

**NUMERICAL COMPUTATION  
OF  
MULTIVARIATE NORMAL AND MULTIVARIATE -T PROBABILITIES  
OVER ELLIPSOIDAL REGIONS**

27JUN01

Paul N. Somerville  
University of Central Florida

**Abstract:** An algorithm for the computation of multivariate normal and multivariate t probabilities over general hyperellipsoidal regions is given. A special case is the calculation of probabilities for central and noncentral F and  $\chi^2$  distributions. A FORTRAN 90 program MVELPS.FOR incorporates the algorithm.

**Key words:** multivariate normal, multivariate t, ellipsoidal regions, noncentral F, noncentral t

### 1. INTRODUCTION

Somerville (1998) developed procedures for calculation of the multivariate normal and multivariate-t distributions over convex regions bounded by hyperplanes. This paper extends the procedures to general ellipsoidal regions. There are a number of potential applications. One is in the field of reliability, and in particular relating to the computation of tolerance factors for multivariate normal populations. Krishnamoorthy, K. and Mathew, Thomas (1999) discussed a number of approximation methods in the their calculation. Another relates to the calculation of probabilities for linear combinations of central and noncentral chisquare and F. Again a number of authors, including Farebrother (1990) have studied the problem. A number of approximation methods exist, including those of Satterthwaite-Welsh, Satterthwaite, F.E. (1946) and Hall-Buckley-Eagleson, Buckley, M. J. and Eagleson, G. K. (1988). Other potential applications are in factor and discriminant analysis.

The program, of course, can also be used for the special case of calculating probabilities for a central or noncentral chisquare or F distribution.

### 2. EVALUATION OF MULTIVARIATE-T AND MULTIVARIATE NORMAL INTEGRALS

Let  $\mathbf{x}^T = (x_1, x_2, \dots, x_k)$  have the distribution  $f(\mathbf{x}) = \text{MVN}(\boldsymbol{\mu}, \Sigma\sigma^2)$ , where  $\Sigma$  is a known positive definite matrix and  $\sigma^2$  is a constant. We wish to evaluate  $f(\mathbf{x})$  over A, an ellipsoidal region given by

$$(\mathbf{x} - \mathbf{m}^*)^T \mathbf{U} (\mathbf{x} - \mathbf{m}^*) = 1.$$

That is, we wish to obtain

$$P = \int_A f(\mathbf{x}) \, d\mathbf{x}.$$

If  $\sigma^2$  is known, then without loss of generality, set  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\sigma = 1$ , and let  $\Sigma$  be the correlation matrix. If  $\sigma^2$  is estimated by  $s^2$  with  $\nu$  degrees of freedom such that  $\nu s^2 / \sigma^2$  is a chi-square variate, then, without loss of generality we may assume  $f(\mathbf{x})$  has the central multivariate-t distribution and  $\Sigma$  is the correlation matrix.

Let  $\Sigma = \mathbf{T}\mathbf{T}^T$  (Cholesky decomposition) and set

$$\mathbf{x} = \mathbf{T} \mathbf{w}.$$

The random variables  $w_1, w_2, \dots, w_k$  are independent standard normal or spherically symmetric t variables. Without loss of generality we may assume that the axes of the ellipsoid are parallel to the coordinate axes and the ellipsoid has the equation

$$(\mathbf{w} - \mathbf{m})^T \mathbf{B}^{-1} (\mathbf{w} - \mathbf{m}) = 1. \tag{2.1}$$

where  $\mathbf{B}$  is a diagonal matrix with diagonal elements given by  $b_i$ .

If the elements  $b_i$  are equal, we have the special case of a spheroidal region. (We show in section 5. that evaluation of multivariate normal and multivariate t distributions over spheroidal regions is equivalent to evaluating noncentral chisquare and noncentral F distributions respectively.)

Case a)  $\sigma^2$  is known (Multivariate Normal)

Let  $r^2 = \mathbf{w}^T \mathbf{w}$ . Since  $\sigma^2$  is known,  $r^2$  is distributed as chi-square with  $k$  degrees of freedom. The procedure is to choose random directions in the  $w$  coordinate system (say  $m$  of them). Using the “nonbinning” or “direct” method, an unbiased estimate of the probability content corresponding to each direction is obtained based on the distance of the boundary of the ellipsoid from the origin. Their arithmetic mean is an estimate of the probability content. To obtain a standard error, the process is repeated (say  $n$  times), and a standard error calculated for the mean of the means.

Using the binning method, an empirical estimate of the distribution of the distances is obtained, and quadrature is used to estimate the probability content of the ellipsoid. As before, the process is repeated so as to provide a standard error.

i) “Nonbinning” method:

If the ellipsoid contains the origin, then for each random direction there is a unique distance (say  $r(i)$ ) to the boundary. An unbiased estimate of the probability content of the ellipsoid is given by

$$\Pr[r \leq r(i)] = \Pr[\chi^2 \leq (r(i))^2].$$

If  $P(\sigma)$  is the probability content of the ellipsoid when  $\sigma$  is assumed known, then

$$P(\sigma) \approx \sum \Pr[\chi^2 \leq r(i)^2]/N$$

where  $N$  is the number of unbiased estimates.

If the ellipsoid does not contain the origin, then for a random direction, a line from the origin in that direction intersects the boundary of the ellipsoid in two points (say  $r(i) \geq r^*(i)$ ) or does not intersect the ellipsoid. If the line intersects the boundary, then an unbiased estimate of the probability content of the ellipsoid is given by

$$\Pr[r \leq r(i)] - \Pr[r \leq r^*(i)] = \Pr[\chi^2 \leq r(i)^2] - \Pr[\chi^2 \leq r^*(i)^2].$$

If the line does not intersect the ellipsoid, an unbiased estimate is 0.

For  $N$  unbiased estimates,

$$P(\sigma) \approx \sum \{\Pr[\chi^2 \leq r(i)^2] - \Pr[\chi^2 \leq r^*(i)^2]\}/N$$

where the summation is over all cases where the line intersects the boundary.

In both cases it is convenient to let  $N = 2 * m$  where having chosen a random direction we use also the opposite direction.

ii) “Binning” method:

For the case where the origin is in the ellipsoid, let  $H(r)$  be the cumulative distribution function of  $r$ , the distance from the origin to the boundary in a randomly chosen direction. The probability content of the ellipsoid is

$$P(\sigma) = \int_a^b (1-H(r)) \Pr[\chi^2 \leq r^2] dr + \Pr[\chi^2 \leq a^2]$$

where  $a$  and  $b$  are respectively the smallest and largest distances of the boundary of the ellipsoid from the origin. Using Gauss-Legendre quadrature,

$$P(\sigma) \approx \sum w_i (1-H(r_i)) \Pr[\chi^2 \leq r_i^2] + \Pr[\chi^2 \leq a^2]$$

where  $r_i$  and  $w_i$  are the abscissae and weights respectively for the Gauss-Legendre quadrature and the values  $r_i$  are used in the binning process to obtain the empirical estimates  $H(r_i)$  of the cumulative distribution function of  $r$ .

If the origin is not in the ellipsoid,

$$P(\sigma) = p_1 \left[ \int_a^b (1-H_1(r)) \Pr[\chi^2 \leq r^2] dr - \int_a^b (1-H_2(r)) \Pr[\chi^2 \leq r^2] dr \right]$$

$$= p_1 \int_a^b \{H_2(r) - H_1(r)\} * \Pr[\chi^2 \leq r^2] dr$$

where  $p_1$  is the probability that a line through the origin in a random direction goes through the ellipsoid, and  $H_1(r)$  and  $H_2(r)$  are respectively the cumulative distribution functions of the larger and smaller distances from the origin to the boundary for intersecting lines. Using Gauss-Legendre quadrature we obtain

$$P(\sigma) \approx p_1 \sum w_i \{H_2(r_i) - H_1(r_i)\} * \Pr[\chi^2 \leq r_i^2].$$

where  $r_i$  and  $w_i$  are the abscissae and weights respectively for the Gauss-Legendre quadrature and the values for  $r_i$  are used in the binning process to obtain the empirical estimates  $H_1(r_i)$  and  $H_2(r_i)$  of the cumulative distribution functions  $H_1(r)$  and  $H_2(r)$ . The value  $p_1$  is simultaneously estimated. Again mocar random directions are used for the estimates  $H_1(r_i)$  and  $H_2(r_i)$ . The procedure is repeated irep times, with the estimate of  $P$  being the mean of the irep estimates, and the repetitions used to obtain a standard error.

We summarize the results for the case where  $\sigma$  is known in the following Table 2.1:

<u>ORIGIN</u>	<u>NONBINNING</u>	<u>BINNING</u>
IN ELLIPSOID	$\sum \Pr[\chi^2 \leq r(i)^2]/N$	$\sum w_i * (1 - H(r_i)) * \Pr[\chi^2 \leq r_i^2] + \Pr[\chi^2 \leq a^2]$
NOT IN	$\sum \{\Pr[\chi^2 \leq r(i)^2] - \Pr[\chi^2 \leq r^*(i)^2]\} / N$	$p_1 \sum w_i * \{H_2(r_i) - H_1(r_i)\} * \Pr[\chi^2 \leq r_i^2]$

TABLE 2.1  
Estimate of  $P(\sigma)$  when  $\sigma$  is known

Case b)  $\sigma^2$  is not known (Multivariate-t)

The ellipsoid is given by  $(\mathbf{w} - \mathbf{m})^T \mathbf{B}^{-1} (\mathbf{w} - \mathbf{m}) = 1$ . Suppose  $s = \sigma$ . Then the probability content of the ellipsoid is  $P(\sigma)$  as given in the previous section. Let  $g(s)$  be the frequency function for  $s$  ( $vs^2/\sigma^2$  has the  $\chi^2$  distribution with  $v$  degrees of freedom). Then the unconditional probability content of the ellipsoid is given by

$$P = \int_0^\infty P(s) g(s) ds.$$

Using Gauss-Laguerre quadrature, we obtain

$$P \approx \sum w_i * P(s_i) * g(s_i)$$

where  $s_i$  and  $w_i$  are the quadrature abscissae and weights respectively. We use the formulae for  $P(\sigma)$  given in the table above.

### 3. DISTANCE OF ORIGIN FROM BOUNDARY

If the equation of the ellipsoidal region is given by  $(\mathbf{w} - \mathbf{m})^T \mathbf{B}^{-1} (\mathbf{w} - \mathbf{m}) \leq g$ , then the origin will be in the region if  $\mathbf{m}^T \mathbf{B}^{-1} \mathbf{m} - g \leq 0$ . Let  $\mathbf{d}$  be a random (normalized) direction. Then the equation of the line from the origin in the direction  $\mathbf{d}$  is given by  $\mathbf{w} = t \mathbf{d}$ . Substituting in the equation for the ellipsoidal boundary we obtain

$$t^2 aa - 2 t bb + cc = 0$$

where  $aa = \mathbf{d}^T \mathbf{B}^{-1} \mathbf{d}$ ,  $bb = \mathbf{m}^T \mathbf{B}^{-1} \mathbf{d}$  and  $cc = \mathbf{m}^T \mathbf{B}^{-1} \mathbf{m} - g$ . Let  $b2ac = bb * bb - aa * cc$ . Then, if the origin is in the ellipsoid, the distances from the origin to the ellipsoid boundary in the directions  $\mathbf{d}$  and  $-\mathbf{d}$  are  $|bb + \sqrt{b2ac}|$  and  $|bb - \sqrt{b2ac}|$ . If the origin is not in the ellipsoid, then if  $b2ac < 0$ , for both directions  $\mathbf{d}$  and  $-\mathbf{d}$  the line does not cross the ellipsoid. If  $b2ac \geq 0$ , the line crosses the boundary in one of the directions  $\mathbf{d}$  or  $-\mathbf{d}$  at distances from the origin  $r(i) = |bb| + \sqrt{b2ac}$  and  $r^*(i) = ||bb| - \sqrt{b2ac}|$ . The line does not cross the boundary in the other direction.

### 4. EXTREMAL DISTANCES FROM ORIGIN TO BOUNDARY

The author has written a Fortran 90 program to calculate the largest and smallest distances from the origin to a general hyper-ellipsoid in  $k$  dimensions. The program is called BOUND.FOR and a paper describing the program will be submitted for publication. The program can also be used to

find all local extrema which may exist. The present program MVELPS.FOR uses the program BOUND.FOR as a subroutine.

## 5. NONCENTRAL CHISQUARE AND F PROBABILITIES

The methodology can be used to obtain cumulative probabilities for chisquare and F distributions. If  $\mathbf{w}$  is MVN( $\mathbf{0}$ ,  $\mathbf{I}$ ) with dimension  $k$ , and  $u$  has the noncentral chisquare distribution with  $k$  degrees of freedom and noncentrality parameter  $\lambda$ , it is not difficult to show that  $\Pr[u \leq f_0]$  is equal to the probability content of  $\mathbf{w}$  over a spheroid with radius  $\sqrt{f_0}$  and center a distance  $\lambda$  from the origin.

To show this, put  $\mathbf{z} = \mathbf{w} + \boldsymbol{\lambda}$ , where  $\boldsymbol{\lambda}$  is a constant. By Theorem 3.3.5 of Anderson (1984)  $u = \mathbf{z}^T \mathbf{z}$  has the chisquare distribution with  $k$  degrees of freedom and noncentrality parameter  $\lambda = \sqrt{\boldsymbol{\lambda}^T \boldsymbol{\lambda}}$ . Then  $u = \sum (w_i + \lambda_i)^2$ , and  $\Pr[u \leq f_0] = \Pr[\sum (w_i + \lambda_i)^2 \leq f_0]$  which is the probability content of a spheroid with radius  $\sqrt{f_0}$  and center a distance  $\lambda$  from the origin.

We now modify our procedure for the special case of calculating noncentral chisquare probabilities. Without loss of generality, let  $\boldsymbol{\lambda} = (\lambda, 0, \dots, 0)$ . Then

$$\Pr[\chi^2(\lambda) \leq c] = \Pr[\sum w_i^2 + 2 \lambda w_1 + \lambda^2 - f_0 \leq 0].$$

Let  $\mathbf{d}$  be a random (normalized) direction in the  $k$ -space, and set  $\mathbf{w}_i = t \mathbf{d}$ . To find the intersection of this line with the spheroid boundary, we substitute  $\mathbf{w}_i = t \mathbf{d}$  in

$$\sum w_i^2 + 2 \lambda w_1 + \lambda^2 - f_0 = 0$$

obtaining  $t^2 + 2 \lambda d_1 t + \lambda^2 - f_0 = 0$ .

Solving for  $t$ , we have

$$t = -\lambda d_1 \pm \sqrt{\lambda^2 d_1^2 - (\lambda^2 - f_0)}.$$

When  $\lambda^2 - f_0 \leq 0$ , the origin is in the spheroid. If the origin is not in the spheroid, and  $\lambda^2 d_1^2 - (\lambda^2 - f_0) \geq 0$ , the line crosses the boundary of the spheroid in two points. If the origin is not in the spheroid, and  $\lambda^2 d_1^2 - (\lambda^2 - f_0) < 0$ , the line does not cross the boundary. The distance from the origin to the boundary is  $|t|$ . Since only one component,  $d_1$ , of  $\mathbf{d}$  is used in computing  $t$ , the program uses each element of  $\mathbf{d}$  and also its negative (total of  $2*k$ ) as  $d_1$  as a means of increasing efficiency.

We next modify the procedure for the special case of calculating noncentral F probabilities.

Suppose  $s^2$  is an estimate of  $\sigma^2$  (assumed to be equal to 1 with no loss of generality) with  $v$  degrees of freedom such that  $v s^2 / \sigma^2$  has the  $\chi^2$  distribution. Then using Theorem 5.4.1 of Anderson (1984),

$$\begin{aligned} F(k, v; \lambda) &= (\mathbf{z}^T \mathbf{z} / k) / ([v s^2 / \sigma^2] / v) \\ &= (\mathbf{z}^T \mathbf{z} / k) / s^2 \end{aligned}$$

has the noncentral F distribution with degrees of freedom  $k$  and  $v$ , and noncentrality parameter  $\lambda$ .

Thus  $F(k, v; \lambda) = \sum (w_i + \lambda_i)^2 / (k s^2)$ ,

and  $\Pr[F(k, v; \lambda) \leq c | s^2] = \Pr[\sum (w_i + \lambda_i)^2 \leq c k s^2 | s]$ .

This is the probability content of a spheroid with center a distance  $\lambda$  from the origin and with radius  $\sqrt{c*k*s^2}$ , for any fixed value of  $s$ . To obtain the unconditional probability, we integrate over  $s$

$$\Pr[F(k, v; \lambda) \leq c] = \int \Pr[F(k, v; \lambda) \leq c | s] g(s) ds$$

where  $g(s)$  is the frequency function for  $s$ , and the integration is from 0 to  $\infty$ .

Having a procedure for calculating  $\Pr[F(k, v; \lambda) \leq c | s]$  (noncentral  $\chi^2$ ), it is convenient to use Gauss Laguerre quadrature to calculate the unconditional probability.

The program MVELPS.FOR has a special shortcut for calculation of noncentral chisquare and F distribution probabilities which takes advantage of the special case of a spheroidal region. There are existing programs for the above distributions and no claims are made as to the superiority or inferiority of any of the programs.

## 6. ACCURACY AND COMPUTATION TIMES

To assess the accuracies attainable using MVELPS.FOR, runs were made for  $k = 3, 6$  and  $10$ , with ellipsoid centers both inside and outside of the ellipsoidal region, and for both the multivariate normal and the multivariate-t distribution. The lengths of the semiaxes and centers, in standardized units are given in Table 6.1. All runs were made using a PC with an AMD 800 processor and a Lahey 95 compiler.

	<u><math>k = 3</math></u>	<u><math>k = 6</math></u>	<u><math>k = 10</math></u>
Axes	(4,2,1)	(1,2,2,3,3,4)	(1,2,2,2,2,3,3,3,3,4)
Center (inside)	(0,1,0)	(0,2,0,0,0,0)	(0,1,0,0,0,0,0,0,0,0)
Center (outside)	(0,3,0)	(0,3,0,0,0,0)	(0,3,0,0,0,0,0,0,0,0)

Table 6.1  
Input for Accuracy Runs

Using 10 replications, each with 1000 random directions, runs were made using both the binning and nonbinning methods. For the binning procedure 20 point quadrature was used for both Gauss-Legendre and Gauss-Laguerre. Standard errors were essentially the same for the binning and nonbinning methods. Table 6.2 gives standard errors and run times in seconds.

	<u><math>k = 3</math></u>		<u><math>k = 6</math></u>		<u><math>k = 10</math></u>	
	<u>df = 10</u>	<u>df = <math>\infty</math></u>	<u>df = 10</u>	<u>df = <math>\infty</math></u>	<u>df = 10</u>	<u>df = <math>\infty</math></u>
Center in	.00050	.00031	.00033	.00026	.00030	.00012
	.82 (.38)	.55 (.17)	.55 (.28)	.44 (.27)	.77 (.38)	.71 (.39)
Center out	.00044	.00021	.00033	.00017	.00016	.00007
	.44 (.16)	.22 (.11)	.50 (.27)	.22 (.22)	.61 (.38)	.38 (.30)

Table 6.2  
Accuracies and Run Times

First line gives standard errors. Second line gives run times in seconds (first number for binning method and number in parentheses for nonbinning)

To ascertain the effect of run size on standard error and running times, further runs were made with result given in Table 6.3. 20 points were used for both Gauss-Legendre and Gauss-Laguerre quadrature for the binning procedure. Semiaxes of (4,2,1) with center (0,1,0) for  $k = 3$  were used. The following table gives the calculated standard errors and times. Times are in seconds with the numbers in parentheses for the nonbinning procedure.

<u>mocar</u>	<u>irep</u>	<u>df = 20</u>		<u>df = <math>\infty</math></u>	
		<u>se</u>	<u>time</u>	<u>se</u>	<u>time</u>
100	10	.008 259	.05 (.00)	.003 631	.00 (.00)
1000	10	.001 444	.06 (.11)	.000 735	.00 (.06)
10000	10	.000 565	.43 (.99)	.000 311	.17 (.49)
1000000	10	.000 196	1.93 (5.65)	.000 093	1.81 (5.25)
10000000	10	.000 063	16.91 (55.97)	.000 044	17.8 (51.79)

Table 6.3  
Run times and standard errors  
for binning and nonbinning procedures

Regression analyses showed near perfect linear relationships between running times and number of random directions, and between the logarithms of standard errors and number of random

directions. For example, for the binning procedure, the approximate linear relationship between running time and number of random directions is

$$\text{time} = .00316 + 5.78\text{E-}6 * (\# \text{ random directions}).$$

It may be noted that except for total random directions of 1000 for the multivariate normal case, the binning procedure consistently requires less running time than the nonbinning procedure. The binning procedure is thus recommended.

It needs to be emphasized that the standard error of the estimate is a function of the shape and position of the ellipsoid (after the “standardizing” transformations). If the ellipsoid is a spheroid centered at the origin, the accuracy is independent of the number of random directions and is a function of the processor and the computer program. The standard error, in general, is a function of the degree of departure of the ellipsoid from a spheroid centered at the origin.

Special runs were made to ascertain the accuracy and running times of the shortcut method for the noncentral F distribution. A total of 180 runs were made, with  $k = 2, 3, 5, 10$  and  $20$ ,  $\lambda$  and  $f_0$  equal to  $.1, .2, .5, 1, 2, 5$  and  $10$ . For each run, 10 estimates, each using 1000 random directions were used. The estimates and standard errors were consistent with existing software programs, e.g. from the M. D. Anderson Cancer Center Statistical Archives (see also <http://www.stat.ucla.edu/statistical/calculator>). Some results follow. When  $\lambda^2 - f_0$  was less than zero (axes origin was inside ellipsoid), the largest standard error estimate was .0001 (estimated probability of .804650 for  $\lambda = 2$  and  $f_0 = 5$ ). When  $\lambda^2 - f_0$  was greater than zero (axes origin was outside ellipsoid), the largest estimated standard error was .000584 (estimated probability of .059056 for  $k = 2, \lambda = 1$  and  $f_0 = .1$ ). Of the 57 cases where the estimated probability was greater than .5, and  $\lambda^2 - f_0$  was less than zero, 42 cases had estimated standard errors less than  $10^{-6}$ . Estimated standard error tended to decrease both for increasing values of  $k$  and increasing values for the estimated probability. The ratio of the estimated standard error divided by the estimated probability tended to increase with increasing values of  $\lambda^2 - f_0$  and decrease with increasing values of the estimated probability.

Average running times ranged from 0.075 and .109 seconds for  $k=2$  and  $3$  respectively to .379 and .843 seconds for  $k=10$  and  $20$  respectively.

The same runs were made for the noncentral chisquare with modestly higher estimated accuracies and modestly smaller running times.

Running times and accuracies for the general ellipsoid case should be well approximated by those for the special spheroid case above. Some examples are given in section 8.

## 7. USING MVELPS.FOR

MVELPS.FOR is a FORTRAN 90 program which calculates the probability content of a general ellipsoid:

$$(\mathbf{x} - \mathbf{m}^*)^T \mathbf{U} (\mathbf{x} - \mathbf{m}^*) = \text{ellipg}$$

when  $\mathbf{x}$  is  $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma} \sigma^2)$ . The element  $\sigma^2$  may be known or estimated with  $\mathbf{v}$  degrees of freedom. The program requires that the files ELIP.IN and ELIP.OUT exist before the program is run. ELIP.IN contains the required input parameters and the program returns the output from the program in the file ELIP.OUT.

Except for the “shortcut method” used for the special case of calculation of noncentral  $\chi^2$  and noncentral F probabilities (described later), the file ELIP.IN requires six “lines” of input. A “line” may actually be a vector or matrix. The six “lines” are:

LINE 1:  $k$  seed  $\mathbf{v}$  mocrar irep meth

LINE 2: ellipg

LINE 3:  $\boldsymbol{\Sigma}$

LINE 4:  $\boldsymbol{\mu}$

LINE 5:  $\mathbf{U}$   
 LINE 6:  $\mathbf{m}^*$

LINES 1 and 2: The elements k, mocar, irep and ellipg have already been discussed. The degrees of freedom v for the estimate of  $\sigma^2$  is given the value “-1” when  $\sigma^2$  is assumed to be known. The seed for the random generator is “seed”. To calculate the probability using the binning method, set “meth” = 1. To calculate the probability using the non-binning or direct method, set “meth” = 0.

LINES 3 and 5:  $\Sigma$  and  $\mathbf{U}$  are symmetric matrices. It is permissible to input only the lower triangular portion of these matrices. Each row occupies a separate line.

If a matrix is a diagonal with all diagonal elements equal to “d”, then one line containing the single element “-d” is sufficient.

LINES 4 and 6: The k elements of the vector make up the line.

The “shortcut method”, used for calculating noncentral  $\chi^2$  and noncentral F probabilities, requires two lines:

LINE 1: k seed v mocar irep meth

LINE 2:  $\lambda$  f0.

LINE 1 is the same as before, except that now “meth” is set to -1.

The noncentrality parameter  $\lambda$  has previously been defined, and f0 is the upper limit of the noncentral  $\chi^2$  or noncentral F function as in Prob[  $F \leq f0$  ].

ELIP.IN can contain several sets of instructions. After the last set of instructions, the program expects a line of four or more -1's.

## 8. EXAMPLES

### Example 1.

Let  $\mathbf{x}$  have the multivariate t distribution with  $k = 3$ , all correlations equal to .5, mean vector  $\mathbf{0}$ , and with 20 degrees of freedom. We want to find the probability content of the ellipsoidal region

$$(\mathbf{x} - \mathbf{m}^*)^T \mathbf{U} (\mathbf{x} - \mathbf{m}^*) = 1,$$

where  $\mathbf{U}$  is a diagonal matrix with diagonal (.25, .2, .1) and  $\mathbf{m}^*$  is (1, .5, .25). Choose 357 as seed and make 10 preliminary estimates each using 10,000 random directions. We would like to first use the binning method and then the nonbinning method. ELIP.IN could be as follows:

```
3 357 20 10000 10 1
```

```
1
```

```
1
```

```
.5 1
.5 .5 1
0 0 0
.25
0 .2
0 0 .1
1 .5 .25
```

```
3 357 20 10000 10 0
```

```
1
```

```
1
```

```
.5 1
.5 .5 1
0 0 0
.25
0 .2
0 0 .1
1 .5 .25
```

-1 -1 -1 -1 -1 -1

The following output (in ELIP.OUT) is obtained:

date and time 20010620222229.380  
Binning method used  
elapsed time in seconds is .440  
no of pops is 3 seed is 357  
df for var est is 20  
value of integral is calculated 10 times  
mean value is 6.80487812E-01  
standard error of mean is 5.13384177E-04  
-----

date and time 20010620222229.82  
elapsed time in seconds is .830  
no of pops is 3 seed is 357  
df for var est is 20  
value of integral is calculated 10 times  
mean value is 6.80493176E-01  
standard error of mean is 5.23232622E-04  
-----

### Example 2

Find the probability content of the ellipsoidal region  $(\mathbf{x} - \mathbf{m}^*)^T \mathbf{U} (\mathbf{x} - \mathbf{m}^*) = 1$ , where  $k = 4$ ,  $\mathbf{m}^* = (1, 1, 1, 1)$  and  $\mathbf{U}$  is diagonal with diagonal elements equal to .2. Assume  $\mathbf{x}$  is MVN(0, I). Using the binning method with 10 estimates each with 1000 random directions and seed of 297, ELIP.IN may be given by

4 297 -1 1000 10 1  
1  
-1  
0 0 0 0  
-.2  
1 1 1 1  
-1 -1 -1 -1

The output from ELIP.OUT is:

date and time 20010620222230.650  
Binning method used  
elapsed time in seconds is .050  
no of pops is 4 seed is 297  
variance assumed known  
no of random dir. is 2000.000000  
value of integral is calculated 10 times  
mean value is 3.08498234E-01  
standard error of mean is 6.20856299E-04  
-----

### Example 3

Find  $\Pr[F \leq 3]$  where  $F$  is the noncentral  $F$  distribution with degrees of freedom 4 and 25 for the numerator and denominator respectively, and the noncentrality parameter is 1.2. Make 12 estimates with 1200 random directions each. ELIP.IN may be given by

4 9133 25 1200 12 -1  
1.2 3  
-1 -1 -1 -1 -1 -1

The output from ELIP.OUT is:

Noncentrality parameter is 1.200000

date and time 20010620222230.700  
elapsed time in seconds is .220  
no of pops is 4 seed is 9133  
df for var est is 25  
value of integral is calculated 12 times  
mean value is .905823290  
standard error of mean is 9.21626452E-06  
-----

#### **Example 4**

Find  $\Pr[\chi^2 \leq 2.5]$  where  $\chi^2$  is the noncentral chisquare distribution with 5 degrees of freedom and noncentrality parameter 1.6. Use a single estimate (no error estimate will be given) with 10,000 random directions. ELIP.IN may be given by

```
5 9571 -1 10000 1 -1  
1.6 2.5
```

The output from ELIP.OUT is:

```
Noncentrality parameter is 1.600000  
date and time 20010620222230.920  
elapsed time in seconds is .110  
no of pops is 5 seed is 9571  
variance assumed known  
no of random dir. is 100000.000000  
value of integral is calculated 1 times  
mean value is 9.27048028E-02  
-----
```

## **9. SUMMARY AND CONCLUSIONS**

A methodology has been developed and a Fortran 90 program written to find the probability content of an hyperellipsoid when the underlying distribution is multivariate normal or multivariate-t. The methodology has been used by the author for two applications. In the computation of tolerance factors for multivariate populations there are four parameters:  $N$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  (Krishnamoorthy, K. and Mathew, Thomas (1999)). The author is writing a Fortran 90 program which will be submitted for publication which calculates any one of the four parameters, given the other three. The author has a Fortran 90 program in an advanced state of development which calculates cumulative probabilities for linear combinations of central and noncentral linear combinations of chisquare and F.

The present program (MVELPS) can be used to calculate probabilities for the special case where the ellipsoids are actually spheroids, namely for central noncentral F and  $\chi^2$  distributions. No claims are made regarding superiority or inferiority to other existing programs for this case.

## **10. REFERENCES**

- Anderson, T. W. (1984) "An Introduction to Multivariate Statistical Analysis", New York, Wiley.
- Buckley, M. J. and Eagleson, G. K. (1988). "An Approximation to the Distribution of Quadratic Forms in Random Normal Variables", *Australian Journal of Statistics* **30A**, 150-159.
- Farebrother, R. W. (1990) "The Distribution of a Quadratic Form in Normal Variables", *Applied Science* **39**, 294-309.
- Krishnamoorthy, K. and Mathew, Thomas (1999) "Comparison of Approximation Methods for Computing Tolerance Factors for a Multivariate Normal Population", *Technometrics*, Vol. 41, No. 3, 234-249.

Satterthwaite, F.E. (1946) "An Approximate distribution of the Estimates of Variance Components", *Biometrika*, **33**, 110-114.

Krishnamoorthy, K. and Mathew, Thomas (1999) "Comparison of Approximation Methods for Computing Tolerance Factors for a Multivariate Normal Population", *Technometrics*, Vol. 41, No. 3, 234-249.

Somerville, P. N. (1998) "Numerical Computation of Multivariate Normal and Multivariate-t Probabilities Over Convex Regions" *Journal of Computational and Graphical Statistics*, Vol. 7, No 4, 529-544.