Computational Statistics & Data Analysis 13 (1992) 197–210
North-Holland

A numerical method for accurately approximating multivariate normal probabilities *

Morgan Wang and W.J. Kennedy
Iowa State University, Ames, IA 50011, USA

Received March 1990
Revised September 1990

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).

Keywords: 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.

* This work was supported by National Science Foundation grant DMS 8907940.

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 $X = [X, \bar{X}] = \{x : \underline{X} \le x \le \bar{X}\}$. Interval arithmetic employs the basic operations, defined as follows, for intervals X and Y,

$$
\begin{aligned}
X + Y &= \left[\underline{X} + \underline{Y},\ \bar{X} + \bar{Y}\right], \\
X - Y &= \left[\underline{X} - \bar{Y},\ \bar{X} - \underline{Y}\right], \\
X \cdot Y &= \left[\min\left(\underline{X}\underline{Y},\ \underline{X}\bar{Y},\ \bar{X}\underline{Y},\ \bar{X}\bar{Y}\right),\ \max\left(\underline{X}\underline{Y},\ \underline{X}\bar{Y},\ \bar{X}\underline{Y},\ \bar{X}\bar{Y}\right)\right], \\
1/Y &= \left[1/\bar{Y},\ 1/\underline{Y}\right],\ 0 \notin Y, \\
X/Y &= X \cdot (1/Y).
\end{aligned}
\tag{1.1}
$$

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 over $X_1, \ldots, X_n$ is

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

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 $IP = [\underline{P}, \bar{P}]$ of the true probability P. Moreover, we will arrange that the width $w(IP) = \bar{P} - \underline{P}$ of this interval is small enough so that $w(IP) \le 2\alpha$ and $w(IP) \le \rho \cdot |\underline{P} + \bar{P}|$, 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 inclusion. 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 probabiiity. 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 evaluation of multivariate normal integrals based on interval arithmetic and Taylor series

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

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

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

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

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

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

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

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

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

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

Let us express P in (2.1) as

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

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

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

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

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 $f(y_1)(h(y_1, c_2(y_1)))_i$ 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

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

with $d_2 = (B_2 - A_2)/(2l_{22}) = b_2(y_1) - c_2(y_1)$. 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 $\epsilon_{1i} \epsilon [a_1, b_1] = X_1$.

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

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.

($\mathbf{R}_1$) The united extension of a real-valued function is inclusion monotonic.

($\mathbf{R}_2$) 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.

($\mathbf{R}_3$) 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.

($\mathbf{R}_4$) 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

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

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 $(h(y_1, \epsilon_2(y_1)))_{n_2}$ 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 $\delta \epsilon X_1$ such that

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

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 $\cup_{y_1 \epsilon X_1} X_2(y_1)$, 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 extension of $(h(\delta, \epsilon_2(\delta)))_{n_2}$, denoted by

$(H(X_1, X_2(X_1)))_{n_2}$, is inclusion monotonic. Therefore, the interval inclusion IE_1 of E_1 is inclusion monotonic, where

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

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 (1963), Milton (1972), Bohrer and Schervish (1981), Schervish (1984), Genz and Kahaner (1986), 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" midpoint used was necessarily very close to the theoretically correct probability. The estimated

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
$ABS \leqslant \alpha = 0.00001$ satisfied (% cases)	100.00	99.35
$REL \leqslant \rho = 0.00001$ satisfied (% case)	12.63	98.08
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.

absolute error and estimated relative error for both Schervish's and the Scalar Algorithms were defined as

$$ABS = |\text{computed value} - \text{"true value"}| \quad \text{and}$$

$$REL = ABS/\text{"true value."}$$

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.

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
$ABS \leqslant \alpha = 0.00000001$ satisfied (% cases)	100.00	100.00
$REL \leqslant \rho = 0.00000001$ 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.

I.D.	Range		Correlation			Inclusion of probability	
	$A(I)$	$B(I)$	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	0.94278893709764
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] $A(1) = -1.2$ $A(2) = 0.5$ $A(3) = -1.0$

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

Table 4
Interval inclusion of integrals used for the third experiment.

I.D.	Range		Correlation			Inclusion of probability	
	$A(l)$	$B(l)$	ρ_{12}	ρ_{13}	ρ_{23}	Lowerbound	Upperbound
1	−2.0	[a]	−0.99	0.99	−0.99	0.3413447446	0.3413447475
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] $\rho_{14}=0.0$, $\rho_{24}=0.0$, $\rho_{34}=0.5$.
[i] $\rho_{14}=0.1$, $\rho_{24}=0.1$, $\rho_{34}=0.1$.
[j] $\rho_{14}=0.3$, $\rho_{24}=0.5$, $\rho_{34}=0.6$.
[k] $\rho_{14}=0.7$, $\rho_{24}=0.7$, $\rho_{34}=0.7$.

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 results 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	0.9400.45399	0.000001	0.000001	10
6	0.9367437845	0.000001	0.000001	11
7	0.9343125193	0.000002	0.000002	09
8	0.9339763586	0.000002	0.000002	09
9	0.9336343452	0.000002	0.000001	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.

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

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 results 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.

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

Table 9
Scalar implementation results for the third experiment (α, $\rho = 10^{-5}$).

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	0.8327168572	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 results for the third experiment (α, $\rho = 10^{-5}$).

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.

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.

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.

References

[1] G. Alefeld and J. Herzberger, Introduction to Interval Computation (Academic Press, New York, 1983).
[2] R.E. Bohrer and M.J. Schervish, An Error-Bounded Algorithm for Normal Probabilities of Rectangular Regions, *Technometrics*, **23**, No. 3 (1981) 297–300.
[3] G.F. Corliss, Applications of Differentiation Arithmetic, Reliability in Computing (Academic Press, 1988) 127–148.
[4] G.F. Corliss and L.B. Rall, Adaptive, Self-Validating Numerical Quadrature, *SIAM, J. Sci. Stat. Comput.*, **8**, No. 5 (1987) 831–847.
[5] A. Genz and D.K. Kahaner, The Numerical Evaluation of Certain Multivariate Normal Integrals, *Journal of Computational and Applied Mathematics*, **16** (1986) 255–258.
[6] S.S. Gupta, Probability Integrals of Multivariate Normal and Multivariate *t*, *Ann. Math. Statist.*, **34** (1963) 792–838.
[7] M.E. Jerrell, Automatic Differentiation Using Almost Any Language, *SIGNUM Newsletter*, **24**, No. 1 (1989) 2–9.
[8] L.W. Johnson and R.D. Riess, Numerical Analysis (Addison-Wesley, Massachusetts, 1982) 14.

[9] C.L. Lawson, Series Reversion as the Reversed Chain Rule, *SIGNUM Newsletter*, **23**, No. 1 (1988) 7–9.
[10] R.C. Milton, Computer Evaluation of Multivariate Normal Integral, *Technometrics*, **14**, No. 4 (1972) 881–889.
[11] R.E. Moore, Methods and Applications of Interval Analysis (SIAM Studies in Applied Mathematics 2, SIAM, Philadelphia, 1979).
[12] M.W. Plant and R.E. Quandt, On the Accuracy and Cost of Numerical Integration in Several Variables, *J. Statist. Comput. Simul.*, **32**, No. 4 (1989) 229–248.
[13] L.B. Rall, Automatic Differentiation: Techniques and Applications, Lecture Notes in Computer Science, Vol. 120 (Springer, Berlin, New York, 1981).
[14] H. Ratschek and J. Rokne, Computer Methods for the Range of Functions (Ellis Horwood Limited, England, 1984).
[15] M.J. Schervish, Multivariate Normal Probabilities with Error Bound, Algorithm AS 195, *J. Royal Statistical Society, Appl. Statist.*, **33**, No. 1 (1984) 81–94.