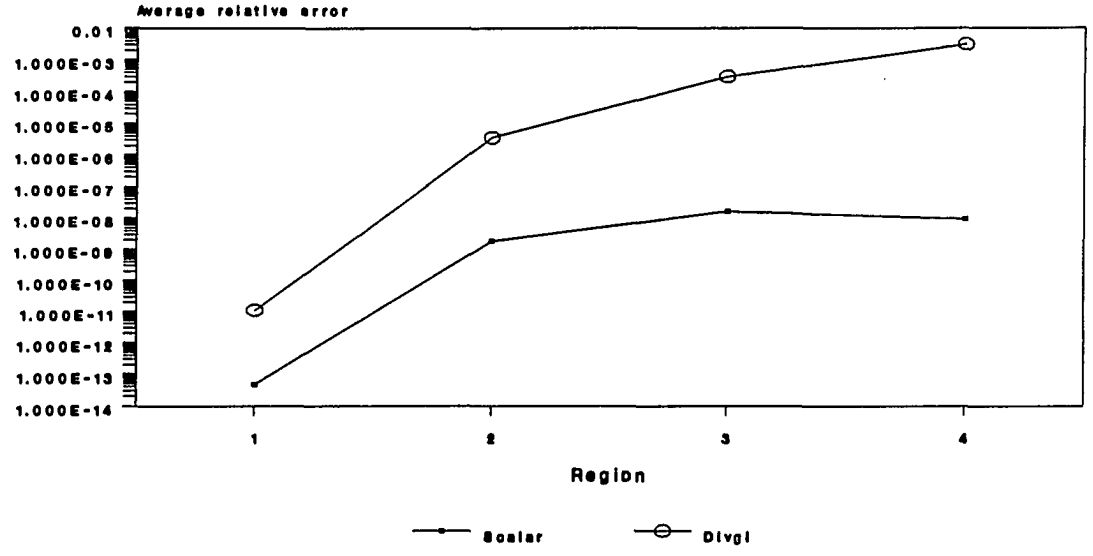

Figure 1 Average relative error by correlation

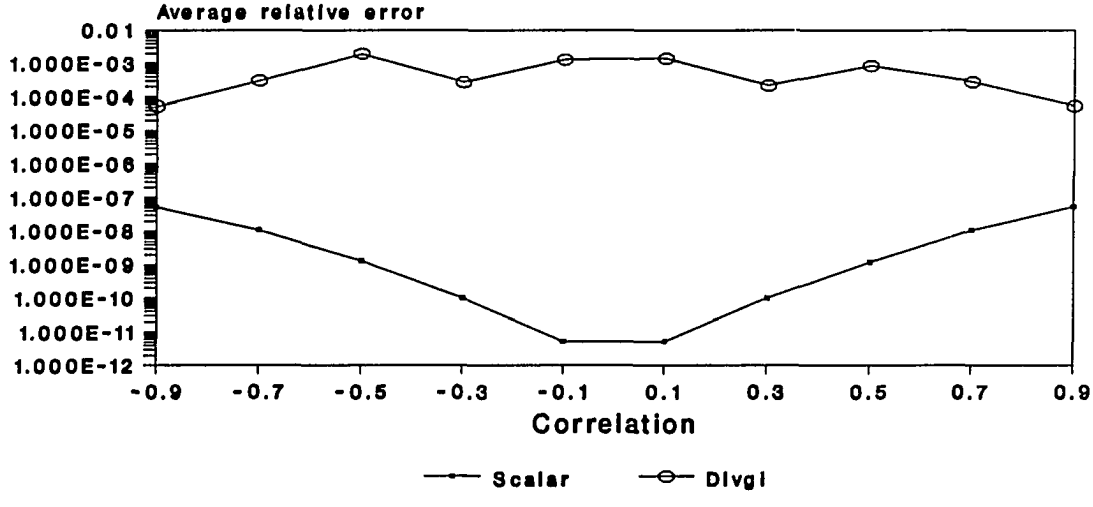

Figure 2 Average relative error by region

Table 1 Means and standard deviations of relative errors over unit squares

	Scalar Implementation		Divig's Method	
	Mean R.E.	Std. R.E.	Mean R.E.	Std. R.E.
0.9	5*10d-8	3*10d-7	5*10d-5	2*10d-4
0.7	1*10d-8	8*10d-8	3*10d-4	2*10d-3
0.5	1*10d-9	1*10d-8	2*10d-3	1*10d-2
0.3	1*10d-10	8*10d-10	3*10d-4	2*10d-3
0.1	5*10d-12	1*10d-15	1*10d-3	6*10d-3

Table 2 Interval inclusion of selected bivariate normal probabilities

	Range			Inclusion of Probability	
I.D.	A(I)	B(I)	ρ	Lowerbound	Upperbound
1	-0.5	0.5	0.999	0.370361572466195d0	0.370361572466196d0
2	-1.0	1.0	0.997	0.667734766057642d0	0.667734766057643d0
3	-1.28	1.28	0.995	0.785428010364473d0	0.785428010364474d0
4	-1.64	1.64	0.993	0.889189871675273d0	0.889189871675274d0
5	-1.96	1.96	0.991	0.943761605250668d0	0.943761605250669d0
6	-2.58	2.58	0.990	0.988513410907603d0	0.988513410907604d0
7 (a)	0.15	0.50	0.999	0.251550397591894d-13	0.251550397591895d-13
8 (b)	0.20	1.25	0.997	0.677641964063935d-09	0.677641964063936d-09
9 (c)	0.35	1.35	0.995	0.703116302535518d-14	0.703116302535519d-14
10 (d)	0.45	1.45	0.993	0.848115623421181d-16	0.848115623421188d-16
11 (e)	0.50	2.25	0.991	0.316093774041186d-15	0.316093774041189d-15
12 (f)	0.50	2.50	0.990	0.592400794268086d-14	0.592400794268090d-14

(a). A(2)=-0.5 B(2)=-0.15 (b). A(2)=-1.25 B(2)=-0.20 (c). A(2)=-1.35 B(2)=-0.35

(d). A(2)=-1.45 B(2)=-0.45 (e). A(2)=-2.25 B(2)=-0.50 (f). A(2)=-2.50 B(2)=-0.50

Table 3 Scalar implementation results for integrals in Table 2

I.D.	Probability	Absolute Error	Relative Error	Time (Sec.)
1	0.370361572466195d0	0.1d-15	0.1d-15	2.0
2	0.667734766057642d0	0.1d-15	0.1d-15	2.0
3	0.785428010364473d0	0.1d-15	0.1d-15	3.0
4	0.889189871675273d0	0.1d-15	0.1d-15	4.0
5	0.943761605250668d0	0.1d-15	0.1d-15	5.0
6	0.988513410907604d0	0.1d-15	0.1d-15	7.0
7	0.251550397591894d-13	0.1d-28	0.1d-15	1.0
8	0.677641964063933d-9	0.2d-24	0.2d-15	1.0
9	0.703116302535517d-14	0.2d-28	0.2d-15	1.0
10	0.848115623421181d-16	0.1d-31	0.1d-15	1.0
11	0.316093774041189d-15	0.2d-30	0.2d-15	1.0
12	0.592400794268090d-14	0.1d-29	0.1d-15	1.0

Table 4 Divgi's algorithm results for integrals in Table 2

I.D.	Probability	Absolute Error	Relative Error	Time (Sec.)
1	0.370361572466195d0	0.1d-15	0.1d-15	1.0
2	0.667734766057642d0	0.1d-15	0.1d-15	1.0
3	0.785428010364473d0	0.1d-15	0.1d-15	1.0
4	0.889189871675273d0	0.1d-15	0.1d-15	1.0
5	0.943761605250668d0	0.1d-15	0.1d-15	1.0
6	0.988513410907604d0	0.1d-15	0.1d-15	1.0
7	0.251465515077598d-13	0.1d-17	0.5d-3	1.0
8	0.677642063581451d-9	0.1d-14	0.2d-4	1.0
9	0.704991620636974d-14	0.1d-17	0.2d-2	1.0
10	0.555111512312578d-16	0.3d-16	0.4d0	1.0
11	0.333066907387547d-15	0.2d-17	0.7d-1	1.0
12	0.593969318174459d-14	0.1d-17	0.1d-2	1.0

6. SUMARRY AND CONCLUSIONS

Comparison of algorithms for computing probabilities and percentiles is frequently required. One item needed when conducting comparative studies in this area is some useable source of "correct answers" to use as a basis for comparisons relative to accuracy.

The present paper reports success in applying elements of interval analysis to obtain a method for self-validating computation of Bivariate Normal probabilities. Development of this interval method suggested the possibility of using a scalar truncated form of the interval method to rapidly approximate the probabilities. Investigation of the scalar implementation, using midpoints of short intervals containing the true value as a basis, showed that the scalar implementation is indeed superior to methods currently used to find probabilities over rectangles. The methods in use employ the CDF and are those tested by Terza and Welland (1988), namely those given by Bouver and Bargmann (1979), Parrish and Bargmann (1981), Daley (1974), Divgi (1979), Drezner (1978), IMSL (1987), and Young and Minder (1974). Several of the algorithms implementing this methodology are based on work reported by Owen (1965) and Donnelly (1973). The utility of self-validating computing methodology seems to be proved, at least in this application. Research is currently underway to extend the self-validating method to the multivariate normal distribution and to develop interval-based methods for self-validation computing relative to other probability distributions. Results obtained to date are very encouraging.

7. ACKNOWLEDGMENTS

The authors wish to thank the Associate Editor and referees for helpful suggestions which improved the presentation in this paper.

8. REFERENCES

Bouver, H. and Bargmann, R. E. (1979), Comparison of Computational Algorithms for the Evaluation of the Univariate and Bivariate Normal Distribution, Proceedings of the Computer Science and Statistics: 12th Annual Symposium on the Interface, 344-348.

Corliss, G. F. and Rall L. B. (1987), Adaptive, Self-Validating Numerical Quadrature, SIAM J. SCI. STAT. COMPUT. 8, 831-847.

Daley, D. J. (1974), Computation of Bi- and Tri-variate Normal Integrals, Applied Statistics 23, 435-438.

Divgi, D. R. (1979), Calculation of Univariate and Bivariate Normal Probability Functions, Annals of Statistics 7, 903-910.

Donnelly, T. G. (1973), Algorithm 462: Bivariate Normal Distribution, CACM 16, 636.

Drezner, Z. (1978), Computation of the Bivariate Normal Integral, Math. Comp. 32, 277-279.

Drezner, Z. and Wesolowsky, G. O. (1990), On the Computation of Bivariate Normal Integral, to appear in JSCS.

IMSL (1987), IMSL STAT/LIBRARY User's Manual, Houston, IMSL Inc., 917-918.

Moore, R. E. (1979), Methods and Applications of Interval Analysis, SIAM Stud. Appl. Math. 2, SIAM, Philadelphia.

Owen, D. B. (1965), Tables for Computing Bivariate Normal Probabilities, Annals of Mathematical Statistics 27, 1075-1090.

Owen, D. B. (1965), A Special Case of the Bivariate Non-Central t-Distribution, Biometrika 52, 437-446.

Parrish, R. S. and Bargmann, R. E. (1981), A Method for the Evaluation of

Cumulative Probabilities of Bivariate Distributions Using the Pearson Family, Statistical Distributions in Scientific Work 5, 241-257.

Terza, J. V. and Welland, U. (1988), Accuracy and Efficiency Comparisons of Bivariate Normal Approximation Algorithms, Working Paper 12-87-2, Dept. of Economics, Pennsylvania State University.

Young, J. C. and Minder, C. E. (1974), Algorithm AS76: An Integral Useful in Calculating Non-Central t and Bi-variate Normal Probabilities, Applied Statistics 23, 455-457.

PART II

A NUMERICAL METHOD FOR ACCURATELY APPROXIMATING MULTIVARIATE NORMAL PROBABILITIES

A Numerical Method for Accurately Approximating Multivariate Normal Probabilities

Morgan Wang and W. J. Kennedy

Department of Statistics

Iowa State University

Ames, Iowa 50011

ABSTRACT

A Taylor series expansion of the multivariate normal integral is used to calculate the value of the integral over rectangular regions. Interval analysis and automatic differentiation provide self-validation for calculated probabilities. In examples, the Taylor series approximation gives more accurate results than the algorithm of Schervish (1984).

Key words: Interval Analysis, Automatic Differentiation, Multivariate Normal Probability

1. INTRODUCTION

This paper has two objectives. The first is to describe a method for self-validating computation of Multivariate Normal Probabilities over rectangular regions. The second objective is to give the results of a comparative study of the accuracy of two different algorithms for approximating such probabilities. *Self-validating numerical* computation is sometimes called *automatic error analysis*, and it can be achieved in more than one way. We will use interval analysis to obtain self-validation. This means we compute an interval that is guaranteed to contain the theoretically correct value of the desired probability. Then the midpoint of the interval is the scalar approximation and the half width of the interval is a guaranteed error bound giving validity to the scalar approximation. Since very short intervals are computed, the approximations obtained provide essentially correct answers to use as a basis for comparing the accuracy of outputs from competing scalar algorithms. Self-validating computing based on interval analysis is far more costly in terms of computer time than the usual floating-point scalar computing, so *it is not a general purpose computing tool*. Rather it is useful in special situations such as the one to be described here. Basic elements of interval analysis are given by Moore (1979), Ratschek and Rokne (1984), Alefeld and Herzberger (1983), and Corliss and Rall (1987). The following brief description sets the notation used in later discussion.

By an interval we mean a closed bounded set of real numbers $X = [\underline{X},\overline{X}] = (x:\underline{X}\leq x\leq\overline{X})$. Interval arithmetic employs the basic operations, defined as follows, for intervals X and Y,

$$X + Y = [\underline{X}+\underline{Y}, \overline{X}+\overline{Y}]$$

$$X - Y = [\underline{X}-\overline{Y}, \overline{X}-\underline{Y}]$$

$$X \cdot Y = [\min(\underline{X}\underline{Y}, \underline{X}\overline{Y}, \overline{X}\underline{Y}, \overline{X}\overline{Y}),\ \max(\underline{X}\underline{Y}, \underline{X}\overline{Y}, \overline{X}\underline{Y}, \overline{X}\overline{Y})] \tag{1.1}$$

$$1 / Y = [1/\overline{Y}, 1/\underline{Y}],\ 0 \notin Y$$

$$X / Y = X \cdot (1/Y)$$

Algebraic properties of interval arithmetic are given by Moore (1979).

Let f be a real valued function of n real variables $x_1, x_2, \ldots, x_n$, defined on real intervals $X_1, X_2, \ldots, X_n$, respectively. The *united extension* of the function f is over $X_1, \ldots, X_n$ is

$$\overline{f}(X_1, \ldots, X_n) = \bigcup_{(x_1, x_2, \ldots, x_n) \in (X_1, X_2, \ldots, X_n)} (f(x_1, x_2, \ldots, x_n)).$$

The *interval extension* of f is an interval valued function F of n interval valued variables such that

$$F(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n)$$

for all real arguments. In other words when all of the arguments of F are degenerate intervals, F coincides with f. An interval valued function F of $X_1, X_2, \ldots, X_n$ is *inclusion monotonic* if for any subsets $Y_1, Y_2, \ldots, Y_n$ of $X_1, X_2, \ldots, X_n$ respectively $F(Y_1, Y_2, \ldots, Y_n) \subseteq F(X_1, \ldots, X_n)$. We will be interested in finding inclusion monotonic interval extensions of certain functions. These will simply be called *interval inclusions*. When the scalar valued function f is a rational function of variables and other functions, an interval inclusion can be obtained by substituting interval operations for the corresponding scalar operations. Of course an interval inclusion for each function involved in the rational expression must be available. The resulting interval valued expression is called the *natural interval extension* of f.

For our application, we will finally compute an interval inclusion IP = $[\underline{P},\overline{P}]$ of the true probability P. Moreover, we will arrange that the width $w(IP) = \overline{P} - \underline{P}$ of this interval is small enough so that $w(IP) \leq 2\alpha$ and $w(IP) \leq \rho \cdot |\underline{P}+\overline{P}|$, where α, ρ are required absolute and relative error bounds, respectively.

Implementation of interval operations in a digital computer must be done very carefully if the properties of real interval analysis are to remain valid when the floating-point system is used to obtain interval inclusions. A recommended procedure is to employ directed rounding. By this we mean that rounding is toward $-\infty$ when computing lower endpoints of the interval and toward $+\infty$ for upper endpoints. Therefore, the computed interval obtained in rounded interval computing will have floating-point endpoints and contain the real interval at every stage of the computations.

In the next section, we will describe a method for obtaining interval inclusion of multivariate normal probability. Then results of a comparative study of two competing algorithms for approximating normal probabilities will be given in the third section. Finally, we will make recommendations of best method in Section 4.

2. SELF-VALIDATING EVALUATING OF MULTIVARIATE NORMAL INTEGRALS BASED ON INTERVAL ARITHMETIC AND TAYLOR SERIES

Given a random vector $\underset{\sim}{z}$ having the N-variate Normal distribution with mean vector $\underset{\sim}{0}$ and positive definite covariance matrix V, probability P over a rectangular region R can be expressed as

$$P = \int_{A_1}^{B_1} \int_{A_2}^{B_2} \cdots \int_{A_N}^{B_N} \frac{1}{(2\pi)^{N/2}|V|^{\frac{1}{2}}} \exp(-\tfrac{1}{2}\underset{\sim}{z}' V^{-1} \underset{\sim}{z}) d\underset{\sim}{z}$$

where $R = \{(z_1,\ldots,z_n) | A_i \le z_i \le B_i, 1 \le i \le N\}$.

Since V is positive definite and symmetric, there exists a lower triangular matrix

$$L = \begin{bmatrix} \ell_{11} & & & \\ \ell_{12} & \ell_{22} & & 0 \\ \vdots & & \ddots & \\ \ell_{1N} & \cdots & \cdots & \ell_{NN} \end{bmatrix}$$

such that $V = LL'$. Then under the transformation $\underset{\sim}{Y} = L^{-1}\underset{\sim}{Z}$, we have

$$P = \int_{a_1}^{b_1} \int_{a_2(y_1)}^{b_2(y_1)} \cdots \int_{a_N(y_1,\ldots,y_{N-1})}^{b_N(y_1,\ldots,y_{N-1})} K \prod_{i=1}^{N} f(y_i) dy_N,\ldots,dy_1 \qquad (2.1)$$

where $f(y_i) = \exp(-y_i^2/2)$, $i=1,2,\ldots,N$; $a_1 = A_1/\ell_{11}$; $b_1 = B_1/\ell_{11}$;

$$a_i(y_1,\ldots,y_{i-1}) = (A_i - \sum_{j=1}^{i-1} \ell_{ij} y_j)/\ell_{ii}, \quad i=2,\ldots,N$$

$$b_i(y_1,\ldots,y_{i-1}) = (B_i - \sum_{j=1}^{i-1} \ell_{ij} y_j)/\ell_{ii}, \quad i=2,\ldots,N$$

$$K = (1/2\pi)^{N/2}.$$

Let us express P in (2.1) as

$$P = \int_{a_1}^{b_1} \int_{a_2(y_1)}^{b_2(y_1)} \cdots \int_{a_{N-1}(y_1,\ldots,y_{N-2})}^{b_{N-1}(y_1,\ldots,y_{N-2})} K \prod_{i=1}^{N-1} f(y_i) \cdot g(y_1,\ldots,y_{N-1}) dy_{N-1},\ldots,dy_1 \tag{2.2}$$

where $g(y_1,\ldots,y_{N-1}) = \int_{a_N(y_1,\ldots,y_{N-1})}^{b_N(y_1,\ldots,y_{N-1})} f(y_N) dy_N$.

To illustrate the general procedure of transforming P to a desired form, we use N=3 for example and express (2.2) as

$$P = K \int_{a_1}^{b_1} f(y_1) \int_{a_2(y_1)}^{b_2(y_1)} f(y_2) g(y_1,y_2) dy_2 dy_1$$

$$- \int_{a_1}^{b_1} f(y_1) \int_{a_2(y_1)}^{b_2(y_1)} h(y_1,y_2) dy_2 dy_1 \tag{2.3}$$

Now we expand $h(y_1,y_2)$ in a Taylor Series with respect to y_2 at $y_2 = c_2(y_1) = (a_2(y_1) + b_2(y_1))/2$ to even order n_2 and denote the ith order Taylor Series coefficient as $(h(y_1,c_2(y_1)))_i$. This series has the form

$$h(y_1,y_2) = \sum_{i=0}^{n_2-1} \left[h(y_1,c_2(y_1))\right]_i (y_2-c_2(y_1))^i + \left[h(y_1,\epsilon_2(y_1))\right]_{n_2} (y_2-\epsilon_2(y_1))^{n_2} \tag{2.4}$$

where $\epsilon_2(y_1)$ is in the closed interval $[a_2(y_1),b_2(y_1)]$ denoted by $X_2(y_1)$.

Since $f(y_1)\left[h(y_1,c_2(y_1))\right]_i$ is a function of y_1 for each $i=0,2,\ldots,n_2-2$, we will denote it by $u_i(y_1)$. Next we replace $h(y_1,y_2)$ in (2.3) by (2.4) and integrate (2.3) with respect to y_2. This gives P in the form

$$P = 2K \sum_{\substack{i=0 \\ i \text{ even}}}^{n_2-2} \int_{a_1}^{b_1} u_i(y_1) \frac{d_2^{(i+1)}}{(i+1)} dy_1 + E_1 \tag{2.5}$$

where the remainder term E_1 is

$$E_1 = 2K \frac{d_2^{(n_2+1)}}{(n_2+1)} \int_{a_1}^{b_1} f(y_1)\left[h(y_1, \epsilon_2(y_1))\right]_{n_2} dy_1 \tag{2.6}$$

with $d_2 = (B_2-A_2)/(2\ell_{22}) = b_2(y_1) - c_2(y_1)$. For each i, we expand $u_i(y_1)$ in Taylor Series with respect to y_1 at $y_1 = c_1 = (a_1+b_1)/2$ up to even order n_1. Thus, we obtain

$$u_i(y_1) = \sum_{j=0}^{n_1-1} (u_i(c_1))_j (y_1-c_1)^j + (u_i(\epsilon_{1i}))_{n_1} (y_1-\epsilon_{1i})^j \tag{2.7}$$

where $\epsilon_{1i} \in [a_1,b_1] = X_1$.

Now, we replace $u_i(y_1)$ in (2.5) by (2.7) and integrate (2.5). This gives P in the form

$$P = 2^2K \cdot \sum_{\substack{i=0 \\ i \text{ even}}}^{n_2-2} \left\{ \frac{d_2^{(i+1)}}{(i+1)} \sum_{\substack{j=0 \\ j \text{ even}}}^{n_1-2} (u_i(c_1))_j \frac{d_1^{(j+1)}}{(j+1)} \right\} + E_1 + E_2 \tag{2.8}$$

where the remainder term E_2 is

$$E_2 = 2^2K \frac{d_1^{(n_1+1)}}{(n_1+1)} \frac{d_2^{(n_2+1)}}{(n_2+1)} \sum_{\substack{i=0 \\ i \text{ even}}}^{n_2-2} (u_i(\epsilon_{1i}))_{n_1} \tag{2.9}$$

with $d_1 = (b_1-c_1) = (c_1-a_1) = (B_1-A_1)/2\ell_{11}$.

This is the form that will be used to derive an interval inclusion IP of P. While constructing IP, every interval extension of a real valued function in the expression of P must have inclusion monotonicity. The following four results from the theory of interval *analysis* are well known and will be given without proof (c.f. Moore (1979), pp 20-24). These results will be used

later to insure inclusion monotonicity.

(R_1) The united extension of a real-valued function is inclusion monotonic.

(R_2) An interval arithmetic function as given by (1.1) is the united extension of the associated real valued function, i.e., it is inclusion monotonic.

(R_3) A *rational interval function*, i.e., a function whose interval values are defined by a specific finite sequence of interval arithmetic operations, is inclusion monotonic.

(R_4) The interval extension of a monotone function, such as the exponential function, coincides with its united extension, i.e., it is inclusion monotonic.

The functions $(u_i(c_1))_j$ in P are each a linear combination of the product of a polynomial and an exponential function, i.e., are each a linear combination of the product of a rational function and monotone function. Therefore, (R_2), (R_3) and (R_4) insure that the natural interval extension of $(u_i(c_1))_j$, denoted by $(U_i(C_1))_j$, is inclusion monotonic for every i and j. It follows that the interval extension of the integration rule of P, which can be written as

$$IR = 2^2K \cdot \sum_{\substack{i=0 \\ i \text{ even}}}^{n_2-2} \left\{ \frac{D_2^{(i+1)}}{(i+1)} \sum_{\substack{j=0 \\ j \text{ even}}}^{n_1-2} (U_i(C_1))_j \frac{D_1^{(j+1)}}{(j+1)} \right\} \tag{2.10}$$

where $D_1 = [d_1, d_1]$, $D_2 = [d_2, d_2]$, $C_1 = [c_1, c_1]$, is inclusion monotonic.

Similarly, the interval extension IE_2 of E_2 is inclusion monotonic, where IE_2 is

$$IE_2 = 2^2 K \frac{D_1^{(n_1+1)}}{(n_1+1)} \frac{D_2^{(n_2+1)}}{(n_2+1)} \sum_{\substack{i=0 \\ i \text{ even}}}^{n_2-2} (U_i(X_1))_{n_1} \tag{2.11}$$

Next consider the error term E_1. We know that $f(y_1)$ is an integrable function which does not change sign in the interval X_1, and $\left(h(y_1, \epsilon_2(y_1))\right)_{n_2}$ is a continuous function on X_1. By the second mean value theorem for definite integrals (c.f. Johnson and Riess (1982)) there exists a $\delta \in X_1$ such that

$$\int_{a_1}^{b_1} f(y_1)\left(h(y_1, \epsilon_2(y_1))\right)_{n_2} dy_1 = \left(h(\delta, \epsilon_2(\delta))\right)_{n_2} \int_{a_1}^{b_1} f(y_1) dy_1 \tag{2.12}$$

An interval inclusion of the univariate integral in (2.12) can easily be formed using Taylor expansion and interval inclusion of both integration rule part and error term as suggested by Corliss and Rall (1987). Then let $X_2(X_1) = [\nabla a_2(X_1), \Delta b_2(X_1)]$ be the interval hull of $\bigcup_{y_1 \in X_1} X_2(y_1)$, where ∇ and Δ denote directed downward and upward roundings, respectively. Using the argument which gave us that the interval extension of $(u_i(c_1))_j$ is inclusion monotonic, we can show the interval inclusion of $\left(h(\delta, \epsilon_2(\delta))\right)_{n_2}$, denoted by $\left(H(X_1, X_2(X_1))\right)_{n_2}$, is inclusion monotonic. Therefore, the interval inclusion IE_1 of E_1 is inclusion monotonic, where

$$IE_1 = 2K \frac{D_2^{(n_2+1)}}{(n_2+1)} \left(H(X_1, X_2(X_1))\right)_{n_2} \cdot IF \tag{2.13}$$

and IF is an interval inclusion of $\int_{a_1}^{b_1} f(y_1)\, dy_1$. Thus the interval inclusion of P is

$$IP = IR + IE_1 + IE_2. \tag{2.14}$$

The steps in deriving IP for the trivariate case are indicative of what can be done in higher dimensions. The authors have derived and programmed self-validating support for dimensions through n=4. Certainly the degree of complexity increases substantially with dimension, but the ability to achieve high quality self-validating approximations in general is not available otherwise.

Equation (2.8) expresses P in terms of derivatives of various functions. If it were necessary to obtain and use algebraic expressions for the various higher order derivatives, use of (2.8) would not be practical. Fortunately we do not need mathematical expressions for these derivatives and Taylor coefficients. The numerical tool called by various names including *automatic differentiation* and *differentiation arithmetic* is available for application here. Automatic differentiation takes advantage of the fact that successively higher order derivatives, each evaluated at the same point c(say), are all that we require. This being the case, we first express the given function algebraically in terms of binary operations and elementary functions for which Taylor coefficients at c are readily produced. Then n-tuples of Taylor coefficients for the various elementary functions are combined according to the usual rules of differential calculus applied to the function in question. The resulting n-tuple is the desired n-tuple of Taylor coefficients for the given function. Interval inclusions of these coefficients are obtained by substituting interval operations for the associated scalar operations. Moore (1979), Rall (1981), Corliss (1988),

Lawson (1988), and Jerrell (1989) give useful descriptions of automatic differentiation. Thus using rounded interval arithmetic and automatic differentiation, an interval inclusion of IP in (2.14) is not difficult to compute.

Software to support rounded interval analysis of desired probabilities was written for an IBM microcomputer equipped with an Intel 80287 Numeric Processor Extension. This software includes interval arithmetic and some utility routines such as the interval logarithm and interval exponential functions, and is available upon request from the authors. It can compute an interval inclusion IP = $[\underline{P},\overline{P}]$ of true probability P such that the width of this interval is extremely small subject to the limitations in expression of numeric values in floating-point. In our applications we used composite quadrature and set an upper bound of 24 terms in each Taylor expansion. If the resulting interval inclusion did not satisfy the error bound, the region of integration was further subdivided. This process continued until an interval inclusion was obtained which had width that did not exceed the specified bound.

3. COMPARISON OF ALGORITHMS FOR MULTIVARIATE NORMAL PROBABILITY OVER A RECTANGULAR REGION

Although there are many scalar algorithms such as Gupta (1964), Milton (1972), Bohrer and Schervish (1981), Schervish (1984), Genz and Kahaner (1988), and Plant and Quandt (1989) for obtaining multivariate normal probabilities, we wish to only consider algorithms which provide an error measure on the computed result and put only one restriction, namely a positive definite covariance matrix, on the parameter space. Among available algorithms, Schervish (1984) is the only one which satisfies these conditions. Therefore, we shall compare the Schervish algorithm with our "scalar implementation" algorithm.

The "scalar implementation" algorithm deletes E_1 and E_2 from (2.8) and evaluates the n-term finite sum (i.e., $n = n_1 = n_2 = \ldots = n_{dim-1}$), our integration rule of (2.8), using scalar arithmetic to provide an estimator $\hat{P}_n$ of the desired probability P. Then $\hat{\alpha} = \hat{P}_n - \hat{P}_{n-2}$ is an estimator of the absolute error, and $\hat{\rho} = \hat{\alpha}/\hat{P}_n$ is an estimator of the relative error, of $\hat{P}_n$. This algorithm is designed to compute $\hat{P}_n$ for successively larger n and terminate when $\hat{\alpha}_n \leq \alpha$ and $\hat{\rho}_n \leq \rho$, where α and ρ are user supplied absolute and relative error bounds, respectively. For the base computing precision employed in the study, we use $10^{-15} < \alpha, \rho < 1$. Obviously, the termination criteria used here may sometimes result in an estimate $\hat{P}_n$ which, with respect to the true probability P, does not satisfy the desired error bounds α and ρ. This deficiency is present in most scalar algorithms. We will use the high quality self-validating answers obtained from interval analysis to see how frequently this termination methodology fails to result in a $\hat{P}_n$ that is within the prescribed error bounds α and ρ.

We emphasize the trivariate normal case in our comparisons because the computational burden in finding interval inclusions is not tremendous, and two and three dimensions most frequently arise in multivariate normal applications. For the trivariate case, thirty different positive definite correlation matrices having correlations in the range $0 < \rho \leq 0.9$ were selected at random and used in two runs of each of two experiments, one run for each of two error bounds. Each experiment included half of the correlation matrices. A third experiment was conducted to compare the algorithms for selected four variable cases and some exceptionally difficult trivariate integrals.

The first experiment has 525 integrals, 35 integrals for each of the 15 correlation matrices. The integrals are defined over cubic regions having unit length in every direction. The origin of each cube was selected at random within the domain [-5,5] in each dimension.

We set the absolute error bound $\alpha = 10^{-5}$ and relative error bound $\rho = 10^{-5}$ in the first run of experiment 1. In the second run $\alpha = 10^{-8}$ and $\rho = 10^{-8}$ were used. The results were compared with our "true value," obtained from self-validating computation of each integral. In all cases the width of the interval inclusion was no greater than 10^{-15}, so the "true value" used was necessarily very close to the theoretically correct probability. The estimated absolute error and estimated relative error for both Schervish's and the Scalar Algorithms were defined as

ABS = |computed value - "true value"|

and

REL = ABS/"true value."

Table 1 summarizes the result for the first run of Experiment one.

Table 2 gives the result from second run of Experiment one. Schervish's algorithm generally requires less time in execution, if we don't take account of those cases for which it does not terminate within 120 seconds. The mean relative error of the Schervish algorithm in both runs is very large. This algorithm appears to be rather unreliable for use over these small regions of integration. Inspection of the second and third lines of these tables shows that the termination criteria used by the scalar algorithm yields answers which almost always satisfied the given error bounds.

The second experiment involved fifteen trivariate integrations, one for each correlation matrix. We use this experiment to evaluate the performance for large regions of integration. Table 3 presents the regions of integration, correlation structures and the computed interval inclusion of the probabilities in this experiment. Table 5 and Table 6 show the computed probabilities, absolute errors, and relative errors when α and ρ are 10^{-5} for each algorithm. The Schervish algorithm failed to terminate for two of the fifteen integrals. Table 7 and Table 8 present the computed probabilities for the same data with α and ρ each given the value 10^{-8}. Again, Schervish's algorithm did not terminate for five of these integrations.

The third experiment included four four-dimensional integrals and eight three-dimensional problems featuring larger correlations. Table 4 gives the necessary description of those twelve integrals. Table 9 and Table 10 shows the results of both algorithms with α and ρ equal to 10^{-5}.

4. CONCLUSIONS

The Taylor expansion of the multivariate normal integral appears to provide a better scalar computational method than that used by Schervish in algorithm AS195 when accuracy is the basis for comparison. This is rather surprising in view of the simplicity of the Taylor expansion for four or fewer dimensions. Although computing time is greater for the Taylor expansion implementation than for Schervish's Method, the difference in execution time is not extremely large. Results for dimension greater than four were not reported. Limited experiments in higher dimensions have been conducted and, as expected, it was determined that very high accuracy was difficult to achieve. The Taylor expansion method is able to obtain satisfactorily accurate results, but only for large computing times.

Satisfactorily complete comparison of the two scalar algorithms was possible in this situation because a source of essentially correct answers was provided by interval analysis and automatic differentiation. Inclusion of multivariate probabilities for $n \leq 5$ is certainly reasonably possible using the methodology described in this paper. For more than five dimensions, enormous computing times are required to obtain desired inclusions. Using a single processor, cases of $n > 5$ should be avoided because they require too long in processing. However, the algorithm described can be operated under a parallel implementation which the authors believe has the potential to reduce processing time to an acceptable level when a large parallel processing configuration is employed. Work is underway to test such application. The method described in this paper is not readily extendable to interval inclusion of the multivariate normal CDF. The authors

have derived a method for obtaining interval inclusion of the bivariate normal CDF. A report of these results is currently being submitted for publication.

Table 1 Trivariate normal integration over unit cubes (α, $\rho = 10^{-5}$)

	Method	
	Schervish's Method[a]	Scalar Implementation
Answer between 0 and 1 obtained (% case)	98.92	100.00
ABS $\leq \alpha =$ .00001 satisfied (% cases)	100.00	99.35
REL $\leq \rho =$ .00001 satisfied (% case)	12.63	98.07
mean ABS value for all answers obtained	3.55×10^{-7}	6.43×10^{-6}
mean REL value for all answers obtained	7.19×10^{18}	1.21×10^{-5}
mean time (seconds) to compute answer	1.27 sec.	3.07 sec.

(a).There were 58 integrals for which Schervish's algorithm did not terminate within 120 seconds. Those integrals with missing value for answer are excluded from the Schervish method summary statistics.

Table 2 Trivariate normal integration over unit cubes (α, $\rho = 10^{-8}$)

	Method	
	Schervish's Method[a]	Scalar Implementation
Answer between 0 and 1 obtained (% case)	98.46	100.00
ABS $\leq \alpha =$ .00000001 satisfied (% cases)	100.00	100.00
REL $\leq \rho =$ .00000001 satisfied (% case)	43.85	100.00
mean ABS value for all answers obtained	2.79×10^{-9}	2.28×10^{-12}
mean REL value for all answers obtained	1.14×10^{18}	1.29×10^{-9}
mean time (seconds) to compute answer	4.92 sec.	8.04 sec.

(a).There were 69 integrals for which Schervish's algorithm did not terminate within 120 seconds. Those integrals with missing value for answer are excluded from the Schervish method summary statistics.

Table 3 Interval inclusion of integrals used for the second experiment

	Range		Correlation			Inclusion of Probability	
I.D.	A(I)	B(I)	ρ_{12}	ρ_{13}	ρ_{23}	Lowerbound	Upperbound
1	-6.0	2.0	0.9	0.9	0.9	0.96170067975686	0.96170067975689
2	-6.0	2.0	0.6	0.0	0.6	0.94278893709753	0.94278893709754
3	-6.0	2.0	0.0	0.0	0.9	0.94584202421904	0.94584202421905
4	-6.0	2.0	0.5	0.5	0.5	0.94253344853592	0.94253344853593
5	-6.0	2.0	0.5	0.0	0.5	0.94001583581975	0.94001583581976
6	-6.0	2.0	0.0	0.0	0.5	0.93674547946271	0.93674547946272
7	-6.0	2.0	0.1	0.1	0.1	0.93431490424365	0.93431490424366
8	-6.0	2.0	0.1	0.0	0.1	0.93397873246545	0.93397873246546
9	-6.0	2.0	0.0	0.0	0.1	0.93363670469361	0.93363670469362
10	-6.0	2.0	-0.3	-0.3	-0.3	0.93190899977300	0.93190899977301
11	-6.0	2.0	-0.3	0.0	-0.3	0.93237338842590	0.93237338842591
12	-6.0	2.0	0.0	0.0	-0.3	0.93283666254131	0.93283666254132
13	(a)	6.0	0.2	0.7	-0.4	0.22060958070880	0.22060958070881
14	(a)	6.0	0.3	0.5	0.7	0.28935499140859	0.28935499140860
15	(a)	6.0	0.1	0.4	0.9	0.27966079658526	0.27966079658559

(a). A(1)=-1.2 A(2)=0.5 A(3)=-1.0

Table 4 Interval inclusion of integrals used for the third experiment

	Range		Correlation			Inclusion of Probability	
I.D.	A(I)	B(I)	ρ_{12}	ρ_{13}	ρ_{23}	Lowerbound	Upperbound
1	-2.0	(a)	-0.99	0.99	-0.99	0.3413447446	0.3413447476
2	(b)	(c)	0.95	0.95	0.95	0.8423030713	0.8423030714
3	(d)	(e)	0.95	0.95	0.95	0.8423030713	0.8423030714
4	-2.0	2.0	0.95	0.95	0.95	0.9328452295	0.9328452296
5	(b)	(c)	0.95	0.90	0.99	0.8439839808	0.8439840328
6	-2.0	6.0	0.99	0.99	0.99	0.9725431790	0.9725438962
7	(f)	(g)	0.95	0.95	0.95	0.9410484396	0.9410484397
8	-2.0	6.0	-0.95	0.95	-0.95	0.9477740878	0.9477740879
9[h]	-1.96	1.96	0.50	0.00	0.50	0.8327171150	0.8327171669
10[i]	-6.0	2.0	0.10	0.10	0.10	0.9140338507	0.9140338593
11[j]	-2.0	2.0	0.10	0.20	0.40	0.8477754123	0.8477754965
12[k]	-2.0	2.0	0.70	0.70	0.70	0.8802218010	0.8802218577

(a).B(1)=0,B(2)=1,B(3)=2. (b).A(1)=-1.2,A(2)=-1.3,A(3)=-1.4. (c).B(1)=2,B(2)=3,B(3)=4.
(d).A(1)=-2,A(2)=-3,A(3)=-4. (e).B(1)=1.2,B(2)=1.3,B(3)=1.4. (f).A(1)=-2,A(2)=-2,A(3)=-6
(g).B(1)=2,B(2)=6,B(3)=2. (h). ρ_{14}=0.0,ρ_{24}=0.0,ρ_{34}=0.5. (i). ρ_{14}=0.1,ρ_{24}=0.1,ρ_{34}=0.1.
(j). ρ_{14}=0.3,ρ_{24}=0.5ρ_{34}=0.6. (k). ρ_{14}=0.7,ρ_{24}=0.7,ρ_{34}=0.7.

Table 5 Scalar implementation results for the second experiment (α, $\rho = 10^{-5}$)

I.D.	Probability	Absolute Error	Relative Error	Time (Sec.)
1	0.9616903379	0.000010	0.000010	53
2	0.9427889546	0.000001	0.000001	26
3	0.9458420291	0.000001	0.000001	28
4	0.9425334524	0.000001	0.000001	26
5	0.9400158165	0.000001	0.000001	27
6	0.9367455063	0.000001	0.000001	22
7	0.9343148825	0.000001	0.000001	21
8	0.9339787106	0.000001	0.000001	21
9	0.9336367282	0.000001	0.000001	22
10	0.9319089423	0.000001	0.000001	42
11	0.9323733142	0.000001	0.000001	34
12	0.9328366876	0.000001	0.000001	22
13	0.2206095784	0.000001	0.000001	16
14	0.2893549849	0.000001	0.000001	13
15	0.2796608132	0.000001	0.000001	13

Table 6 Schervish's algorithm result for the second experiment (α, $\rho = 10^{-5}$)

I.D.	Probability	Relative Error	Absolute Error	Time (Sec.)
1	0.9616994778	0.000001	0.000001	12
2	0.9427877439	0.000001	0.000001	13
3	(a)	(a)	(a)	(a)
4	0.9425321387	0.000001	0.000001	12
5	0.9400145399	0.000001	0.000001	10
6	0.9367437845	0.000001	0.000001	11
7	0.9343125193	0.000002	0.000002	09
8	0.9339763585	0.000002	0.000002	09
9	0.9336343452	0.000002	0.000002	09
10	0.9319080196	0.000001	0.000001	09
11	0.9323721476	0.000001	0.000001	09
12	0.9328339930	0.000003	0.000003	09
13	0.2206092694	0.000001	0.000001	09
14	0.2893538000	0.000001	0.000003	10
15	(a)	(a)	(a)	(a)

(a).Failed to terminate within two hours.

Table 7 Scalar implementation results for the second experiment (α, $\rho = 10^{-8}$)

I.D.	Probability	Absolute Error	Relative Error	Time (Sec.)
1	0.9617006780	0.000000001	0.000000001	586
2	0.9427889371	0.000000001	0.000000001	135
3	0.9458420241	0.000000001	0.000000001	119
4	0.9425334485	0.000000001	0.000000001	123
5	0.9400158358	0.000000001	0.000000001	124
6	0.9367454793	0.000000001	0.000000001	78
7	0.9343149042	0.000000001	0.000000001	93
8	0.9339787324	0.000000001	0.000000001	92
9	0.9336367046	0.000000001	0.000000001	77
10	0.9319089998	0.000000001	0.000000001	133
11	0.9323733883	0.000000001	0.000000001	122
12	0.9328366624	0.000000001	0.000000001	81
13	0.2206095808	0.000000001	0.000000001	47
14	0.2893549915	0.000000001	0.000000001	35
15	0.2796607965	0.000000001	0.000000001	36

Table 8 Schervish's algorithm result for the second experiment (α, $\rho = 10^{-8}$)

I.D.	Probability	Absolute Error	Relative Error	Time (Sec.)
1	0.9617006798	0.000000001	0.000000001	31
2	(a)	(a)	(a)	(a)
3	(a)	(a)	(a)	(a)
4	0.9425334498	0.000000001	0.000000001	42
5	0.9400158370	0.000000001	0.000000001	67
6	0.9367445481	0.000000001	0.000000001	69
7	0.9343149058	0.000000001	0.000000001	52
8	0.9339787340	0.000000001	0.000000001	53
9	0.9336367062	0.000000001	0.000000001	52
10	0.9319089972	0.000000001	0.000000001	56
11	0.9323733896	0.000000001	0.000000001	43
12	0.9328366637	0.000000001	0.000000001	55
13	(a)	(a)	(a)	(a)
14	(a)	(a)	(a)	(a)
15	(a)	(a)	(a)	(a)

(a).Failed to terminate within two hours.

Table 9 Scalar implementation results for the third experiment (α, ρ $=10^{-8}$)

I.D.	Probability	Absolute Error	Relative Error	Time (Min.)
1	0.3413447460	0.000001	0.000001	39
2	0.8423030714	0.000001	0.000001	15
3	0.8423030714	0.000001	0.000001	14
4	0.9328452295	0.000001	0.000001	17
5	0.8439840069	0.000001	0.000001	24
6	0.9725473658	0.000004	0.000004	45
7	0.9410484396	0.000001	0.000001	35
8	0.9477740879	0.000001	0.000001	68
9	0.8327168672	0.000001	0.000001	22
10	0.9140338584	0.000001	0.000001	19
11	0.8477754719	0.000001	0.000001	20
12	0.8802218037	0.000001	0.000001	36

Table 10 Schervish's algorithm result for the third experiment (α, ρ $= 10^{-8}$)

I.D.	Probability	Relative Error	Absolute Error	Time (Min.)
1	(a)	(a)	(a)	(a)
2	(a)	(a)	(a)	(a)
3	(a)	(a)	(a)	(a)
4	0.9328451942	0.000001	0.000001	02
5	(a)	(a)	(a)	(a)
6	(a)	(a)	(a)	(a)
7	(a)	(a)	(a)	(a)
8	(a)	(a)	(a)	(a)
9	0.8327170367	0.000001	0.000001	07
10	0.9140321608	0.000002	0.000002	09
11	0.8477749281	0.000001	0.000001	07
12	0.8802219221	0.000001	0.000001	20

(a).Failed to terminate within two hours.

5. ACKNOWLEDGEMENT

The authors wish to thank the referees for their careful and useful evaluation of the original manuscript. The final version is decidedly better due to their efforts.

6. REFERENCES

Alefeld, G. and Herzberger, J. (1983), Introduction to Interval Computation, New York: Academic Press.

Bohrer R. E. and Schervish, M. J. (1981), An Error-Bounded Algorithm for Normal Probabilities of Rectangular Regions, Technometrics, Vol. 23, No. 3, 297-300.

Corliss, G. F. (1988), Applications of Differentiation Arithmetic, in Reliability in Computing, edited by R. E. Moore, New York: Academic Press, 127-148.

Corliss, G. F. and Rall, L. B. (1987), Adaptive, Self-Validating Numerical Quadrature, SIAM, J. Sci. Stat. Comput. Vol. 8, No. 5, 831-847.

Genz, A. and Kahaner, D. K. (1986), The Numerical Evaluation of Certain Multivariate Normal Integrals, Journal of Computational and Applied Mathematics 16, 255-258.

Gupta, S. S. (1963), Probability Integrals of Multivariate Normal and Multivariate t, Ann. Math. Statist., 34, 792-838.

Jerrell, M. E. (1989), Automatic Differentiation Using Almost Any Language, SIGNUM Newsletter Vol. 24, No. 1, 2-9.

Johnson, L. W. and Riess, R. D. (1982), Numerical Analysis, Massachusetts: Addison-Wesley.

Lawson, C. L. (1988), Series Reversion as the Reversed Chain Rule, SIGNUM Newsletter, Vol. 23, No. 1, 7-9.

Milton, R. C. (1972), Computer Evaluation of Multivariate Normal Integral, Technometrics, Vol. 14, No. 4, 881-889.

Moore, R. E. (1979), Methods and Applications of Interval Analysis, SIAM

Studies in Applied Mathematics 2, SIAM, Philadelphia.

Plant, M. W. and Quandt, R. E. (1989), On the Accuracy and Cost of Numerical Integration in Several Variables, J. Statist. Comput. Simul. Vol. 32, No. 4, 229-248.

Rall, L. B. (1981), Automatic Differentiation: Techniques and Applications, Lecture Notes in Computer Science, Vol. 120, New York: Springer-Verlag.

Ratschek, H. and Rokne, J. (1984), Computer Methods for the Range of Functions, England: Ellis Horwood Limited.

Schervish, M. J. (1984), Multivariate Normal Probabilities with Error Bound, Algorithm AS 195, J. Royal Statistical Society, Appl. Statist., Vol. 33, No. 1, 81-94.

PART III

COMPARISON OF ALGORITHMS FOR BIVARIATE NORMAL CDF PROBABILITIES BASED ON SELF-VALIDATING RESULTS FROM INTERVAL ANALYSIS

Comparison of Algorithms for Bivariate Normal CDF Probabilities Based on Self-Validated Results from Interval Analysis

Morgan C. Wang and William J. Kennedy

Department of Statistics

Iowa State University

Ames, Iowa 50011

ABSTRACT

Comparison of algorithms for computing probabilities and percentiles is often carried out in an effort to identify the best algorithm for various applications. One requirement when conducting comparative studies is some useable source of "satisfactory approximations to correct answers" to use as a basis when making accuracy comparisons. This paper reports success in employing interval analysis with Owen's method (1956) to obtain a self-validating computational method for Bivariate Normal Probabilities. Results from applying this method can be used to provide a basis for accuracy studies of algorithms for Bivariate Normal Probabilities. A study to compare leading methods for evaluating the bivariate normal CDF probabilities, using the self-validated base values, was carried out. The paper reports a choice of better method.

Key Words: *interval analysis, Bivariate Normal Probabilities*

1. INTRODUCTION

There are two major objectives of this paper. The first objective is to describe a method for self-validating computation of Bivariate Normal CDF values. The second objective is to give the result of a comparative study of accuracy of competing methods for approximating this CDF.

Self-validating numerical computation can be achieved in several ways. We will employ interval analysis and implement rounded interval arithmetic in the computer to obtain self-validation. This means we will use directed roundings and proceed to compute an interval having floating-point endpoints that is guaranteed to contain the theoretically correct value of the desired probability. Then the midpoint of this computed interval is a scalar approximation, and the half width of this interval is a guaranteed error bound giving validity to this scalar approximation. We will arrange to obtain very short intervals so that the approximations obtained provide essentially correct answers to use as a basis for comparing the accuracy of outputs from competing "scalar" algorithms.

Basic elements of interval analysis and a complete list of references can be found in Kennedy (1990). Implementation of interval arithmetic, and computation of interval inclusions for various functions are described in Wang and Kennedy (1990).

In the next section, we will describe a way for obtaining interval inclusion of the Bivariate Normal CDF using the scalar algorithm given by Owen (1956). Then results of a comparative study of three competing scalar algorithms for approximating these probabilities will be given in the third section. Finally, we will make recommendations of best method in Section 4.

2. SELF-VALIDATING EVALUATING OF BIVARIATE NORMAL INTEGRALS BASED ON INTERVAL ANALYSIS

Given a random vector $\underset{\sim}{z} = (z_1, z_2)'$ having the Bivariate Normal Distribution with mean vector $\underset{\sim}{0}$, unit variance, and correlation ρ, the probability P of z_1 less than h and z_2 less than k can be expressed as

$$P(h,k;\rho) = \int_{-\infty}^{h}\int_{-\infty}^{k} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp-\left[\frac{z_1^2+z_2^2-2\rho z_1 z_2}{2(1-\rho^2)}\right] dz_1 dz_2. \qquad (2.1)$$

One way to construct an interval inclusion of (2.1) is to use the formula given by D. B. Owen (1956). This formula is

$$P(h,k;\rho) = \tfrac{1}{2}g(h) + \tfrac{1}{2}g(k) - f\big(h,a(h,k)\big) - f\big(k,a(k,h)\big) - \tfrac{1}{2}b(h,k) \qquad (2.2)$$

where $$g(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt \qquad (2.3)$$

$$a(x,y) = (x-y\rho)/y\sqrt{1-\rho^2} \qquad (2.4)$$

$$b(x,y) = \begin{cases} 1 & x < 0 \text{ and } y \geq 0, \text{ or } x \geq 0 \text{ and } y < 0. \\ 0 & \text{otherwise} \end{cases} \qquad (2.5)$$

$$f(x,y) = \text{sign}(y) \int_{0}^{|y|} \frac{\exp[-\tfrac{1}{2}x^2(1+t^2)]}{2\pi(1+t^2)}\, dt \qquad (2.6)$$

We wish to find an interval inclusion of (2.2), i.e., we wish to obtain an interval inclusion of g(x), a(x,y), b(x,y) and f(x,y), and combine these using interval arithmetic. Since we know that both a(x,y) and b(x,y) are rational functions the interval inclusion of these functions is easily obtained. Interval inclusion of g(x) can be obtained using a Taylor series and automatic differentiation arithmetic as described by Kennedy (1990). A method for computing an interval inclusion of f(x,y) will be developed in the

remainder of this section.

Before starting our discussion, some properties of f(x,y), which are given by Owen (1956), will be reviewed. Also, some notation relative to interval analysis will be given. The properties P1-P6, and notational elements N1-N5 are as follows.

(P1). $f(x,y) = -f(x,-y)$, all $x \in \mathbb{R}$ and $y \in \mathbb{R}$.

(P2). $f(x,0) = 0$.

(P3). $f(x,y) = \frac{1}{2}g(x) + \frac{1}{2}g(xy) - g(x)g(xy) - f(xy,1/y)$, if $x \in \mathbb{R}$ and $y \geq 1$.

(P4). $$f(x,y) = \frac{1}{2\pi}\left[\mathrm{Tan}^{-1}y - \sum_{j=0}^{\infty} c_j\, y^{2j+1}\right]$$

where $$c_j = (-1)^j \frac{1}{2j+1}\left[j - \exp(\frac{-x^2}{2}) \sum_{i=0}^{j} \frac{x^{2i}}{2^i i!}\right] \quad (\text{if } 0 < y \leq 1)$$

(P5). $f(x,y)$ increases as y increases for any fixed $x \in \mathbb{R}$.

(P6). $|f(x,y)|$ decreases as x increases for any fixed $y \neq 0$.

(N1). ∇ and Δ denote the downward and upward roundings from real numbers to floating-point numbers, respectively.

(N2). $X = [\underline{X}, \overline{X}]$ denotes the closed real interval having upper bound $\overline{X}$ and lower bound $\underline{X}$. $X = [x,x]$ is the real interval of the corresponding real number x.

(N3). $XL = [\underline{X}, \underline{X}]$, i.e., XL coincides with X whenever $\underline{X} = \overline{X}$.

(N4). $XU = [\overline{X}, \overline{X}]$.

(N5). $F(X,Y)$ denotes an interval inclusion of $f(x,y)$.

Now, we define $d_k = \sum_{j=0}^{k} c_j y^{2j+1}$ and $f_k(x,y) = \frac{1}{2\pi}(\tan^{-1}y - d_k)$ where the c_j are coefficients given in (P4). Notice that $c_j y^{2j+1}$ satisfies the

following three conditions: (1) $(|c_j y^{2j+1}|)$ is a monotone decreasing sequence for all $y \in (0,1]$. (2) $(c_j y^{2j+1})$ is an alternating sequence. (3) $\lim_{j\to\infty} |c_j y^{2j+1}| = 0$, for all $y \in (0,1]$. Therefore, using the alternating series test, the limit of d_k, denoted by d_∞ exists. Moreover, d_∞ is such that, $d_1 < d_3 < \ldots < d_\infty < \ldots < d_2 < d_0$. Consequently the limit of $f_k(x,y)$, denoted by $f_\infty(x,y)$, exists and equals $f(x,y)$. Furthermore we have that $f_0(x,y) < f_2(x,y) < \ldots < f(x,y) < \ldots < f_3(x,y) < f_1(x,y)$. Therefore, we see that

$$f(x,y) \in [f_{2k}(x,y),\ f_{2k+1}(x,y)] \text{ for all } k = 0,1,2,\ldots. \qquad (2.7)$$

Notice that the results given above are true only when $y \in (0,1]$. Thus, we will need to employ a transformation using formulas (P1) and (P3) whenever y is outside this range. For example, $f(x,-2)$ is equivalent to $-\frac{1}{2}g(x) - \frac{1}{2}g(2x) + g(x)g(2x) + f(2x,\frac{1}{2})$ by P1 and P3. Therefore we can obtain $f(x,-2)$ using $f(2x,\frac{1}{2})$ in the expression given above.

The interval inclusion of a(h,k) in (2.2) can be a nondegenerate interval. Therefore we must extend our discussion to include an interval-valued function f with interval-valued arguments. Fortunately, using properties (P4) and (P5), we have the following relationships for the interval of values of f over intervals $X = [\underline{X},\overline{X}]$ and $Y = [\underline{Y},\overline{Y}]$.

$$\begin{cases} f(X,Y) \subseteq [f(\overline{X},\underline{Y})\ ,\ f(\underline{X},\overline{Y})] & \text{if } \underline{Y} \geq 0 \\ f(X,Y) \subseteq [f(\underline{X},\underline{Y})\ ,\ f(\overline{X},\overline{Y})] & \text{if } \overline{Y} \leq 0 \\ f(X,Y) \subseteq [f(\underline{X},\underline{Y})\ ,\ f(\underline{X},\overline{Y})] & \text{if } \underline{Y} < 0 < \overline{Y}\ . \end{cases} \qquad (2.8)$$

The relationships in (2.8) show that indeed we need only deal with the scalar-valued function f having two real-valued arguments to obtain an interval inclusion of f(X,Y).

Finally, we note that the function $f_k(x,y)$ is a linear combination of

monotone functions and rational functions. Thus, for each k the interval inclusion of $f_k(x,y)$ can be obtained very easily. Therefore, an interval inclusion of f(x,y), denoted by F(X,Y), can be formed using the following formula.

$$\begin{cases} F(X,Y) \subseteq [\underline{F(XU,YL)} \ , \ \overline{F(XL,YU)}] & \text{if } \underline{Y} \geq 0 \\ F(X,Y) \subseteq [\underline{F(XL,YL)} \ , \ \overline{F(XU,YU)}] & \text{if } \overline{Y} \leq 0 \\ F(X,Y) \subseteq [\underline{F(XL,YL)} \ , \ \overline{F(XL,YU)}] & \text{if } \overline{Y} > 0 > \underline{Y} \end{cases} \tag{2.8}$$

Consequently, the interval inclusion IP = $[\underline{P},\overline{P}]$ of (2.1) could be obtained by correctly using interval arithmetic and interval inclusion of g(x), f(x,y), a(x,y) and b(x,y).

3. COMPARISON OF ALGORITHMS FOR APPROXIMATING THE BIVARIATE NORMAL CDF

Experience gained using the interval inclusion algorithm described in the preceding section suggested that a good algorithm using only scalar (not interval) arithmetic might result if the interval inclusion algorithm were simply coded for scalar arithmetic. The number of terms to carry in the series expansion had to somehow be determined. Fortunately, the size of successive intervals $[f_{2j}(x,y), f_{2j+1}(x,y)]$ in (2.7) decreases with increasing j. Therefore the midpoints, denoted by $\hat{f}_j(x,y)$, form a sequence of approximations to $f(x,y)$. In practice, the expansion was continued until the interval length stabilized at a sufficiently small value. This algorithm will be called the "scalar implementation."

Terza and Welland (1988) report results of a comparative study of accuracy of several algorithms for bivariate normal CDF approximation. Two of the best algorithms identified in this study are the IMSL routine DBNRDF and the Divgi (1979) algorithms. These two algorithms were compared with the "scalar implementation" to determine which generally provides the most accurate results. The basis for accuracy comparisons were self-validated values obtained using the interval inclusion algorithm.

Three experiments were constructed to test the performance of these scalar algorithms for different ρ values. In all of the integrals evaluated in each of these experiments, the width of the interval inclusion of computed probability was not larger than 10^{-15}, so the "true value" middle point of computed interval inclusion was necessarily very close to the theoretically correct probability. Therefore, the "true value" formed the basis for our precision study.

The first experiment involved one thousand integrals, one hundred integrals for each ρ which was randomly selected between -0.9 and 0.9. The corner point (h,k) for each integral, where h,k were defined in (2.1), was selected at random within the region $[-5,5]\times[-5,5]$. The estimated absolute error and estimated relative error for these scalar algorithms were computed according to

$$\text{A.E.} = |\text{computed value} - \text{"true value"}|$$

and

$$\text{R.E.} = \text{A.E.}/\text{"true value"}.$$

The results of this experiment are presented in Table 1. Clearly, both the "scalar implementation" and Divgi's Algorithm perform equally well and there is no reason to prefer one over the other.

The second experiment used three hundred integrals, fifty integrals for each $\rho \in (0.95, 0,99, 9.999, -0.95, -0.99, -0.999)$. The corner point for each integral in this experiment was chosen at random from the domain $[-3,3]\times[-3,3]$. Table 2 summarizes the results of this experiment. Obviously, Divgi's Algorithm is preferable because it requires the least amount of computing time and provides outstanding accuracy.

The third experiment included twelve integrals with large correlation. Table 3 gives the necessary description of those integrals and the computed interval inclusions. Table 4, Table 5, and Table 6 shows the results of three scalar algorithms. For these integrals there is little difference in accuracy of the three algorithms. Divgi's method continued to require less computing time.

4. CONCLUSIONS

Interval and differentiation arithmetic were used to compute interval inclusions of desired probabilities. The lengths of computed intervals were made sufficiently small so that probabilities guaranteed to be correct essentially to machine precision were obtained. The utility of these arithmetics in this kind of application is obvious. However, due to substantially greater computing time, production of self-validated results using interval computations is probably not a replacement for traditional scalar computing since sufficiently accurate scalar algorithms are usually obtainable.

Results of the study of accuracy of competing scalar algorithms indicate that either Divgi's algorithm or our scalar implementation perform very well. The scalar implementation generally requires more machine time, so Divgi's algorithm appears to be the best choice.

Table 1 Means and standard deviations of absolute and relative errors from first experiment

	Scalar Version	Divgi's Method	IMSL
Mean A.E.	4.55E-17	8.03E-17	8.16E-09
Std. A.E.	0.00000 [a]	0.000000 [a]	3.61E-08
Mean R.E.	4.98E-08	8.17E-08	3.26E-02
Std. R.E.	6.60E-7	1.70E-05	1.53E-01
Average Run Time (Sec.)	0.06700	0.04800	0.07000

(a) Standard Deviations are very small.

Table 2 Means and standard deviations of absolute and relative errors from second experiment

	Scalar Version	Divgi's Method	IMSL
Mean A.E.	4.86E-17	3.22E-17	7.13E-09
Std. A.E.	0.00000 [a]	0.00000 [a]	4.09E-08
Mean R.E.	1.84E-09	4.88E-10	3.42E-02
Std. R.E.	1.97E-08	4.43E-09	2.27E-01
Average Run Time (Sec.)	0.18300	0.05000	0.05300

(a). Standard Deviations are Very Small.

Table 3 Interval inclusion of integrals used for the third experiment

I.D.	HH	KK	RHO	Inclusion of Probability	
				Lowerbound	Upperbound
1	0.0000	0.0000	-0.9999	0.0022508095471555	0.0022508095476480
2	0.1000	0.0000	0.9999	0.4999999999999994	0.4999999999999995
3	0.1250	0.0000	0.9999	0.4999999999999999	0.5000000000000000
4	4.0000	0.0000	-0.9999	0.4999683287581668	0.4999683287581669
5	0.0000	4.0000	-0.9999	0.4999683287581668	0.4999683287581669
6	8.0000	8.0000	0.9999	0.9999999999999993	0.9999999999999994
7	7.0000	9.0000	-0.9999	0.9999999999987201	0.9999999999987202
8	-3.875	7.6250	-0.9999	0.0000533123497388	0.0000533123497389
9	-5.000	5.0000	0.9999	0.0000002866515718	0.0000002866515719
10	-0.0125	-0.00675	-0.9999	0.0002252159041540	0.0002252159041541
11	-2.500	-3.7500	0.9999	0.0000884172852008	0.0000884172852009
12	5.0000	-5.0000	0.9999	0.0000002866515718	0.0000002866515719

Table 4 Sclalr implementation results for the third experiment

I.D.	Probability	Absolute Error	Relative Error
1	0.0022508095474046	3.500E-15	1.555E-12
2	0.4999999999999994	1.000E-16	2.000E-16
3	0.500000000000000	1.000E-16	2.000E-16
4	0.4999683287581669	1.000E-16	2.000E-16
5	0.4999683287581669	1.000E-16	2.000E-16
6	0.9999999999999993	1.000E-16	1.000E-16
7	0.9999999999987201	1.000E-16	1.000E-16
8	0.0000533123497389	1.000E-16	1.876E-12
9	0.0000002866515719	1.000E-16	3.496E-10
10	0.0002252159041540	1.000E-16	4.444E-13
11	0.0000884172852008	1.000E-16	1.131E-12
12	0.0000002866515719	1.000E-16	3.497E-10

Table 5 Divgi's implementation results for the third experiment

I.D.	Probability	Absolute Error	Relative Error
1	0.0022508095474046	3.500E-15	1.555E-12
2	0.4999999999999994	1.000E-16	2.000E-16
3	0.500000000000000	1.000E-16	2.000E-16
4	0.4999683287581669	1.000E-16	2.000E-16
5	0.4999683287581669	1.000E-16	2.000E-16
6	0.9999999999999993	1.000E-16	1.000E-16
7	0.9999999999987201	1.000E-16	1.000E-16
8	0.0000533123497389	1.000E-16	1.876E-12
9	0.0000002866515719	1.000E-16	3.496E-10
10	0.0002252159041538	2.000E-16	8.888E-13
11	0.0000884172852008	1.000E-16	1.131E-12
12	0.0000002866515719	1.000E-16	3.497E-10

Table 6 Results of IMSL's subroutine DBNRDF for the third experiment

I.D.	Probability	Absolute Error	Relative Error
1	0.0022508095474056	4.000E-15	1.777E-12
2	0.4999999999999995	1.000E-16	2.000E-16
3	0.4999999999999995	1.000E-15	2.000E-15
4	0.4999683287581674	5.000E-16	1.000E-15
5	0.4999683287581674	5.000E-16	1.000E-15
6	0.9999999999999993	1.000E-16	1.000E-16
7	0.9999999999987211	1.050E-15	1.060E-15
8	0.0000533123497389	1.000E-16	1.876E-12
9	0.0000002866515709	1.000E-15	3.496E-09
10	0.0002252159041551	1.100E-15	4.943E-12
11	0.0000884172852008	1.000E-16	1.131E-12
12	0.0000002866515709	9.000E-16	3.119E-09

5. REFERENCES

Divgi, D. R. (1979), Calculation of Univariate and Bivariate Normal Probability Functions, Annals of Statistics 7, 903-910.

IMSL (1987), IMSL STAT/LIBRARY User's Manual, Houston, IMSL Inc., 917-918.

Kennedy, W. J. (1990), Statistical Computing Feature: Special Purpose Numerical Tools for Approximating Functions, Statistical Computing and Statistical Graphics Newsletter. V.1, 3-6.

Owen, D. B. (1956), Tables for Computing Bivariate Normal Probabilities, Annals of Mathematical Statistics, V.27, 1075-1090.

Terza, J. V. and Welland, W. (1988), Accuracy and Efficiency Comparisons of Bivariate Normal Approximation Algorithms, Working Paper 12-87-2, Dept. of Economics, Pennsylvania State University.

Wang, M. and Kennedy, W. J. (1990), Comparison of Algorithms for Bivariate Normal Probability Over a Rectangle Based on Self-Validated Results from Interval Analysis, to appear in Journal of Statistical Computation and Simulation.

PART IV.

SELF-VALIDATING COMPUTATIONS OF PROBABILITIES FOR SELECTED CENTRAL AND NONCENTRAL UNIVARIATE PROBABILITY FUNCTIONS

Self-validating computations of probabilities for selected central and non-central univariate probability functions

Morgan C. Wang and William J. Kennedy

Department of Statistics

Iowa State University

Ames, IA 50011

ABSTRACT

Self-validating computational methods provide a scalar approximation to the desired value and a guaranteed error bound. Such methods are especially useful whenever computed results must satisfy given accuracy requirements. This paper gives methods for obtaining self-validating results when computing probabilities and percentiles of univariate continuous distributions. Probability functions dealt with explicitly in the paper are Normal, Incomplete Gamma, Incomplete Beta, and Non-central Chi-Square. Self-validation is achieved through use of interval arithmetic computations.

Key Words: *Interval analysis, probability functions, automatic differentiation, continued fractions, and self-validation.*

1. INTRODUCTION

Self-validating numerical computation is sometimes called automatic error analysis, and it can be achieved in more than one way. We will use interval analysis to accomplish self-validation of probabilities and percentiles. This means that we compute an interval which is guaranteed to contain the theoretically correct value of the desired probability or percentile. Then the midpoint of the interval is the scalar approximation and the half-width of this interval is a guaranteed error bound giving validity to the scalar approximation. Since we strive to obtain intervals having very small width, the approximation obtained is guaranteed to be within a given short distance from the unknown true value.

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 evaluating a new algorithm, a reliable source of essentially true values is needed. Usually existing tables do not provide sufficiently accurate entries, or cover a sufficiently large region of the variable and parameter space, to be satisfactory for this application. Rust and Voit (1990) mention this lack of basis for comparing accuracy. 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, of course, depend in part on the level of accuracy of the computed probability. Self-validation of intermediate and final result will be useful in this case.

Computations involving intervals do, at first sight, seem to be a

complicated and inconvenient process. In fact, this is not the case given today's computer hardware which includes standard floating-point support. A few simple subprograms give interval analysis capabilities.

Successful interval computing algorithms yield intervals having very small width. Numerical difficulties during the course of computation, such as cancellation and rounding, result in wide intervals. Although the interval obtained is guaranteed to contain the true unknown answer, the interval may be too wide to be useful. Thus "good" algorithms are needed both in scalar and in interval computations. It is not necessarily the case that a good scalar computing algorithm will perform well when implemented using interval computations, and the converse is also true. In fact, some of the interval-based methods recommended in this paper are not among the best scalar computing methods. In the remainder of this section we will give definitions and describe procedures which form the basis for development of self-validating computing for probabilities and percentiles.

Basic elements of interval analysis are given by Moore (1979), Alefeld and Herzberger (1983), Ratschek and Rokne (1984), and Corliss and Rall (1987). By an interval we mean a closed bounded set of real numbers $A = \{x: \underline{A} \le x \le \overline{A}\} = [\underline{A}, \overline{A}]$. The set of all real compact intervals is denoted by $I\mathbb{R}$ and its elements are expressed by upper case letters.

Let $A = [\underline{A}, \overline{A}]$ and $B = [\underline{B}, \overline{B}]$ be any two elements in $I\mathbb{R}$, then the interval arithmetic operations are defined as

$$AoB = \{aob: a \in A \text{ and } b \in B\} \tag{1.1}$$

where o is one of the arithmetic operators +, -, *, and / ; and A/B is defined only when $0 \notin B$. Algebraic properties of interval arithmetic are given

by Moore (1979) and Ratschek and Rokne (1984).

Suppose f is a real valued function of n real variables $x_1, x_2, \ldots, x_n$, defined on real intervals $X_1, X_2, \ldots, X_n$, respectively. The <u>united extension</u> of the function f over $X_1, X_2, \ldots, X_n$ is

$$\bar{f}(X_1,X_2,\ldots,X_n) = \bigcup_{(x_1,x_2,\ldots,x_n)\in(X_1,X_2,\ldots,X_n)} \{f(x_1,x_2,\ldots,x_n)\}. \quad (1.2)$$

An <u>interval extension</u> of f is an interval valued function F of n interval valued variables such that $F(x_1,x_2,\ldots,x_n) = f(x_1,x_2,\ldots,x_n)$ for all real arguments. In other words when all of the arguments of F are degenerate intervals (i.e., $A = [\underline{A},\bar{A}]$ such that $\underline{A} = \bar{A}$), F coincides with f. One special case of an interval extension is the <u>natural interval extension</u> which can be obtained by substituting interval operations for the corresponding scalar operations. An interval valued function F of $X_1, X_2, \ldots, X_n$ is <u>inclusion monotonic</u> if for any subsets $Y_1, Y_2, \ldots, Y_n$ of $X_1, X_2, \ldots, x_n$, respectively, $F(Y_1,Y_2,\ldots,Y_n) \subseteq F(X_1,X_2,\ldots,X_n)$. We will be interested in finding inclusion monotonic interval extensions of certain real valued functions. These will simply be called <u>interval inclusions</u>. Generally, for a given function, there are many interval inclusions.

The following four results from the theory of interval analysis are well known and will be given without proof (c.f., Moore (1979), pp 20-24). These results will be used later to ensure inclusion monotonicity of the natural interval extension of a function.

(R_1) The united extension of a real valued function is inclusion monotonic.

(R_2) An interval arithmetic function as given by (1.1) is the united extension of the associated real valued function, i.e., it is inclusion monotonic.

(R_3) A rational interval function, i.e., a function whose interval values are defined by a specific finite sequence of interval arithmetic operations, is inclusion monotonic.

(R_4) The natural interval extension of a continuous monotone function, such as the exponential function, coincides with its united extension, i.e., it is inclusion monotonic.

Implementation of interval operations on a digital computer must be done very carefully if the properties of real interval analysis are to remain valid when the floating-point system is used to obtain interval inclusions. A recommended procedure is to employ directed rounding. By this we mean that rounding is toward $-\infty$ when computing the lower endpoint of the interval and toward $+\infty$ for the upper endpoint. Therefore, the computed intervals obtained in rounded interval computing will have floating-point endpoints and contain the theoretically correct real intervals at every stage of the computations so long as there is not underflow or overflow in floating expression.

The software for rounded interval arithmetic used for this research was developed on an IBM compatible personal computer containing an INTEL 80287 NPX. This software includes basic rounded interval arithmetic operations and some utility routines such as the interval exponential function and interval logarithm function.

In the next section, general methodology for self-validating approximation of the value of univariate integrals and the inverse of univariate integrals will be discussed. An application of this methodology to selected univariate cumulative distribution functions will be given in Section 3.

2. SELF-VALIDATING EVALUATING OF UNIVARIATE INTEGRALS AND INVERSES OF UNIVARIATE INTEGRALS

First, we consider the problem of finding an interval inclusion Jf of the integral $P = \int_a^b f(x)dx$ where f(x) is a Riemann integrable function and x is a single real valued variable. Two cases are considered. The first case involves a Taylor series expansion of f(x) whenever both a and b are finite numbers, and the second case employs a continued fraction to handle the situation when either |a| or |b| is infinite.

To describe the first case, we need to assume that this integral can be expressed as an infinite series of the form

$$P = \sum_{i=0}^{\infty} w_i g_i(x_i) \tag{2.1}$$

$$= r_n + e_n$$

where $$r_n = \sum_{i=0}^{n} w_i g_i(x_i) \tag{2.2}$$

and $$e_n = \sum_{n+1}^{\infty} w_i g_i(x_i) \tag{2.3}$$

denote the integration rule and error term, respectively. For integrals considered in the next section we will be able to define and computationally deal with a series expansion.

Let $X_i = [\nabla x_i, \Delta x_i]$, $W_i = [\nabla w_i, \Delta w_i]$, and $G_i(X_i)$ be an interval inclusion of $g_i(x_i)$ for $i = 0,1,\ldots,n$, where ∇ and Δ stand for downward and upward directed rounding, respectively. Directed roundings insure that the true values x_i and w_i will be in the intervals X_i and W_i which have floating-point endpoints and are used when computing. The natural interval extension

$$Ir_n = \sum_{i=0}^{n} W_i G_i(X_i) \tag{2.4}$$

of integration rule (2.2), obtained by substituting rounded interval operations for corresponding scalar operations, and replacing $g_i(x_i)$ and w_i by $G_i(X_i)$ and W_i, is inclusion monotonic. This is to say that Ir_n is an interval inclusion of the integration rule r_n.

An interval inclusion Ie_n of the error term e_n of (2.3) depends heavily on the behavior of the tail of this infinite series. Therefore, the way to construct an interval inclusion of the error term can vary from one integral to another. Thus, we leave the details of this discussion for the next section, but we assume for now that Ie_n can be obtained.

Let $If_n = Ir_n + Ie_n$ be the sum of the interval inclusions of the terms r_n and e_n for given n. Then we have that

$$P = \int_a^b f(x)dx \in If_n, \text{ for each } n. \tag{2.5}$$

An application of (2.5) for increasing values of n can be made. The intersection of the interval inclusions If_n for increasing n will yield a sequence of approximations

$$Jf_0 = If_0$$

$$Jf_n = Jf_{n-1} \cap If_n, \quad n = 1,2,\ldots, . \tag{2.6}$$

Computations will terminate when the successive intervals Jf_n cease to decrease in width. The resulting interval is self-validating for P.

The second case, when the interval of integration does not have finite length, is based on use of continued fractions. A continued fraction is given by two sequences of real numbers $(a_n; n\geq 1)$ and $(b_n; n\geq 0)$, and can be written as

$$v = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \ldots}}}$$

or more compactly as $v = b_0 + \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \ldots$. The mth convergent of a continued fraction which converges is defined as $v_m = b_0 + \frac{a_1}{b_1+} \frac{a_2}{b_2+} \ldots \frac{a_m}{b_m}$ for $m \geq 0$. The continued fraction is said to converge to a finite value v if $\mathrm{Lim}_{m \to \infty} v_m = v < \infty$. One way to evaluate a convergent continued fraction is to use the forward recurrence formula (Kennedy and Gentle 1980) which is defined through the following equations

$$\begin{aligned} c_m &= b_m c_{m-1} + a_m c_{m-2} \\ d_m &= b_m d_{m-1} + a_m d_{m-2} , \quad m = 1,2,\ldots, \end{aligned} \tag{2.7}$$

with $c_{-1} = 1$, $d_{-1} = 0$, $c_0 = b_0$, $d_0 = 1$ and the mth convergent is given by $v_m = c_m/d_m$.

The following result (R_5) was stated and proved by Dudley (1987), and will be given without proof. This result will be used to obtain an interval inclusion of certain continued fractions.

(R_5) For a continued fraction $b_0 + \frac{a_1}{b_1+} \frac{a_2}{b_2+} \ldots$, such that $a_m \geq 0$ and $b_m \geq 0$ for all $m \geq 1$, we have

$$v_0 \leq v_2 \leq \ldots \leq v \leq \ldots \leq v_3 \leq v_1 \tag{2.8}$$

when it converges to $v < \infty$; if it does not converge, (2.8) remains true without the stipulation "$\leq v \leq$."

Now, let $Ia_n = [\nabla a_n, \Delta a_n]$, $Ib_n = [\nabla b_n, \Delta b_n]$, and define an interval

extension of the forward recurrence formulas

$$Ic_m = Ib_m \cdot Ic_{m-1} + Ia_m \cdot Ic_{m-2}$$
$$Id_m = Ib_m \cdot Id_{m-1} + Ia_m \cdot Id_{m-2}, \quad m = 1,2,\ldots, \tag{2.9}$$

with $Ic_{-1} = [1,1]$, $Id_{-1} = [0,0]$, $Ic_0 = [Vb_0, \Delta b_0]$, $Id_0 = [1,1]$, and an interval inclusion of the mth convergent $Iv_m = Ic_m/Id_m = [IVL_m, IVU_m]$. Applying the result (R_5), we have $IVL_0 \leq IVL_2 \leq \ldots \leq v \leq \ldots \leq IVU_3 \leq IVU_1$. Suppose we now define $JV_m = [IVL_{2m}, IVU_{2m+1}]$, $m = 0,1,2,\ldots$. Clearly, (JV_m) is a nested sequence. Therefore width of successive intervals can be used as a basis for termination of computations. The resulting interval is self-validating for v.

These methods can be used, with suitable extension, to compute the probabilities of higher dimensional probability functions such as the bivariate normal (Wang and Kennedy (1990)) and multivariate normal probability functions ((Wang and Kennedy (to appear)).

Next, we consider the problem of finding the yth percentile of the distribution of a continuous random variable at a given probability P, i.e., finding the root of the equation $g(y) = P(x \leq y) - P = 0$. We will use the interval Newton-Raphson iteration given by Moore (1979) to obtain an interval inclusion of this root. We will begin with an appropriate initial interval, which can be arbitrarily wide, to start the iteration. Convergence of the interval-based iteration is guaranteed whereas in the scalar iteration convergence depends on the starting value. In other words, the initial value problem of the traditional scalar iteration method does not exist for the interval version iteration. The interval method is denoted as (R_6) and is

stated as follows.

(R_6) If an interval $X^{(0)}$ contains a root y of $g(y) = P(x \leq y) - P = 0$, then so does $X^{(k)}$ for all $k = 0,1,2,\ldots$, defined by

$$X^{(k+1)} = X^{(k)} \cap N(X^{(k)})$$

where $$N(X^{(k)}) = m(X^{(k)}) - g(m(X^{(k)}))/G'(X^{(k)}),$$

$m(X^{(k)})$ is the midpoint of $X^{(k)}$,

and G' is an interval inclusion of the probability density function.

Furthermore, the intervals $X^{(k)}$ form a nested sequence converging to y if $0 \notin G'(X^{(0)})$. The computation can be terminated when successive interval widths become acceptably small. The resulting interval $X^{(k)}$ is self-validating for y.

3. APPLICATIONS TO SELECTED UNIVARIATE CUMMLATIVE DISTRIBUTION FUNCTIONS

3.1 Univariate normal distribution

Let

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{3.1}$$

$$= \int_{-\infty}^{x} \phi(t) dt$$

denote the CDF of the normal distribution. One method for self-validating computation of $\Phi(x)$ utilizes the continued fraction (Wall 1948),

$$\int_{x}^{\infty} \phi(t) dt = \phi(x)\left(\frac{1}{x+} \frac{1}{x+} \frac{2}{x+} \frac{3}{x+} \ldots\right), \text{ for } x > 0. \tag{3.2}$$

The conditions of (R_5) are satisfied for the fraction, and an interval inclusion of the monotone function ϕ is easily obtained. Thus an interval inclusion of $\Phi(x)$ is easily computed.

Although (3.2) works very well when $x \geq 1.5$, the speed of convergence is relatively slow for smaller x. Therefore, we need to employ a Taylor series expansion over the range $0 < x < 1.5$. Consider the slightly more general problem of finding an interval inclusion IP of the probability

$$P = \int_{a}^{b} \phi(t) dt \tag{3.3}$$

over a finite interval [a,b], with $a > 0$.

The derivatives of $\phi(t)$ exist to all orders in the interval $X = [a,b]$. Thus, we can expand ϕ at the midpoint $c = (a+b)/2$ of X. Then, for any even number n and some ϵ between x and c, we have

$$\phi(x) = \sum_{i=0}^{\infty} (\phi(c))_i (x-c)^i$$

$$= \sum_{i=0}^{n-1} (\phi(c))_i (x-c)^i + (\phi(\epsilon))_n (x-c)^n$$

$$\in \sum_{i=0}^{n-1} (\phi(c))_i (x-c)^i + (F(X))_n (x-c)^n \qquad (3.4)$$

where $(F(X))_n$ is an interval inclusion of $(\phi(\epsilon))_n$ and $(\phi(c))_i = \frac{\phi^{(i)}(c)}{i!}$, $i = 0,1,2,\ldots,n-1$. Let H be the interval hull $(\nabla(c-a), \nabla(c-b), \Delta(c-a), \Delta(c-b))$, $C = [\nabla c, \Delta c]$, and $(F(C))_i$ be an interval inclusion of $(\phi(c))_i$, $i = 0,1,2,\ldots,n-1$. Integrating both sides of (3.4) with respect to x, we obtain

$$P = \int_a^b \phi(t)dt \in \sum_{i=0}^{n-1} (\phi(c))_i \frac{(x-c)^{i+1}}{i+1} \Bigg|_a^b + (F(X))_n \frac{(x-c)^{n+1}}{n+1} \Bigg|_a^b$$

$$\subseteq IP_n = Ir_n + Ie_n \qquad (3.5)$$

where $$Ir_n = 2 \sum_{\substack{i=0 \\ i \text{ even}}}^{n-2} (F(C))_i \frac{H^{i+1}}{(i+1)} \qquad (3.6)$$

and $$Ie_n = 2 (F(X))_n \frac{H^{n+1}}{n+1} \qquad (3.7)$$

denote the integration rule, and error term, respectively.

We can obtain the desired interval inclusion IP, if an interval inclusion of the Taylor coefficients in (3.5) can be easily and efficiently computed. Fortunately, there is a numerical tool called "automatic differentiation" or "differentiation arithmetic" which employs recursion to give successively higher order derivatives at a point c. Detailed

descriptions of automatic differentiation can be found in Moore (1979), Rall (1981), Corliss (1988), Lawson (1988), Jerrell (1989), and Kennedy (1990). Implementation of automatic differentiation using interval arithmetic provides the means for computing interval inclusion of the Taylor coefficients in (3.5) at the point c or over the intervals C and X. We will not give details of this computation because they have been provided in the literature cited.

Now, we can define

$$JP_n = IP_n \cap IP_{n-1}, \quad n = 2,3,\ldots$$
$$ABS_n = 0.5*W(JP_n) = 0.5(\overline{JP}_n - \underline{JP}_n), \quad n = 2,3,\ldots$$
$$REL_n = ABS_n/m(JP_n) = 2ABS_n/(\overline{JP}_n + \underline{JP}_n), \quad n = 2,3,\ldots \qquad (3.8)$$

where ABS_n and REL_n denote selected absolute error and relative error measures associated with JP_n, respectively. This iteration can be terminated whenever either JP_n ceases to reduce in width or when both ABS_n and REL_n satisfy prespecified absolute and relative error tolerances. The resulting interval is self-validating for P in (3.5). Table 1 gives some examples of interval inclusions obtained using this method.

The ability to compute very short interval inclusions of normal probabilities provides the basis for obtaining interval inclusion of normal percentiles. Let x_p denote the unknown percentile for given probability p. We will proceed to find an interval inclusion for the root of the equation $g(x) = \Phi(x) - p = 0$. Since $g'(x) = \phi(x)$, an interval inclusion $G'(x)$ of $g'(x)$ is easily obtained for any finite x. If $X^{(0)}$ is any interval containing x_p, we have that $0 \notin G'(X^{(0)})$. The interval Newton-Raphson procedure, described in the previous section can be applied and is guaranteed

to find an interval inclusion of x_p to within any specified tolerance limited only by the precision of the computer floating-point system used in the implementation. Selection of the initial interval $X^{(0)}$ can, for example, be made using the Hastings (1955) approximation. Table 2 gives some interval inclusions obtained using this method in a personal computer.

3.2 Gamma distribution

The cumulative probability function of a gamma random variable can be expressed as

$$g(\alpha,x) = \frac{1}{\Gamma(\alpha)} \int_0^x e^{-t} \cdot t^{\alpha-1} dt \ , \tag{3.9}$$

where $\alpha > 0$, $x > 0$, and $\Gamma(\alpha)$ is the complete gamma function. Instead of evaluating (3.9) directly, we will first consider computing an interval inclusion of the complement of this cumulative probability function,

$$g^c(\alpha,x) = \frac{1}{\Gamma(\alpha)} \int_x^\infty e^{-t} \cdot t^{\alpha-1} dt \ . \tag{3.10}$$

When α is an integer, $g^c(\alpha,x)$ can be expressed as

$$g^c(\alpha,x) = \left(\sum_{i=0}^{\alpha-1} \frac{x^i}{i!} \right) \cdot e^{-x} \tag{3.11}$$

which is the product of a rational function $\Sigma_{i=0}^{\alpha-1} (x^i/i!)$ and a monotone function e^{-x}, so an interval inclusion of (3.11) can be obtained using (R_3) and (R_4).

For a non-integer $\alpha = b+n$ where n is the integer part of α and $0 < b < 1$, $g^c(\alpha,x)$ can be expressed as

$$g^c(\alpha,x) = g^c(b,x) + \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} \sum_{i=0}^{n-1} \frac{x^i}{\prod_{j=0}^{i} (b+j)}$$

$$- \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} \left(\frac{1}{x+} \frac{1-b}{x+} \frac{1}{1+} \frac{2-b}{2+} \frac{2}{1+} \frac{3-b}{x+} \cdots \right)$$

$$+ \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} \sum_{i=0}^{n-1} \frac{x^i}{\prod_{j=0}^{i} (b+j)}$$

$$- \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} [fr1(b,x)+fr2(b,n,x)]. \qquad (3.12)$$

In (3.12), fr2(b,n,x) is a rational function and fr1(b,x) is a convergent continued fraction (William and Thron 1980), which satisfies the conditions in (R_5). An interval inclusion of the complete gamma function was developed by the authors as a part of the software support library. Using this inclusion, we can proceed to employ interval arithmetic in an obvious way to obtain an interval inclusion of (3.12). Consequently, we can obtain the interval inclusion of (3.10) and (3.9).

Although the computational method based on (3.12) works very well for computing an interval inclusion of $g^c(\alpha,x)$, the substraction operation in $1 - g^c(\alpha,x)$ can produce serious cancellation when $g^c(\alpha,x)$ is very close to one. The result of this is that the interval inclusion of $g(\alpha,x)$ is too wide to be very useful. To overcome this problem we implement the Taylor series expansion method whenever cancellation is detected. When there is no singularity in the integral the interval Taylor series expansion method works very well. Unfortunately, we need to overcome the singularity problem which occurs when $\alpha < 1$. To do this we make the following modification. The function $g(\alpha,x)$ in (3.9) can be expressed as

$$- \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} \left(\frac{1}{x+} \frac{1-b}{x+} \frac{1}{1+} \frac{2-b}{2+} \frac{2}{1+} \frac{3-b}{x+} \cdots \right)$$

$$+ \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} \sum_{i=0}^{n-1} \frac{x^i}{\prod_{j=0}^{i} (b+j)}$$

$$- \frac{e^{-x} \cdot x^{b-1}}{\Gamma(b)} [fr1(b,x)+fr2(b,n,x)]. \qquad (3.12)$$

In (3.12), fr2(b,n,x) is a rational function and fr1(b,x) is a convergent continued fraction (William and Thron 1980), which satisfies the conditions in (R_5). An interval inclusion of the complete gamma function was developed by the authors as a part of the software support library. Using this inclusion, we can proceed to employ interval arithmetic in an obvious way to obtain an interval inclusion of (3.12). Consequently, we can obtain the interval inclusion of (3.10) and (3.9).

Although the computational method based on (3.12) works very well for computing an interval inclusion of $g^c(\alpha,x)$, the substraction operation in $1 - g^c(\alpha,x)$ can produce serious cancellation when $g^c(\alpha,x)$ is very close to one. The result of this is that the interval inclusion of $g(\alpha,x)$ is too wide to be very useful. To overcome this problem we implement the Taylor series expansion method whenever cancellation is detected. When there is no singularity in the integral the interval Taylor series expansion method works very well. Unfortunately, we need to overcome the singularity problem which occurs when $\alpha < 1$. To do this we make the following modification. The function $g(\alpha,x)$ in (3.9) can be expressed as

$$g(\alpha,x) = \frac{1}{\Gamma(\alpha)} \int_0^{x_0} e^{-t} \cdot t^{\alpha-1} dt + \frac{1}{\Gamma(\alpha)} \int_{x_0}^{x} e^{-t} t^{\alpha-1} dt$$

$$= g(\alpha,x_0) + \frac{1}{\Gamma(\alpha)} \int_{x_0}^{x} e^{-t} t^{\alpha-1} dt \ . \qquad (3.13)$$

An interval inclusion of $(1/\Gamma(\alpha)) \int_0^{x_0} e^{-t}\, t^{\alpha-1}\, dt$ can be obtained using the Taylor series expansion as was described in Subsection 3.1. The first term in the right member of (3.13), namely $g(\alpha,x_0)$, is equivalent to

$$\frac{1}{\Gamma(\alpha)} \int_0^{x_0} \left[\sum_{i=0}^{\infty} \frac{(-1)^i t^i}{i!} \right] t^{\alpha-1} dt$$

$$= \frac{1}{\Gamma(\alpha)} \sum_{i=0}^{\infty} (-1)^i \frac{t^{\alpha+i}}{i!(\alpha+i)} \Bigg|_0^{x_0}$$

$$= \frac{x_0^{\alpha}}{\Gamma(\alpha)} \sum_{i=0}^{\infty} (-1)^i \frac{x_0^i}{i!(\alpha+i)} \ . \qquad (3.14)$$

Let $c_i = (-1)^i [x^i/i!(\alpha+i)]$ and $g_k(\alpha,x_0) = [x_0^{\alpha}/\Gamma(\alpha)] \sum_{i=0}^{k} c_i$. Now c_i satisfies the following three conditions: (1) $\{c_i\}$ is an alternating sequence, (2) $\{|c_i|\}$ is monotone decreasing, and (3) $\lim_{i\to\infty} |c_i| = 0$. Therefore, using the alternating series test, the limit of $g_k(\alpha,x_0)$ is $g(\alpha,x_0)$. Moreover, $g(\alpha,x)$ is such that

$$g_1(\alpha,x_0) < g_3(\alpha,x_0) < \ldots < g(\alpha,x_0) < \ldots < g_2(\alpha,x_0) < g_0(\alpha,x_0).$$

Therefore, we see that

$$g(\alpha,x_0) \in [g_{2k+1}(\alpha,x_0), g_{2k}(\alpha,x_0)] \text{ for } k = 0,1,2,\ldots \ . \qquad (3.15)$$

Since $\sum_{i=0}^{k} c_i$ is a rational function for each finite k, the interval inclusion of $\sum_{i=0}^{k} c_i$ is ensured by (R_3). Consequently, an interval inclusion of $g_{2k+1}(\alpha,x_0)$ and $g_{2k}(\alpha,x_0)$ denoted by $G_{2k+1}(\alpha,x_0)$ and $G_{2k}(\alpha,x_0)$,

respectively, can be obtained. Finally, $g(\alpha,x_0) \in [\underline{G_{2k+1}(\alpha,x_0)}\ ,\ \overline{G_{2k}(\alpha,x_0)}]$ for all $k = 0,1,2,\ldots$. Consequently, the interval inclusion of (3.13) can be obtained. The methods described above have been shown in practice to perform entirely satisfactorily. Some example interval inclusions are given in Table 3.

To find interval inclusion of percentiles of the gamma distribution we can employ the same methodology given previously for percentiles of the univariate normal. Example interval inclusions of percentiles are given in Table 4.

3.3 Central and non-central chi-square distributions

The CDF of a Chi-Square random variable having ν degrees of freedom is

$$\chi^2(y,\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^y e^{-t/2} t^{\nu/2-1} dt. \tag{3.16}$$

Using the well known relationship between this function and the Incomplete Gamma function we have

$$g(\nu/2,y/2) = \frac{1}{\Gamma(\nu/2)} \int_0^{y/2} e^{-t} t^{\nu/2-1} dt. \tag{3.17}$$

Therefore interval inclusions of chi-square probabilities and percentiles are readily obtained using the methods described previously. Table 5 shows some results specific to this distribution.

Now, we extend the discussion to the non-central chi-square distribution. The cumulative distribution function of a non-central chi-square random variable with ν degree of freedom and non-centrality parameter λ can be expressed as

$$h(z,\lambda,\nu) = \sum_{j=0}^{\infty} \left[\frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!} \frac{1}{2^{\frac{\nu}{2}+j} \Gamma(\frac{\nu}{2}+j)} \int_0^z t^{\frac{\nu}{2}+j-1} e^{-t/2} dt \right] \tag{3.18}$$

$$- \sum_{j=0}^{\infty} \left[\frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!} g(\tfrac{\nu}{2}+j, z/2) \right]$$

which is a weighted sum of central chi-square probabilities with weights equal to the probabilities in a Poisson distribution with expected value $\lambda/2$.

The formula (3.18) can be rewritten as $h(z,\lambda,\nu) = r_k + e_k$

where $$r_k = \sum_{j=0}^{k-1} \frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!} \cdot g(\tfrac{\nu}{2}+j, \tfrac{z}{2}) \tag{3.19}$$

and $$e_k = \sum_{j=k}^{\infty} \frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!} \cdot g(\tfrac{\nu}{2}+j, \tfrac{z}{2}) \tag{3.20}$$

which has the form of integration rule and error term, respectively.

Well known facts in this case are (1) $g(\nu/2+j,z/2)$ decreases as j increases (Tricomi 1950), and (2) $\Sigma_{j=k}^{\infty} e^{-\lambda/2}(\lambda/2)^j/j! = 1/\Gamma(k) \int_0^{\lambda/2} e^{-t} t^{k-1} dt = g(k,\lambda/2)$ (Johnson 1959). Therefore, we have the following relationship

$$0 \le e_k = \sum_{j=k}^{\infty} \frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!} \cdot g(\tfrac{\nu}{2}+j, \tfrac{z}{2})$$

$$\le g(\tfrac{\nu}{2}+k, \tfrac{z}{2}) \cdot \sum_{j=k}^{\infty} \frac{e^{-\lambda/2} \cdot (\frac{\lambda}{2})^j}{j!}$$

$$= g(\tfrac{\nu}{2}+k, \tfrac{z}{2}) \cdot g(k, \tfrac{\lambda}{2}) \ . \tag{3.21}$$

Thus, an interval inclusion of the error term (3.20) of (3.18) is assured because we can compute an inclusion of the incomplete gamma function. An interval inclusion of the finite linear combination (3.19) is assured by

methods previously described. Consequently, we have an interval inclusion of (3.18). This method has proved to give excellent interval inclusions over a wide range of the variable and parameter involved.

Again, interval inclusion of percentiles of a non-central random variable can be obtained in a manner similar to that used for the univariate normal distribution. Selected results are shown in Tables 6, 7, and 8.

3.4 Beta distribution

The Incomplete Beta function with parameters $\alpha>0$ and $\beta>0$ is defined by

$$I(\alpha,\beta,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt \ , \tag{3.22}$$

where $\Gamma(\cdot)$ is the complete gamma function.

We can obtain an interval inclusion of (3.22) by simply using the Taylor series expansion method which was described in Section 3.3. However, to improve the computation speed, we make the following modification. Instead of expanding $t^{\alpha-1}(1-t)^{\beta-1}$, we only expand $(1-t)^{\beta-1}$. For the case $0 < x \leq 0.5$ and $0 < \beta \leq 1$, equation (3.22) can be expressed as

$$I(\alpha,\beta,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x^\alpha}{\alpha} + \sum_{i=1}^{\infty} \frac{\prod_{a=1}^{i} (j-\beta)}{i!} \frac{x^{\alpha+i}}{(\alpha+i)} \right)$$

$$= r_n + e_n$$

where

$$r_n = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x^\alpha}{\alpha} + \sum_{i=1}^{n-1} \frac{\prod_{j=1}^{i} (j-\beta)}{i!} \frac{x^{\alpha+i}}{\alpha+i} \right) \tag{3.23}$$

and
$$e_n = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{i=n}^{\infty} \frac{\prod_{j=1}^{i} (j-\beta)}{i!} \frac{x^{\alpha+i}}{(\alpha+i)} \tag{3.24}$$

denote an integration-type rule, and error term, respectively. The expression in (3.23) is a rational expression, so an interval inclusion is easily obtained. The error term (3.24) satisfies

$$0 \le e_n \le \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{x^{\alpha+n}}{(\alpha+n)} \frac{1}{(1-x)} . \tag{3.25}$$

Therefore, an interval inclusion of e_n is $[0,\overline{I}e_n]$ where $\overline{I}e_n$ is the upper endpoint of an interval inclusion of $\left(\Gamma(\alpha+\beta)x^{\alpha+n}\right)/\left(\Gamma(\alpha)\Gamma(\beta)(\alpha+n)(1-x)\right)$. Consequently, we have the interval inclusion of $I(\alpha,\beta,x)$ for each n. Since the width of the interval inclusion decreases as n increases, we can monitor the width and cease iteration at a desirable point.

For the case, $0 < x \le 0.5$ and $\beta > 1$, we decompose β and express $I(\alpha,\beta,x)$ in the form

$$I(\alpha,\beta,x) = I(\alpha,r,x) + \sum_{i=1}^{s} \frac{x^{\alpha}(1-x)^{\beta-i}\Gamma(\alpha+\beta-i)}{\Gamma(\alpha)\Gamma(\beta+1-i)} \tag{3.26}$$

where r is the fractional part of β and s is the integer part. An interval inclusion of (3.26) results from the facts that $\Sigma_{i=1}^{s}\left(x^{\alpha}(1-x)^{\beta-i}\Gamma(\alpha+\beta-i)\right)/\left(\Gamma(\alpha)\Gamma(\beta+1-i)\right)$ is a rational expression, and $I(\alpha,r,x)$ has $0 < r < 1$.

When $x > 0.5$, we may use the well known relationship $I(\alpha,\beta,x) = 1 - I(\beta,\alpha,1-x)$. However, if $I(\beta,\alpha,1-x)$ is close to unity, cancellation will become a problem resulting in unacceptably wide interval inclusion. When this occurs we employ the Taylor series method applied to the integrand of $I(\alpha,\beta,x)$ in (3.22).

Interval inclusion of percentiles in this case are again computed using

the interval version of the Newton-Raphson method. To speed up this iteration, we use the scalar approximation suggested by Majumder and Bhattacharjee (1964) to find an initial interval $X^{(0)}$ having small width.

Tables 9 and 10 show the results of computing inclusions of percentiles using input intervals previously computed as inclusions of indicated probabilities.

4. CONCLUSIONS

Interval analysis and automatic differentiation include numerical tools which can profitably be applied to obtain self-validating approximations to probabilities and percentiles. In situations where accuracy guaranteed to within given bounds is needed, these are among the few available numerical tools.

The methods suggested for use in this paper have been extensively tested on the functions considered. Excellent results were obtained over very large regions of the variable and parameter space in every case. When a failure occurred, an excessively large interval normally resulted and 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.

Table 1 Interval inclusion of selected normal probabilities

I.D.	XL	XU	Inclusion of Probability	
			Lowerbound	Upperbound
1	- ∞	-35.000	0.112491070647240d-267	0.112491070647241d-267
2	- ∞	-12.000	0.177648211207767d-32	0.177648211207768d-32
3	- ∞	-5.0000	0.286651571879193d-6	0.286651571879194d-6
4	- ∞	-4.0000	0.316712418331199d-4	0.316712418331200d-4
5	- ∞	-3.0000	0.114420683102269d-2	0.114420683102270d-2
6	- ∞	-2.0000	0.227501319481792d-1	0.227501319481793d-1
7	- ∞	-1.0000	0.158655253931457d0	0.158655253931458d0
8	- ∞	1.0000	0.841344746068542d0	0.841344746068544d0
9	- ∞	2.0000	0.977249868051820d0	0.977249868051821d0
10	- ∞	3.0000	0.998650101968369d0	0.998650101968371d0
11	- ∞	4.0000	0.999968328758166d0	0.999968328758167d0
12	- ∞	4.4500	0.999995706485529d0	0.999995706485531d0
13	12.0000	12.5000	0.177274954777880d-32	0.177274954777881d-32
14	-1.0000	-0.99999	0.241971934371664d-5	0.241971934371665d-5
15	-2.5000	1.64000	0.9432877512001200	0.943287751200121d0

Table 2 Interval inclusion of selected normal percentiles

I.D.	P-values	Inclusion of Percentile	
		Lowerbound	Upperbound
1	0.991	0.2365618126864291d1	0.2365618126864301d1
2	0.001	-0.3090232306167811d1	-0.3090232306167821d1
3	0.286651571879193d-6	-0.5000000000000001d1	-0.4999999999999991d1
4	0.316712418331199d-4	-0.4000000000000001d1	-0.3999999999999991d1
5	0.114420683102269d-2	-0.3000000000000001d1	-0.2999999999999991d1
6	0.227501319481792d-1	-0.2000000000000001d1	-0.1999999999999991d1
7	0.158655253931457d0	-0.1000000000000001d1	-0.9999999999999991d0
8	0.841344746068542d0	0.1000000000000001d1	0.1000000000000011d0
9	0.977249868051820d0	0.2000000000000001d1	0.2000000000000011d1
10	0.998650101968369d0	0.3000000000000001d1	0.3000000000000011d1
11	0.999968328758166d0	0.4000000000000001d1	0.4000000000000011d1
12	0.999995706485529d0	0.4449999999999991d1	0.4450000000000011d1

Table 3 Interval inclusion of selected gamma probabilities

I.D.	x	α	Inclusion of Probability	
			Lowerbound	Upperbound
1	0.0001	0.0001	0.9991370418689930d0	0.999137041868994d0
2	1.0000	50.500	0.173150041448188d-65	0.173150041448189d-65
3	4.9900	130.50	0.121024456548534d-131	0.121024456548536d-131
4	5.1000	1.0000	0.9939032534344840d0	0.993903253434485d0
5	15.100	10.000	0.356013376959444d0	0.356013376959445d0
6	135.10	110.00	0.9881925296857490d0	0.988192529685751d0
7	15.100	9.5000	0.9506923635975500d0	0.950692363597551d0
8	126.55	100.50	0.9925626958862300d0	0.992562695886231d0
9	5.1000	0.9900	0.9940467696603010d0	0.994046769660302d0
10	4.5000	10.000	0.1709273289937850d-1	0.170927328993786d-1
11	87.750	102.50	0.6681167091523150d-1	0.668116709152317d-1
12	85.500	100.00	0.6772902784653810d-1	0.677290278465383d-1

Table 4 Interval inclusion of selected gamma percentiles

I.D.	%	α	Inclusion of Percentiles	
			Lowerbound	Upperbound
1	99.000	101.00	0.1258386528836050d3	0.125838652883606d3
2	99.000	0.9000	0.4372270680091970d1	0.437227068009198d1
3	99.900	136.50	0.1754699564308500d3	0.175469956430851d3
4	99.900	0.1000	0.3363677011718730d1	0.336367701171875d1
5	99.990	11.000	0.2776229438785260d2	0.277622943878527d2
6	99.990	181.50	0.2359284506386780d3	0.235928450638679d3
7	90.000	0.3000	0.8848107733602430d0	0.884810773360245d0
8	90.000	41.000	0.4939016466028120d2	0.493901646602813d2
9	0.0010	11.000	0.2040810821899770d1	0.204081082189978d1
10	0.0100	46.500	0.2529434555583180d2	0.252943455558319d2
11	0.1000	1.0000	0.1000500333583530d-2	0.100050033358354d-2
12	0.1000	0.3000	0.6972699096783340d-10	0.697269909678335d-10

Table 5 Interval inclusion of selected chi-square probabilities

I.D.	x	υ	Inclusion of Probability Lowerbound	Upperbound
1	50.500	1.0000	0.999999999998808d0	0.999999999998809d0
2	0.5000	16.000	0.303127472288458d-9	0.303127472288460d-9
3	10.500	21.000	0.283383551233326d-1	0.283383551233327d-1
4	15.500	36.000	0.113265810760853d-2	0.113265810760854d-2
5	25.500	51.000	0.107320450769665d-2	0.107320450769666d-2
6	5.5000	66.000	0.252067679434142d-23	0.252067679434143d-23
7	50.500	76.000	0.106285755383325d-1	0.106285755383327d-1
8*	45.500	101.00	0.363040096521666d-6	0.363040096571562d-6
9	45.500	101.00	0.363040096553920d-6	0.363040096553922d-6
10*	50.500	121.00	0.192892701215020d-8	0.192892708269584d-8
11	50.500	121.00	0.192892704396170d-8	0.192892704396171d-8
12*	50.500	156.00	0.304118709382183d-16	0.354534110402760d-16
13	50.500	156.00	0.333062956747127d-16	0.333062956747142d-16
14	30.500	6.0000	0.999968426157000d0	0.999968426157001d0
15	0.5000	156.00	0.755288523755819d-162	0.755288523755820d-162

(*). Cancellation increases the interval size.

Table 6 Interval inclusion of selected noncentral chi-square probabilities

I.D.	x	υ	λ	Inclusion of Probability Lowerbound	Upperbound
1	0.00393	1.0000	6.0000	0.249846372425803d-2	0.249846372425805d-2
2	9.23636	5.0000	1.0000	0.8272971875117554d0	0.8272971875117555d0
3	24.72497	11.000	21.000	0.2539481822183l2d0	0.2539481822183l3d0
4	44.98534	31.000	6.0000	0.812519878506496d0	0.812519878506498d0
5	38.56038	51.000	1.0000	0.851949736185911d-1	0.851949736185913d-1
6	82.35814	100.00	16.000	0.118434882274782d-1	0.118434882274783d-1
7	331.78852	300.00	16.000	0.735595671030670d0	0.735595671030672d0
8	459.92612	500.00	21.000	0.279702360080005d-1	0.279702360080007d-1
9	0.00016	1.0000	1.0000	0.612142892988142d-2	0.612142892988143d-2
10	0.00393	1.0000	1.0000	0.303381422975377d-1	0.303381422975379d-1
11	(a)	1.0000	1.0000	0.612142892988142d-2	0.303381422975379d-1
12	(a)	1.0000	(b)	0.828445318038297d-3	0.304174999263582d-1

(a). x=[0.00016 , 0.00393] (b). λ = [1 , 6].

Table 7 Interval inclusion of selected non-central chi-square densities

I.D.	x	υ	λ	Inclusion of Density Lowerbound	Upperbound
1	0.00612	1.0000	1.0000	0.309304159557992d1	0.309304159557993d1
2	0.11483	3.0000	16.000	0.571906677085898d-4	0.571906677085899d-4
3	0.55430	5.0000	21.000	0.316969716194547d-5	0.316969716194549d-5
4	24.724970	11.000	1.0000	0.576930663314747d-2	0.576930663314748d-2
5	15.655460	31.000	11.000	0.291947650837170d-3	0.291947650837171d-3
6	77.385960	51.000	21.000	0.252695349938571d-1	0.252695349938572d-1
7	70.06489	100.00	21.000	0.658785129946774d-4	0.658785129946775d-4
8	331.78852	300.00	6.0000	0.888682666567772d-2	0.888682666567774d-2
9	429.38754	500.00	21.000	0.181489687585154d-3	0.181489687585155d-3
10	459.92612	500.00	21.000	0.211916059304653d-2	0.211916059304655d-2
11	540.93031	500.00	21.000	0.975821289846907d-2	0.975821289846909d-2
12	576.49281	500.00	21.000	0.290406258695947d-2	0.290406258695949d-2

Table 8 Interval inclusion of selected non-central chi-square percentiles

I.D.	%	υ	λ	Inclusion of Percentile Lowerbound	Upperbound
1	1.00000	11.000	10.000	0.673260712284854d1	0.673260712284857d1
2	5.00000	11.000	10.000	0.972921029106716d1	0.972921029106721d1
3	10.00000	11.000	10.000	0.116293773040816d2	0.116293773040816d2
4	90.00000	11.000	10.000	0.315133389312785d2	0.315133389312786d2
5	95.00000	11.000	10.000	0.352916032402770d2	0.352916032402771d2
6	99.00000	11.000	10.000	0.430420059386981d2	0.430420059386982d2
7	1.00000	31.000	25.000	0.303344757534414d2	0.303344757534415d2
8	5.00000	31.000	25.000	0.366569594327401d2	0.366569594327402d2
9	10.00000	31.000	25.000	0.403407783130309d2	0.403407783130310d2
10	90.00000	31.000	25.000	0.727834344489928d2	0.727834344489929d2
11	95.00000	31.000	25.000	0.783231522558848d2	0.783231522558849d2
12	99.00000	31.000	25.000	0.893603958338555d2	0.893603958338556d2

Table 9 Interval inclusion of selected beta probabilities

I.D.	x	α	β	Inclusion of Probability Lowerbound	Upperbound
1	0.050	0.1000	0.1000	0.377508482874951	0.377508482874953
2	0.200	0.3000	0.9000	0.593638436633204	0.593638436633206
3	0.950	0.9000	0.7000	0.886741184991061	0.886741184991062
4	0.050	1.2500	81.250	0.974466810319941	0.974466810319942
5	0.050	101.25	101.25	0.292566253276879d-74	0.292566253276880d-74
6	0.350	1.2500	21.250	0.999802030898906	0.999802030898907
7	0.950	41.250	1.2500	0.174271224125545	0.174271224125546
8	0.950	101.25	21.250	0.999999528446403	0.999999528446405
9	0.050	9.0000	9.0000	0.328714227224861d-7	0.328714227224862d-7
10	0.050	7.0000	3.0000	0.257187500000000d-7	0.257187500000001d-7
11	0.050	.99999	1.00001	0.500019851861818d-1	0.500019851861819d-1
12	0.050	1.0000	1.00001	0.500004872861717d-1	0.500004872861718d-1

Table 10 Interval inclusion of selected beta percentiles

I.D.	Input Probability Lowerbound	Upperbound	Computed Percentile Lowerbound	Upperbound
1	0.377508482874951	0.377508482874953	0.04999999999	0.05000000001
2	0.593638436633204	0.593638436633206	0.19999999999	0.20000000001
3	0.886741184991061	0.886741184991062	0.94999999999	0.95000000001
4	0.974466810319941	0.974466810319942	0.04999999999	0.05000000001
5	0.292566253276d-74	0.292566253277d-74	0.04999999999	0.05000000001
6	0.999802030898906	0.999802030898907	0.34999999999	0.35000000001
7	0.174271224125545	0.174271224125546	0.94999999999	0.95000000001
8	0.999999528446403	0.999999528446405	0.94999999999	0.95000000001
9	0.32871422722488d-7	0.32871422722487d-7	0.04999999999	0.05000000001
10	0.25718750000000d-7	0.25718750000001d-7	0.04999999999	0.05000000001
11	0.050001985186181 8	0.050001985186181 9	0.04999999999	0.05000000001
12	0.050004872861717	0.050004872861718	0.04999999999	0.05000000001

5. REFERENCES

Alefeld, G. and Herzberger, J. (1983), Introduction to Interval Computations, New York: Academic Press.

Corliss, G. F. (1988), "Application of Differentiation Arithmetic", Reliability in Computing: the role of interval methods in scientific computing, New York: Academic Press.

Corliss, G. F. and Rall, L. B. (1987), "Adaptive, Self-Validating Numerical Quadrature," SIAM Journal on Scientific and Statistical Computing, 8, 831-847.

Dudley, R. M. (1987), "Some Inequalities for Continued Fractions," Mathematics of Computation, 49, 585-593.

Hastings, C. J. (1955), Approximations for Digital Computers, Princeton: Princeton University Press.

Jerrel, M. E. (1989), "Automatic Differentiation Using Almost Any Language," SIGNUM Newsletter, 24, 2-9.

Johnson, N. L. (1959), "On an extension of the connection between Poisson and χ^2 distributions," Biometrika, 46, 352-363.

Kennedy, W. J. (1990), "Special Purpose Numerical Tools for Approximating Functions," Statistical Computing and Statistical Graphics Newsletter: American Statistical Association, 1, 3-6.

Kennedy, W. J. and Gentle, J. E. (1980), Statistical Computing, New York and Basel: Marcel Dekker.

Lawson, C. L. (1988), "Series Reversion as the Reversed Chain Rule," SIGNUM Newsletter, 23, 7-9.

Majumder, K. L. and Bhattacharjee, G. P. (1964), "The Incomplete Beta

Integral," Algorithm AS63, 117-120.

Moore, R. E. (1979), Methods and Applications of Interval Analysis, Philadelphia: SIAM.

Rall, L. B. (1981), "Automatic Differentiation: techniques and applications," Lecture Notes in Computer Science, 120, New York: Springer-Verlag.

Ratschek, H. and Rokne, J. (1984), Computer Methods for the Range of Functions, England: Ellis Horwood Ltd.

Rust, Philip F. and Voit, Eberhard O. (1990), "Statistical Densities, Cumulatives, Quantiles, and Power obtained by S-System Differential Equations," Journal of the American Statistical Association, 85, 572-578.

Tricomi, F. G. (1950), "Sulla funzione gamma incomplete," Ann. Math. Pura Appl., 31, 263-279.

Wall, H. S. (1948), Analytic Theory of Continued Fractions, New York: Van Nostrand.

Wang, M. and Kennedy, W. J. (1990), "Comparison of Algorithms for Bivariate Normal Probability over a Rectangle Based on Self-Validating Results from Interval Analysis," Journal of Statistical Computation and Simulation, 37, 13-25.

Wang, M. and Kennedy, W. J. (to appear), A Numerical Method for Accurately Approximating Multivariate Normal Probabilities, Computational Statistics and Data Analysis.

William, B. J. and Thron, W. J. (1980), Continued Fractions: Analytic Theory and Applications, London: Allison-Wesley.

GENERAL SUMMARY

Self-validating numerical integration methods based on interval analysis, automatic differentiation, continued fractions, and Taylor series method already successfully applied in the computation of probabilities and percentiles for selected distribution functions.

The methods suggested for use in this dissertation have been extensively tested on the distribution functions considered. Excellent results were obtained over a very large regions of the variables and parameters space in every case. The methods may fail when floating-point underflow or overflow occurs. However this leads to an unusually wide resulting interval. And an unusually wide resulting interval can be served as an automatic error detector.

Self-validating numerical methods are not only needed in the computations of probabilities and percentiles. Moreover, these methods studied in this dissertation can not solve the problem of self-validating numerical integration over an infinite range in general. Therefore, further research is necessary.

We believe that there exist at least two different ways to develop the general method for solving self-validating numerical integration over an infinite range. The first way is to use S-systems which were originally developed for analysis of organizationally complex systems such as cellular and molecular networks in biology. The basic idea of S-systems can be found in Irvine and Savageau (1990). Some applications of S-systems in statistical computing can be found in Rust and Voit (1990). Another way is to find an efficient method to automatically transferorm an infinite series to a

corresponding continued fraction which has some desired properties.

In general, numerical tools such as automatic differentiation and continued fractions are very useful in statistical computing. For example, the information matrix of a maximum likelihood estimator can be obtained easily by applying the automatic differentiation technique. Therefore, these basic automatic differentiation functions should be included in the standard statistical computing environments such as IMSL and S-PLUS.

Finally, we want to say that this is a fruitful area in statistical computing and continuing research in this area is underway.

REFERENCES

Alberth, O. (1988), Precise Numerical Analysis, Dubuque: Wm. C. Brown Publishers.

Alefeld, G. and Herzberger, J. (1983), Introduction to Interval Computation, New York: Academic Press.

Bowman, K. O. and Shenton, L. R. (1989), Continued Fractions in Statistical Applications, New York and Basel: Marcel Dekker.

Bohrer, R. E. and Schervish, M. J. (1981), An Error-Bounded Algorithm for Normal Probabilities of Rectangular Regions, Technometrics, Vol. 23, No. 3 297-300.

Bouver, H. and Bargmann, R. E. (1979), Comparison of Computational Algorithms for the Evaluation of the Univariate and Bivariate Normal Distribution, Proceedings of the Computer Science and Statistics: 12th Annual Symposium on the Interface, 344-348.

Corliss, G. F. (1987), Performance of Self-validating Adaptive Quadrature, in Numerical Integration, edited by P. Keast and G. Fairweather,New York: D. Reidel Publishing Company.

Corliss, G. F. (1988), Applications of Differentiation Arithmetic, in Reliability in Computing, edited by R. E. Moore, New York: Academic Press, pp. 127-148.

Corliss, G. F. and Rall, L. B. (1987), Adaptive, Self-Validating Numerical Quadrature, SIAM J. SCI. STAT. COMPUT. , 8, 831-847.

Daley, D. J. (1974), Computation of Bi. and Tri-variate Normal Integrals, Applied Statistics 23, 435-438.

David, P. J. and Rabinowitz, P. (1984), Methods of Numerical Integration, New

York: Academic Press.

Delves, L. M. and Mohamed, J. L. (1988), Computation Methods for Integral Equations, Cambridge: Cambridge University Press.

Divig, D. R. (1979), Calculation of Univariate and Bivariate Normal Probability Functions, Annals of Statistics 23, 435-438.

Donnelly, T. G. (1973), Algorithm 462: Bivariate Normal Distribution, CACM 16, 636.

Drezner, Z. (1978), Computation of the Bivariate Normal Integral, Math. Comp. 32, 277-279.

Drezner, Z. and Wesolowsky, G. O. (1990), On the Computation of Bivariate Normal Integral, J. Statist. Comput. Simul., Vol. 35, 101-107.

Dudley, R. M. (1987), Some Inequalities for Continued Fractions, MATH. COMP. , 49, 585-593.

Gay, D. M. (1988), Interval Least Squares - a Diagnostic Tool, in Reliability in Computing, edited by R. E. Moore, New York: Academic Press, 183-205.

Genz, A. and Kahaner, D. K. (1986), The Numerical Evaluation of Certain Multivariate Normal Integrals, Journal of Computational and Applied Mathematics, 255-258.

Gupta, S. S. (1963), Probability Integrals of Multivariate Normal and Multivariate t, Ann. Math. Statist., 34, 792-838.

IMSL (1987), IMSL STAT/LIBRARY User's Manu, Houston, IMSL Inc., 917-918.

Irvine, D. H. and Savageau, M. A. (1990), Efficient Soloutin of Nonlinear Ordinary Differential Equations Expressed in S-Systems Canonical Form, SIAM Journal on Numerical Analysis, Vol. 27, No. 3, 704-735.

Jerrel, M. E. (1989), Automatic Differentiation Using Almost Any Language,

SIGNUM Newsletter Vol. 24, No. 1, 2-9.

Johnson, L. W. and Riess, R. D. (1982), Numerical Analysis, Massachusetts: Addison-Wesley.

Johnson, N. L. (1959), On an Extension of the Connection Between Poisson and Chi-Square Distributions, Biometrika, 46, 352-363.

Kahan, W. and Leblanc, E. (1985), Anomalies in the IBM ACRITH Package, in Proceedings of 7th Symposium on Computer Arithmetic, edited by Kai Hwang, Silver Springs, MD, IEEE Computer Society Press, 322-331.

Kennedy, W. J. (1990), Statistical Computing Feature: Special Purpose Numerical Tools for Approximating Functions, Statistical Computing and Statistical Graphics Newsletter, Vol. 1, 3-6.

Kennedy, W. J. and Gentle, J. E. (1980), Statistical Computing, New York and Basel: Marcel Dekker.

Krenz, G. S. (1987), Using Weight Function in Self-validating Quadrature, in Numerical Integration, edited by P. Keast and G. Fairweather, New York: D. Reidel Publishing Company.

Lawson, C. L. (1988), Series Reversion as the Reversed Chain Rule, SIGNUM Newsletter, 23, 7-9.

Majumder, K. L. and Bhattacharjee, G. P. (1964), The Incomplete Beta Integral, Algorithm AS 63, 117-120.

Milton, R. C. (1972), Computer Evaluation of Multivariate Normal Integral, Technometrics, Vol. 14, No. 4, 881-889.

Moore, R. E. (1979), Methods and Applications of Interval Analysis: SIAM Studies in Applied Mathematics, 2, Philadelphia: Society for Industrial

Nickel, K. (1975), Interval Mathmetics, Proceedings of the International

Symposium, Karlsruhe, 1975, Berlin: Springer-Verlag.

Nickel, K. (1980), Interval Mathmetics, Proceedings of the International Symposium on Interval Mathmetics, Freiburg, 1980, New York: Academic Press.

Nickel, K. (1985), Interval Mathmetics, Proceedings of the International Symposium on Interval Mathmetics, 1985, New York: Springer-Verlag.

Owens, D. B. (1965), Tables for Computing Bivariate Normal Probabilities, Annals of Mathematical Statistics 27, 1075-1090.

Owens, D. B. (1965), A Special Case of the Bivariate Noncentral T-Distribution, Biometrika 52, 437-446.

Parrish, R. S. and Bargmann, R. E. (1981), A method for the Evaluation of Cumulative Probabilities of Bivariate Distributions Using the Pearson Family, Statistical Distributions in Scientific Work 5, 241-257.

Plant, M. W. and Quandt, R. E. (1989), On the Accuracy and Cost of Numerical Integration in Several Variables, J. Statist. Comput. Simul. Vol. 32, No. 4, 229-248.

Rall, L. B. (1981), Automatic Differentiation: Techniques and Applications, New York: Springer-Verlag.

Ratchek, H. (1980), Centred Forms. SIAM journal of Numerical Analysis, 17, 656-662.

Ratchek, H. and Rokne, J. (1984), Computer Methods for the Range of Functions, Chichester, England: Eliss Horwood Limited.

Rust, P. F. and Voit, E. O. (1990), Statistical Densities, Cumulatives, Quantitiles, and Power Obtained by S-System Differential Equations, Journal of American Statistical Association, 85, 572-578.

Schervish, M. J. (1984), Multivariate Normal Probabilities with Error Bound, Algorithm AS 195, J. Royal Statistical Society, Appl. Statist., Vol. 33, No. 1, 81-94.

Terza, J. V. and Welland, U. (1988), Accuracy and Efficiency Comparisons of Bivariate Normal Approximation Algorithms, Working Paper 12-87-2, Dept. of Economics, Pennsylvania State University.

Tricomi, F. G. (1950), Sulla Funzione Gamma Incomplete, Ann. Math. Pura Appl., 31, 263-279.

Young, J. C. and Minder, C. E. (1974), Algorithm AS76: An Integral Useful in Calculating Non-Central t and Bi-variate Normal Probabilities, Applied Statistics 23, 455-457.

Wall, H. S. (1948), Analytic Theory of Continued Fractions, New York: Van Nostrand.

William, B. J. and Thron, W. J. (1980), Continued Fractions: Analytic Theory and Applications, London: Allison-Wesley.

ACKNOWLEDGEMENTS

To William J. Kennedy, he wisely guided this research and provided an excellent statistical computing research environment.

To Vince Sposito, Kenneth J. Koehler, and Mervyn G. Marasinghe, they guided me to learn data analysis skill.

To Kathy Shelley and Bud Meador, they encouraged me to continue my education at the dark early stages before preliminary examination.

To office mates in statistical computing section, they create an optimal environment to learn and share statistical computing experience.

To Darlene Wicks, she provided the considerable word processing skills that would needed to produce this dissertation.

To Yin-Fang Yiin, Jay-Ming Wang, and Jay-Yun Wang, for putting up with my periodic, seemingly antisocial behavior at critical points when I had to dig in to get things done.

To my mother, father, mother in law, and father in law, for encouraging me to obtain a doctoral degree.