Journal of
Statistical Theory and Practice
Volume 5, No. 4, December 2011

Recursive Integration Methodologies with Applications to the Evaluation of Multivariate Normal Probabilities

A. J. Hayter, *Department of Business Information and Analytics, University of Denver, Denver, USA.*
Email: Anthony.Hayter@du.edu

Received: August 11, 2010 *Revised:* March 4, 2011

Abstract

This paper considers the evaluation of probabilities defined in terms of the multivariate normal distribution. The multivariate normal distribution can have any covariance matrix and any mean vector. Probabilities defined by a set of inequalities of linear combinations of the multivariate normal random variables are considered. It is shown how these probabilities can be evaluated from a series of one-dimensional integrations. This approach affords a practical algorithm for the evaluation of these probabilities which is considerably more efficient than more direct numerical integration approaches. Consequently, it enlarges the class of probabilities of this kind which are computationally feasible.

AMS Subject Classification: 62H99.

Key-words: Multivariate normal distribution; Recursive integration; Numerical integration; Computational intensity; Tree structure; Tri-diagonal matrix; Orthant probability.

1. Introduction

Let $\phi_k(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ represent the probability density function of the random variables $\boldsymbol{X} = (X_1,\ldots,X_k)$ which have a k-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and positive-definite covariance matrix $\boldsymbol{\Sigma}$. This article addresses the evaluation of probabilities defined by a set of inequalities of linear combinations of $\boldsymbol{X}$. That is, for $c_{ij} \in \Re$, $1 \le i \le n$, $1 \le j \le k+1$, attention is directed towards the evaluation of

$$P(c_{i1}X_1 + \ldots + c_{ik}X_k \le c_{i,k+1}; 1 \le i \le n) = \int_S \cdots \int \phi_k(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})d\boldsymbol{x} \qquad (1.1)$$

where the set S is defined as

$$S = \{\boldsymbol{x} : c_{i1}x_1 + \ldots + c_{ik}x_k \le c_{i,k+1}; 1 \le i \le n\}.$$

* 1559-8608/11-4/$5 + $1pp - see inside front cover

It can be assumed that $k \leq n$, since otherwise the required probability can be expressed in terms of a multivariate normal distribution with a smaller dimension.

It can be seen that by employing a linear transformation of $\boldsymbol{X}$ these probabilities can be expressed with different c_{ij}, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and specifically $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_k$, the k-dimensional identity matrix, may be assumed. Notice also that for the purpose of this representation, a two-sided bound on a linear combination of the random variables $\boldsymbol{X}$ can be written as two one-sided inequalities.

While the evaluation of equation (1.1) ostensibly requires the evaluation of a k-dimensional integral expression, it is shown how it can actually be performed by a series of recursive one-dimensional integral calculations. The number of these one-dimensional integral calculations depends upon the dimension k of the multivariate normal distribution and the number of inequalities n that determines the integration region. Whenever the number of one-dimensional integral calculations required is not exorbitant, this approach affords a practical algorithm for the numerical integration of (1.1). Furthermore, it enlarges the set of probabilities of the kind (1.1) whose evaluation is computationally feasible, in comparison to more direct numerical integration approaches to the k-dimensional integral expression.

Recursive relationships have been used in many areas of probability and statistics and it has been recognized that they can provide a very useful analysis tool, as discussed in Hayter (2006). Recursive integration methods have been used in sequential analysis, where the analysis of the performance properties of sequential testing procedures involves the calculation of probabilities over a large number of potential stages of the sequential process. These probabilities can be defined recursively from one stage to another as illustrated in Armitage, McPherson and Rowe (1969) and Jennison and Turnbull (1991), for example. Recursive integration has also been applied to critical point and power calculations for one-way layouts, as in Hayter and Liu (1996a) for example.

The basic idea of recursive integration is that a high dimensional integral expression can be evaluated by a (relatively few) series of low dimensional integral evaluations. An important property of the recursive integration procedure is the actual dimension of the low dimensional integrations which are required. The procedure described here is actually a one-dimensional recursive integration since each step has the computational intensity of a one-dimensional integral calculation. This is sufficient for the evaluation of the probability in equation (1.1).

A special case of equation (1.1) is addressed in Miwa, Hayter and Kuriki (2003) where it is shown that a k-dimensional orthant probability for a general multivariate normal distribution can be expressed as no more than $(k-1)!$ orthant probabilities for multivariate normal distributions with a tri-diagonal covariance matrix. Recursive methodologies are then used to evaluate each of these special orthant probabilities. In that paper a geometrical approach was employed to decompose the orthant region.

In addition, a lot of research has been done on the related problem of calculations for multivariate t distributions, particularly when the inequalities of interest involve contrasts of the means. See, for example, Bretz, Genz and Hothorn (2001), Genz and Bretz (1999, 2002) and Somerville (1997, 1998). Genz and Bretz (2009) also provides an excellent review on the computation of multivariate normal and t probabilities. However, for the general problem considered in this paper, most of the methods proposed in the literature require some element of simulation, and the objective of this paper is to illustrate the usefulness of numerical integration techniques that exploit the potential of recursive methodologies.

The recursive integration methodology is introduced in section 2 with reference to probabilities with $k = 2$, and the recursive integration methodology is discussed more generally in section 3. Section 4 contains the main result employed for dimensional reduction, and its application to the evaluation of (1.1) is discussed in section 5. Finally, section 6 contains some examples of the implementation of this approach.

2. The evaluation of two-dimensional normal probabilities

In this section the case of a two-dimensional multivariate normal distribution, $k = 2$, is discussed. Consider the probability

$$P(\boldsymbol{X} = (X_1, X_2) \in S) = \int_S \int \phi_2(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{x} \tag{2.1}$$

where S is any subset of $\Re^2$. The random variables $\boldsymbol{X} = (X_1, X_2)$ can be expressed as linear combinations of two independent standard normal random variables $\boldsymbol{Y} = (Y_1, Y_2)$, and so the probability (2.1) can be written as

$$\int_{S'} \int \phi_2(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{I}_2) d\boldsymbol{y} \tag{2.2}$$

where S' is the set S after transforming from $\boldsymbol{X}$ to $\boldsymbol{Y}$.

Suppose that the set S' can be expressed as

$$S' = \{(y_1, y_2) : l \le y_1 \le u, l(y_1) \le y_2 \le u(y_1)\} \tag{2.3}$$

for some (possibly infinite) lower and upper bounds l and u for y_1, and bounds $l(y_1)$ and $u(y_1)$ for y_2 which may depend upon y_1. If the set S' cannot be directly expressed in this form, then the method discussed here is applicable to sets S' that can be decomposed into sets which can each be expressed in the form of equation (2.3), or alternatively, for which the bounds for y_2 in the set S' can be formed as the union of disjoint intervals. Then if $\phi(x)$ and $\Phi(x)$ are respectively the probability density function and cumulative distribution function of a standard normal distribution, the representation (2.3) can be used to express the probability (2.2) as

$$\int_{y_1=l}^{u} \phi(y_1) \left[\Phi(u(y_1)) - \Phi(l(y_1))\right] dy_1. \tag{2.4}$$

If the set S' cannot be directly expressed in the form of equation (2.3), then the required probability can be expressed as a sum of terms of the form given in expression (2.4).

The evaluation of expression (2.4) essentially requires two one-dimensional integral evaluations. This is because the calculation of the function $\Phi(x)$ (if it is unknown) requires a one-dimensional integral evaluation, and once it is known and stored, the subsequent evaluation of (2.4) requires an additional one-dimensional integration of the variable y_1. The key point here is that for the numerical integration of expression (2.4), the function $\Phi(x)$ needs to be evaluated for a certain fine grid of points over the real line. However, these evaluations

only need to be performed once, and stored. Once this has been accomplished, the function $\Phi(x)$ can then be considered to be "known" and it is available for future use. Specifically, it can then be used to evaluate expression (2.4) as a one-dimensional integral evaluation.

This benefit in dimensional reduction, from an ostensibly two-dimensional integral evaluation in expression (2.1) to expression (2.4) which requires two one-dimensional integral evaluations, occurs because the integrand in expression (2.2) factors into separate functions for y_1 and y_2. In general, if the integrand $f(y_1, y_2)$ does not factor, then the dimensional reduction is not possible because for each value of y_1, the evaluation of

$$\int_{y_2=l(y_1)}^{u(y_1)} f(y_1, y_2) dy_2 \tag{2.5}$$

is required, and this cannot be obtained by referring to any one-dimensional function. Consequently, expression (2.5) needs to be evaluated by a new one-dimensional integral calculation for each value of y_1, so that the overall computational intensity is that of a two-dimensional integral.

This two-dimensional example illustrates the principle of the recursive integration methodology for the evaluation of high dimensional integral expressions. If the dependence of the integral expression on one of the variables can be isolated to just a low dimensional function, then the evaluation of that function allows a dimensional reduction in the integral. If this process can be repeated in a recursive fashion, then the high dimensional integral expression can be evaluated as a series of low dimensional integral evaluations. This recursive integration methodology is discussed in more detail in section 3.

Finally, it should be emphasized that the discussion in this section with $k = 2$ allows the set S to have any shape. In section 5, the evaluation of the probability (1.1) for higher dimensions $k \geq 3$ is restricted to sets S defined by linear inequalities of the random variables $\boldsymbol{X}$, as indicated in the introduction.

3. Recursive Integration

In this section a recursive methodology is described for the evaluation of a certain class of multi-dimensional integrals.

3.1. Description of the recursive integration methodology

Consider a set of random variables arranged in a tree structure. Let X_1 be the root, with branches $X_{11}, \ldots, X_{1n_1}$, and in general let the random variable $X_{i_1 \ldots i_r}$ have branches $X_{i_1 \ldots i_r j}$, $1 \leq j \leq n_{i_1 \ldots i_r}$. Suppose that the total number of random variables Γ is finite, and that the Γ random variables are independent, although each may have any distribution on the real line, not necessarily identical.

Consider the probability $P(S)$ that the Γ random variables lie in a subset S of $\Re^\Gamma$, where S is defined by

$$X_1 \in I_1 \subseteq \Re$$

and

$$X_{i_1 \ldots i_r} \in I_{i_1 \ldots i_r}(X_{i_1 \ldots i_{r-1}}) \tag{3.1}$$

for all of the other random variables. In other words, the set S can be specified by a region for the root random variable X_1, together with a region for every other random variable which respectively depends only on the value of the random variable from which it is branched. Attention is directed towards regions $I_{i_1\ldots i_r}$ that can be expressed as a finite union of intervals on the real line.

The evaluation of the probability $P(S)$ ostensibly involves a Γ-dimensional integral calculation, but by using recursive integration it can actually be evaluated as a series of one-dimensional integral calculations. In order to show how this can be accomplished, it is helpful to define the function $G_{i_1\ldots i_r}(x)$ to be the probability that the random variable $X_{i_1\ldots i_r}$ lies in the interval $(-\infty,x)$, and that all of the random variables $X_{i_1\ldots i_r\ldots}$ which are derived from $X_{i_1\ldots i_r}$ through one or more branches (these are all of the random variables in the sub-tree which has $X_{i_1\ldots i_r}$ as its root) satisfy the restrictions placed upon them by the set S.

Suppose that a random variable which has no subsequent branches is described as a terminating random variable. Consider then a random variable $X_{i_1\ldots i_r}$ which is not a terminating random variable, and which has branches $X_{i_1\ldots i_r j}$, $j \in T_{i_1\ldots i_r}$, which are terminating branches, and branches $X_{i_1\ldots i_r j}$, $j \in T^*_{i_1\ldots i_r}$, which are not terminating branches. Here $T_{i_1\ldots i_r} \cup T^*_{i_1\ldots i_r} = \{1,\ldots,n_{i_1\ldots i_r}\}$, with $n_{i_1\ldots i_r} \geq 1$, and either the set $T_{i_1\ldots i_r}$ or the set $T^*_{i_1\ldots i_r}$ may be empty.

If the set $T^*_{i_1\ldots i_r}$ is empty, then the random variable $X_{i_1\ldots i_r}$ has branches $X_{i_1\ldots i_r j}$, $1 \leq j \leq n_{i_1\ldots i_r}$, none of which have any subsequent branches. In this case the first step is to calculate and store the cumulative distribution functions $F_{i_1\ldots i_r j}$, $1 \leq j \leq n_{i_1\ldots i_r}$, of each of these terminating random variables. Next, the function

$$G_{i_1\ldots i_r}(x) = \int_{y=-\infty}^{x} f_{i_1\ldots i_r}(y) \left(\prod_{j=1}^{n_{i_1\ldots i_r}} P(X_{i_1\ldots i_r j} \in I_{i_1\ldots i_r j}(y)) \right) dy \tag{3.2}$$

can be calculated and stored, where $f_{i_1\ldots i_r}(x)$ is the probability density function of the random variable $X_{i_1\ldots i_r}$. In the evaluation of this function $G_{i_1\ldots i_r}(x)$, the cumulative distribution functions $F_{i_1\ldots i_r j}$, $1 \leq j \leq n_{i_1\ldots i_r}$, can be referred to for the calculations of $P(X_{i_1\ldots i_r j} \in I_{i_1\ldots i_r j}(y))$, $1 \leq j \leq n_{i_1\ldots i_r}$. Notice that, as required, the function $G_{i_1\ldots i_r}(x)$ represents the probability that the random variable $X_{i_1\ldots i_r}$ lies in the interval $(-\infty,x)$, and that all of the random variables $X_{i_1\ldots i_r j}$, $1 \leq j \leq n_{i_1\ldots i_r}$, satisfy the restrictions placed upon them by the set S.

Now more generally, consider a random variable $X_{i_1\ldots i_r}$ for which the set $T^*_{i_1\ldots i_r}$ is not empty, and for which the functions $G_{i_1\ldots i_r j}(x)$, $j \in T^*_{i_1\ldots i_r}$, have been calculated and stored. In this case the first step is again to calculate and store the cumulative distribution functions $F_{i_1\ldots i_r j}$, $j \in T_{i_1\ldots i_r}$, of the terminating random variables. Next, the function

$$G_{i_1\ldots i_r}(x) = \int_{y=-\infty}^{x} f_{i_1\ldots i_r}(y) \left(\prod_{j\in T_{i_1\ldots i_r}} P(X_{i_1\ldots i_r j} \in I_{i_1\ldots i_r j}(y)) \right) \left(\prod_{j\in T^*_{i_1\ldots i_r}} H_{i_1\ldots i_r j}(y) \right) dy \tag{3.3}$$

can be calculated and stored.

In the calculation of expression (3.3) the cumulative distribution functions $F_{i_1 \ldots i_r j}$, $j \in T_{i_1 \ldots i_r}$, can be referred to in order to calculate the probabilities $P(X_{i_1 \ldots i_r j} \in I_{i_1 \ldots i_r j}(y))$, $j \in T_{i_1 \ldots i_r}$. In addition, the function $H_{i_1 \ldots i_r j}(y)$ is defined to be the probability that the random variable $X_{i_1 \ldots i_r j}$ lies in the interval $I_{i_1 \ldots i_r j}(y)$ and that all of the random variables $X_{i_1 \ldots i_r j \ldots}$ which are derived from $X_{i_1 \ldots i_r j}$ through one or more branches satisfy the restrictions placed on them by the set S. However, $H_{i_1 \ldots i_r j}(y)$ can be found from the interval $I_{i_1 \ldots i_r j}(y)$ in conjunction with reference to the function $G_{i_1 \ldots i_r j}(x)$, since if

$$I_{i_1 \ldots i_r j}(y) = \bigcup_{i=1}^{n(y)} (l_i(y), u_i(y))$$

say, where the intervals $(l_i(y), u_i(y))$ are disjoint (and the number of intervals $n(y)$ may or may not depend upon the value y), then

$$H_{i_1 \ldots i_r j}(y) = \sum_{i=1}^{n(y)} \left[G_{i_1 \ldots i_r j}(u_i(y)) - G_{i_1 \ldots i_r j}(l_i(y)) \right].$$

Again it can be seen that the function $G_{i_1 \ldots i_r}(x)$ calculated from expression (3.3) meets its required definition, which is that it represents the probability that the random variable $X_{i_1 \ldots i_r}$ lies in the interval $(-\infty, x)$, and that all of the random variables $X_{i_1 \ldots i_r \ldots}$ which are derived from $X_{i_1 \ldots i_r}$ through one or more branches satisfy the restrictions placed upon them by the set S.

In this manner the functions $G_{i_1 \ldots i_r}(x)$ can be calculated recursively for each of the non-terminating random variables, moving backwards from the periphery of the tree towards the root. Thus, if the tree is considered to grow away from the root, the recursion is applied in an opposite direction to the growth. Eventually, it will be possible to calculate and store all of the functions $G_{1j}(x)$, $j \in T_1^*$. If the cumulative distribution functions $F_{1j}(x)$ of each of the terminating random variables X_{1j}, $j \in T_1$, are then calculated and stored, the required probability can be calculated as

$$P(S) = \int_{y \in I_1} f_1(y) \left(\prod_{j \in T_1} P(X_{1j} \in I_{1j}(y)) \right) \left(\prod_{j \in T_1^*} H_{1j}(y) \right) dy \tag{3.4}$$

where the cumulative distribution functions $F_{1j}(x)$ are referred to for the evaluation of $P(X_{1j} \in I_{1j}(y))$, $j \in T_1$, and the functions $G_{1j}(x)$ and regions $I_{1j}(y)$ are referred to for the evaluation of $H_{1j}(y)$, $j \in T_1^*$.

It is important to realize that each step in this calculation of $P(S)$ has the computational intensity of a one-dimensional integral evaluation. The last step using expression (3.4) is clearly a one-dimensional integral computation, and every other step can be considered to be the calculation and storage of a function $F(x)$ from a known function $f(x)$ with

$$F(x) = \int_{y=-\infty}^{x} f(y) dy.$$

For terminating random variables this is the calculation of the random variable's cumu-

lative distribution function from its probability density function, which may be unnecessary if the cumulative distribution function is already stored or is otherwise available. For non-terminating random variables $F(x)$ is the function $G(x)$ where the function $f(y)$ is the integrand of expression (3.2) or (3.3) which can be obtained from the probability density function of the random variable and the previously stored functions.

The function $F(x)$ needs to be calculated and stored over a suitable grid of points on $\Re$, so that any values of the function can be obtained from the stored values with sufficient accuracy, using an interpolation method if necessary. If $x_i < x_{i+1}$ are adjacent grid points, then an elementary way to do this is to calculate the value of $F(x_{i+1})$ from the value of $F(x_i)$ using the relationship

$$F(x_{i+1}) \simeq F(x_i) + (x_{i+1} - x_i)\frac{f(x_i) + f(x_{i+1})}{2}.$$

More sophisticated methods can be used to calculate and store the function $F(x)$, but the essential point is that the computational intensity will be linearly related to the number of grid points employed.

Consequently, the evaluation of the probability $P(S)$ is achieved through a series of at most Γ steps, where the computational intensity of each step is linearly related to the number of grid points required to store a one-dimensional function in such a way that the function can be derived from these stored values with sufficient accuracy. In other words, the evaluation of the probability $P(S)$ is achieved with at most Γ steps, where each step has the computational intensity of a one-dimensional numerical integration calculation. Notice that the computational intensity of this recursive integration calculation is linearly related to the number of random variables Γ, whereas more generally a probability calculation would have a computational intensity that grows exponentially with the number of random variables involved.

The recursive integration method described here is applicable because the random variables are independent and because the integration region S has the specified form (3.1). If either of these two conditions are violated, then the method cannot be applied as described. However, it is possible to define more general recursive integration methodologies, where the computational intensities of the steps are those of a p-dimensional integral calculation, say. Therefore, if the method described here with $p = 1$ is not applicable to a certain problem, it may still be helpful to employ a recursive integration approach with a higher value of p. Hayter (2006) provides information about recursive integration for general values of p. In this article attention is restricted to the recursive integration methodology with $p = 1$ described in this section, since this is sufficient for the evaluation of the probability in equation (1.1).

The discussion in this section has been based on the assumption that the random variables are continuous. However, the recursive integration methodology applies in a similar fashion to a set of discrete random variables, and to a mixture of discrete and continuous random variables. For discrete random variables the integrations are replaced by sums.

Finally, it is important to notice that the region $X_{i_1\ldots i_r} \in I_{i_1\ldots i_r}(X_{i_1\ldots i_{r-1}})$ can also be considered to provide a region for $X_{i_1\ldots i_{r-1}}$ which depends upon the value of $X_{i_1\ldots i_r}$, so that the growth direction of the branches can be reversed. Thus, the set S may actually be thought of as being regions $(X_i, X_j) \in S_{ij}$ for every branch in the tree, where X_i and X_j are the ran-

dom variables at the two ends of the branch. Therefore, any one of the Γ random variables may actually be designated as the root random variable, with the tree growing away from that random variable. For a given problem, some choices of the root random variable may provide a recursive integration which is simpler than that provided by other choices.

3.2. Applications of the recursive integration methodology

Probability calculations for two independent random variables can be performed using this recursive integration approach since the set S necessarily satisfies the required condition (3.1). That is why an expression such as (2.4) can be used whenever the integrand in equation (2.2) factors into separate functions for y_1 and y_2, as discussed in section 2. In this case there is a root random variable with one terminating branch.

The subsequent sections of this article consider the issue of how the probability of the multivariate normal distribution given in equation (1.1) can be evaluated by recursive integration. In certain cases the probability can be expressed directly through a set of independent normal random variables with a tree structure and appropriate integration regions, so that the recursive integration method can be directly applied. However, more generally, the probability in equation (1.1) has to be decomposed into terms to which the recursive integration method can be applied.

A well known example (see Tong (1990), page 192, for example) where the recursive integration method can be directly applied to the probability in equation (1.1) is when the off-diagonal terms of the covariance matrix Σ can be expressed as $\sigma_{ij} = \lambda_i \lambda_j$, $i \neq j$, for some $\lambda_i \in \Re$, $1 \leq i \leq k$, and where the probability is

$$P(l_i \leq X_i \leq u_i; 1 \leq i \leq k) \tag{3.5}$$

for some (possibly infinite) lower and upper bounds l_i and u_i. If Z is a standard normal random variable, and $Z_i \sim N(\mu_i, \tau_i^2)$, $1 \leq i \leq k$, are independent normal random variables which are also independent of Z, then with $\tau_i^2 = \sigma_i^2 - \lambda_i^2$ where σ_i^2 is the variance of X_i, the random variables $\boldsymbol{X}$ can be expressed as $X_i = \lambda_i Z + Z_i$, $1 \leq i \leq k$. Consequently, the required probability is

$$P(S) = \int_{z=-\infty}^{\infty} \phi(z) \left(\prod_{i=1}^{k} \left[\Phi\left(\frac{u_i - \mu_i - \lambda_i z}{\tau_i} \right) - \Phi\left(\frac{l_i - \mu_i - \lambda_i z}{\tau_i} \right) \right] \right) dz.$$

In this example Z is the root random variable with terminating branches Z_i, $1 \leq i \leq k$. Also, $I_1 = \Re$ and $I_{1i}(Z) = (l_i - \lambda_i Z, u_i - \lambda_i Z)$, $1 \leq i \leq k$. It is also possible to consider Z_1, say, as the root random variable, with one branch Z which has $k-1$ subsequent terminating branches, although this results in a more complicated recursive integration.

Another important example is the probability

$$P(Y_1 \in I_1, Y_2 \in I_2(Y_1), Y_3 \in I_3(Y_2), \ldots, Y_k \in I_k(Y_{k-1})). \tag{3.6}$$

If the random variables Y_i, $1 \leq i \leq k$, are independent then this problem has the form to which the recursive integration methodology can be applied, since it can be considered to be a tree consisting of a root Y_1 and a subsequent chain of branches progressing to a single

terminating random variable Y_k. If $I_1 = \Re$ and $I_i(x) = [x, \infty)$, $2 \leq i \leq k$, the special case

$$P(Y_1 \leq Y_2 \leq Y_3 \leq \ldots \leq Y_k) \tag{3.7}$$

is obtained. The application of recursive integration to the evaluation of the probability (3.7) was explained in Hayter and Liu (1996b).

When evaluating expression (3.6) it may be advantageous to consider a random variable other than Y_1 to be the root random variable. For example, the recursive integration calculation may be more efficient if the random variable Y_2 is considered to be the root variable, in which case Y_2 has one terminating branch Y_1, and another branch Y_3 which progresses to a chain of branches terminating in Y_k. Expression (3.6) with the Y_i being independent normal random variables has particular importance to this article because in subsequent sections it will be shown how the probability in equation (1.1) can be decomposed into such terms. Whenever the number of these terms is not exorbitant, this affords a useful method for the numerical evaluation of (1.1).

If in expression (3.6) the intervals are $I_1 = (l_1, u_1)$ and $I_i(x) = (l_i + x, u_i + x)$, $2 \leq i \leq k$, then the probability is of the form (3.5) where $X_1 = Y_1$ and $X_i = Y_i - Y_{i-1}$, $2 \leq i \leq k$. If the random variables $\boldsymbol{Y}$ are independent normal random variables, then the random variables $\boldsymbol{X}$ have a tri-diagonal covariance matrix (the off-diagonal terms σ_{ij} of the covariance matrix are zero for $|i - j| \geq 2$). This process can be reversed, so that whenever the covariance matrix $\boldsymbol{\Sigma}$ is tri-diagonal and the integration region is a rectangular box as in (3.5), the probability can be written in the form of expression (3.6) with the Y_i being independent normal random variables, so that it can be evaluated by the recursive integration method. This was shown in Miwa, Hayter and Kuriki (2003).

Finally, consider a general tree of random variables $X_{i_1 \ldots i_r}$ which are independent normal random variables, and suppose that the regions defining the set S are $I_1 = (l_1, u_1)$ and $I_{i_1 \ldots i_r}(x) = (l_{i_1 \ldots i_r} + c_{i_1 \ldots i_r} x, u_{i_1 \ldots i_r} + c_{i_1 \ldots i_r} x)$, for real numbers $l_{i_1 \ldots i_r}$, $u_{i_1 \ldots i_r}$ and $c_{i_1 \ldots i_r}$. Then consider a set of random variables $\boldsymbol{X}$ which are the root random variable X_1, together with additional random variables X_i, $2 \leq i \leq \Gamma$, which are obtained as $X_{i_1 \ldots i_r} - c_{i_1 \ldots i_r} X_{i_1 \ldots i_{r-1}}$ for all of the other random variables $X_{i_1 \ldots i_r}$ in the tree. In other words, each branch of the tree generates a random variable X_i, $2 \leq i \leq \Gamma$, which is obtained as this linear combination of the two random variables at the ends of the branch. This generates a rectangular box type probability as in (3.5) where $\boldsymbol{X}$ has a multivariate normal distribution, and which can be evaluated directly by the recursive integration method.

The covariance matrix of these random variables $\boldsymbol{X}$ is a combination of elements with a product form $\lambda_i \lambda_j$ corresponding to the original tree random variables which are branched from a common random variable, and elements with a "tri-diagonal form" corresponding to the original tree random variables which are chained together along a series of branches. Of course, in general it is not possible to put the probability in equation (1.1) into this format, so that the recursive integration method cannot be applied directly. Nevertheless, the probability in equation (1.1) can be decomposed into terms of this form.

4. Dimensional reduction for normal probabilities

In this section a dimensional reduction procedure is described for multi-dimensional integrals of a multivariate normal probability density function. This procedure can be employed

to decompose the probability in equation (1.1) into terms to which the recursive integration methodology of section 3 can be applied.

If $g : \Re \to \Re$ is an integrable function and $S \subseteq \Re^{r+1}$, define

$$T_{r+1}(S, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \cdots \int_{(\boldsymbol{x}, y) \in S} \phi_r(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, g(y)\, dy\, d\boldsymbol{x}. \tag{4.1}$$

In addition, let the set $R \subseteq \Re^{r+1}$ be given by

$$R = \{(\boldsymbol{x}, y) \in \Re^{r+1} : \boldsymbol{x} \in Q;\ y \leq \boldsymbol{c}_i \boldsymbol{x} + b_i, i \in I_1;\ \boldsymbol{c}_i \boldsymbol{x} + a_i \leq y, i \in I_2;$$

$$\boldsymbol{c}_i \boldsymbol{x} + a_i \leq y \leq \boldsymbol{c}_i \boldsymbol{x} + b_i, i \in I_3\} \tag{4.2}$$

where $Q \subseteq \Re^r$ is a non-empty set, and the sets I_1, I_2 and I_3 are disjoint. Also, the a_i, $i \in I_2 \cup I_3$, and b_i, $i \in I_1 \cup I_3$, are any (finite) real numbers, with $a_i < b_i$, $i \in I_3$. Finally, $\boldsymbol{c}_i \in \Re^r$, $i \in I_1 \cup I_2 \cup I_3$, with $\boldsymbol{c}_i \neq \boldsymbol{c}_j$ for $i \neq j$ (and with these vectors not being a multiple of each other). However, note that it is possible for one of the $\boldsymbol{c}_i$ to be equal to $\mathbf{0}$.

The initial objective of this section is to show how the integral over y can be performed for $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, resulting in expressions of the form $T_r(R^*, g^*, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ where the integral dimension has been reduced by one. A subsequent objective will be to show how this dimension reduction can be applied repeatedly.

4.1. Decomposition of the integration region R

In general, the integration region for y in the evaluation of $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ depends on the value of $\boldsymbol{x}$ through a series of linear combinations $\boldsymbol{c}_i \boldsymbol{x}$. It is useful to be able to express $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ as a linear combination of terms for which the integration region for y depends on the value of $\boldsymbol{x}$ only through a single linear combination. This can be accomplished by defining

$$A_i(\delta) = \{\boldsymbol{x} :\ \boldsymbol{x} \in Q;\ \boldsymbol{c}_{j_2} \boldsymbol{x} + a_{j_2} \leq \boldsymbol{c}_i \boldsymbol{x} + \delta \leq \boldsymbol{c}_{j_1} \boldsymbol{x} + b_{j_1},$$

$$j_1 \in I_1 \cup I_3,\ j_2 \in I_2 \cup I_3\} \tag{4.3}$$

and

$$G_{i_1 i_2} = \{\boldsymbol{x} :\ \boldsymbol{x} \in Q;\ \boldsymbol{c}_{j_2} \boldsymbol{x} + a_{j_2} \leq \boldsymbol{c}_{i_2} \boldsymbol{x} + a_{i_2} \leq \boldsymbol{c}_{i_1} \boldsymbol{x} + b_{i_1} \leq \boldsymbol{c}_{j_1} \boldsymbol{x} + b_{j_1},$$

$$j_1 \in I_1 \cup I_3,\ j_2 \in I_2 \cup I_3\} \tag{4.4}$$

for $i_1 \in I_1 \cup I_3$ and $i_2 \in I_2 \cup I_3$. The sets $A_i(\delta)$ will be empty for certain values of δ, and some of the sets $G_{i_1 i_2}$ may be empty.

If $I_2 \cup I_3 = \emptyset$ and $I_1 \neq \emptyset$ it follows that

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i\in I_1} T_r(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(b_i),\ y \leq \boldsymbol{c}_i\boldsymbol{x}+b_i\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{4.5}$$

and if $I_1 \cup I_3 = \emptyset$ and $I_2 \neq \emptyset$ it follows that

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i\in I_2} T_r(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(a_i),\ \boldsymbol{c}_i\boldsymbol{x}+a_i \leq y\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}). \tag{4.6}$$

Furthermore, if $I_1 \cup I_3 \neq \emptyset$ and $I_2 \cup I_3 \neq \emptyset$ it follows that

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) =$$

$$\sum_{i_1\in I_1\cup I_3}\ \sum_{i_2\in I_2\cup I_3} T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2},\ \boldsymbol{c}_{i_2}\boldsymbol{x}+a_{i_2} \leq y \leq \boldsymbol{c}_{i_1}\boldsymbol{x}+b_{i_1}\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{4.7}$$

where $T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2},\ \boldsymbol{c}_{i_2}\boldsymbol{x}+a_{i_2} \leq y \leq \boldsymbol{c}_{i_1}\boldsymbol{x}+b_{i_1}\},g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ can be expressed as either

$$T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2}, y \leq \boldsymbol{c}_{i_1}\boldsymbol{x}+b_{i_1}\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}) - T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2},\ y \leq \boldsymbol{c}_{i_2}\boldsymbol{x}+a_{i_2}\},g,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

or as

$$T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2}, \boldsymbol{c}_{i_2}\boldsymbol{x}+a_{i_2} \leq y\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}) - T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in G_{i_1i_2},\ \boldsymbol{c}_{i_1}\boldsymbol{x}+b_{i_1} \leq y\},g,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

and the first of these two expressions gives

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i\in I_1\cup I_3} T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(b_i),\ y \leq \boldsymbol{c}_i\boldsymbol{x}+b_i\},g,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$$- \sum_{i\in I_2\cup I_3} T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(a_i),\ y \leq \boldsymbol{c}_i\boldsymbol{x}+a_i\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{4.8}$$

while the second expression gives

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i\in I_2\cup I_3} T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(a_i),\ \boldsymbol{c}_i\boldsymbol{x}+a_i \leq y\},g,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$$- \sum_{i\in I_1\cup I_3} T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in A_i(b_i),\ \boldsymbol{c}_i\boldsymbol{x}+b_i \leq y\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}). \tag{4.9}$$

However, the special case $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$ can be treated differently, since in this case

$$T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma}) = T_{r+1}(\{(\boldsymbol{x},y) : \boldsymbol{x} \in Q,\ \boldsymbol{c}_1\boldsymbol{x}+a_1 \leq y \leq \boldsymbol{c}_1\boldsymbol{x}+b_1\},g,\boldsymbol{\mu},\boldsymbol{\Sigma}). \tag{4.10}$$

Notice, therefore, that whenever $I_1 \cup I_2 \cup I_3 \neq \emptyset$, equations (4.5), (4.6) and either (4.8), (4.9) or (4.10) can be used to express $T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ with terms $T_{r+1}(R',g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ where in the set R', the region for y depends on $\boldsymbol{x}$ only through a single linear combination. Also,

it should be mentioned that as long as none of the inequalities for $i \in I_1 \cup I_2 \cup I_3$ in the definition of the set R is redundant, then none of the integration regions R' for the terms in the decomposition of $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is empty.

4.2. Integrating the y variable

Now that the dependence of the integration region on y has been isolated to a single inequality, the y variable can be integrated in a straightforward manner. There are two cases to consider, relating to whether or not the integration region for y depends on $\boldsymbol{x}$.

The first case is when the integration region for y does not depend on $\boldsymbol{x}$. This occurs when $I_1 \cup I_2 \cup I_3 = \emptyset$, and also for a set R' with $\boldsymbol{c}_i = \boldsymbol{0}$. For these situations let Z and W_i, $1 \leq i \leq r-1$, be linear combinations of the random variables $\boldsymbol{X} = (X_1, \ldots, X_r) \sim N_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $\boldsymbol{W} = (W_1, \ldots, W_{r-1}) \sim N_{r-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ for a positive-definite covariance matrix Σ^*, and $Z \sim N(\beta, \tau^2)$ with Z independent of $\boldsymbol{W}$. For example, if $\boldsymbol{\Sigma}$ has variance terms σ_i^2 and covariance terms σ_{ij}, let $Z = X_r$ and $W_i = X_i - \sigma_{ir} X_r / \sigma_r^2$, $1 \leq i \leq r-1$.

If $I_1 \cup I_2 \cup I_3 = \emptyset$ then

$$T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \cdots \int_{(\boldsymbol{w}, z) \in R^*} \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \left(\phi_1(z; \beta, \tau^2) \int_{y=-\infty}^{\infty} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.11}$$

where R^* is the set Q expressed in terms of the variables $(\boldsymbol{w}, z)$ instead of $\boldsymbol{x}$. Also,

$$T_{r+1}(A_i(\delta) \cap \{y : y \leq \delta\}, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \cdots \int_{(\boldsymbol{w}, z) \in R^*} \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \left(\phi_1(z; \beta, \tau^2) \int_{y=-\infty}^{\delta} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.12}$$

and

$$T_{r+1}(A_i(\delta) \cap \{y : \delta \leq y\}, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \cdots \int_{(\boldsymbol{w}, z) \in R^*} \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \left(\phi_1(z; \beta, \tau^2) \int_{y=\delta}^{\infty} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.13}$$

where here R^* is the set $A_i(\delta)$ expressed in terms of the variables $(\boldsymbol{w}, z)$ instead of $\boldsymbol{x}$. Finally,

$$T_{r+1}(Q \cap \{(\boldsymbol{x},y) : \delta_1 \le y \le \delta_2\}, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \underset{(\boldsymbol{w},z) \in R^*}{\ldots} \int \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$$

$$\left(\phi_1(z; \beta, \tau^2) \int_{y=\delta_1}^{\delta_2} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.14}$$

where again R^* is the set Q expressed in terms of the variables $(\boldsymbol{w}, z)$ instead of $\boldsymbol{x}$.

The second case to consider is when the integration region for y does depend on $\boldsymbol{x}$. For sets R' of the form $A_i(\delta) \cap \{(\boldsymbol{x},y) : y \le \boldsymbol{c}_i\boldsymbol{x} + \delta\}$ and $A_i(\delta) \cap \{(\boldsymbol{x},y) : \boldsymbol{c}_i\boldsymbol{x} + \delta \le y\}$ where $\boldsymbol{c}_i \ne \boldsymbol{0}$, let $Z = \boldsymbol{c}_i\boldsymbol{X} + \delta$ which is distributed $N(\beta, \tau^2)$, say, and again let $\boldsymbol{W} = (W_1, \ldots, W_{r-1})$ be linear combinations of the random variables $\boldsymbol{X} \sim N_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $\boldsymbol{W} \sim N_{r-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is independent of Z with a positive-definite covariance matrix Σ^*. This can be achieved by taking $W_i = o_{i1}X_1 + \ldots + o_{ir}X_r$, $1 \le i \le r-1$, where the vectors $\boldsymbol{o}_i = (o_{i1}, \ldots, o_{ir})$, $1 \le i \le r-1$, span the $r-1$ dimensional space orthogonal to $\boldsymbol{c}_i\boldsymbol{\Sigma}$.

Consequently

$$T_{r+1}(A_i(\delta) \cap \{(\boldsymbol{x},y) : y \le \boldsymbol{c}_i\boldsymbol{x} + \delta\}, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \underset{(\boldsymbol{w},z) \in R^*}{\ldots} \int \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$$

$$\left(\phi_1(z; \beta, \tau^2) \int_{y=-\infty}^{z} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.15}$$

and

$$T_{r+1}(A_i(\delta) \cap \{(\boldsymbol{x},y) : \boldsymbol{c}_i\boldsymbol{x} + \delta \le y\}, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \underset{(\boldsymbol{w},z) \in R^*}{\ldots} \int \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$$

$$\left(\phi_1(z; \beta, \tau^2) \int_{y=z}^{\infty} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.16}$$

where again R^* is the set $A_i(\delta)$ expressed in terms of the variables $(\boldsymbol{w}, z)$ instead of $\boldsymbol{x}$. Also, if $R' = Q \cap \{(\boldsymbol{x},y) : \boldsymbol{c}_1\boldsymbol{x} + \delta_1 \le y \le \boldsymbol{c}_1\boldsymbol{x} + \delta_2\}$ define $Z = \boldsymbol{c}_1\boldsymbol{X}$, with $\boldsymbol{W}$ again chosen to be independent of Z, so that

$$T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \underset{(\boldsymbol{w},z) \in R^*}{\ldots} \int \phi_{r-1}(\boldsymbol{w}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$$

$$\left(\phi_1(z; \beta, \tau^2) \int_{y=z+\delta_1}^{z+\delta_2} g(y) dy \right) dz \, d\boldsymbol{w} \tag{4.17}$$

where R^* is the set Q expressed in terms of the variables $(\boldsymbol{w}, z)$ instead of $\boldsymbol{x}$.

In summary, the $r+1$ dimensional expression $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ has been expressed with r dimensional terms $T_r(R^*, g^*, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, where the function $g^*(z)$ is defined to be

$$g^*(z) = \phi_1(z; \beta, \tau^2) \int_{y \in D(z)} g(y)dy \tag{4.18}$$

where $D(z)$ is either $(-\infty, \infty)$, $(-\infty, \delta)$, (δ, ∞), (δ_1, δ_2), $(-\infty, z)$, (z, ∞), or $(z+\delta_1, z+\delta_2)$, for some constants δ, δ_1 and δ_2, and some mean and variance β and τ^2 of a normal distribution. If $I_1 \cup I_2 \cup I_3 = \emptyset$, equation (4.11) is used. If $I_2 \cup I_3 = \emptyset$ but $I_1 \neq \emptyset$, then equation (4.5) and equations (4.12) and (4.15) are used, and if $I_1 \cup I_3 = \emptyset$ but $I_2 \neq \emptyset$, then equation (4.6) and equations (4.13) and (4.16) are used. Finally, if $I_1 \cup I_3 \neq \emptyset$ and $I_2 \cup I_3 \neq \emptyset$, then either equations (4.8), (4.12) and (4.15) or equations (4.9), (4.13) and (4.16) are used. However, if $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$ then equation (4.10) and equations (4.14) and (4.17) are used.

An important point is that if the function $g(y)$ is known, either from a formula or because it has been stored over a grid of points on the real line, then the computational intensity of calculating and storing the function $g^*(z)$ is no more than that of a one-dimensional integral calculation. In fact, the function $g^*(z)$ can be obtained from the function

$$G(x) = \int_{y=-\infty}^{x} g(y)dy$$

which needs to be calculated and stored over a suitable grid of points on $\Re$. As discussed in section 3, if $x_i < x_{i+1}$ are adjacent grid points, then an elementary way to do this is to calculate the value of $G(x_{i+1})$ from the value of $G(x_i)$ using the relationship

$$G(x_{i+1}) \simeq G(x_i) + (x_{i+1} - x_i)\frac{g(x_i) + g(x_{i+1})}{2}.$$

Once the function $G(x)$ has been calculated and stored, the functions $g^*(z)$ can be obtained from it for each of the terms $T_r(R^*, g^*, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ in the decomposition of $T_{r+1}(R, g, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Consequently, the reduction in the integral dimension from $r+1$ to r has been achieved with only the computational intensity of a one-dimensional integration.

4.3. Repeated dimensional reductions

If the set $Q \subseteq \Re^r$ is defined by inequalities for linear combinations of $\boldsymbol{x}$, then the set $R^* \subseteq \Re^r$ can be expressed as

$$R^* = \{(\boldsymbol{w}, z) \in \Re^r : \boldsymbol{w} \in Q^*;\ z \leq \boldsymbol{c}_i^* \boldsymbol{w} + b_i^*, i \in I_1^*;\ \boldsymbol{c}_i^* \boldsymbol{w} + a_i^* \leq z, i \in I_2^*;$$

$$\boldsymbol{c}_i^* \boldsymbol{w} + a_i^* \leq z \leq \boldsymbol{c}_i^* \boldsymbol{w} + b_i^*, i \in I_3^*\} \tag{4.19}$$

where the set $Q^* \subseteq \Re^{r-1}$ is defined by inequalities for linear combinations of $\boldsymbol{w}$. Therefore once T_{r+1} has been expressed with T_r terms, each of these T_r terms can then be expressed with T_{r-1} terms. This process can be repeated until T_{r+1} is eventually expressed with T_1 terms.

However, it follows from the ideas in section 2 that the computational intensity of eval-

uating $T_2(R,g,\mu,\sigma^2)$ is that of a one-dimensional integral calculation. This is because if $R \subseteq \Re^2$ is of the form given in equation (4.2), it can be expressed as

$$R = \{(x,y) : l \leq y \leq u, l(y) \leq x \leq u(y)\} \tag{4.20}$$

for some functions $l(y)$ and $u(y)$ and values l and u, possibly infinite, so that

$$T_2(R,g,\mu,\sigma^2) = \int_{y=l}^{u} \left[\Phi\left(\frac{u(y)-\mu}{\sigma}\right) - \Phi\left(\frac{l(y)-\mu}{\sigma}\right)\right] g(y)dy.$$

In fact the set R is defined by linear inequalities of (x,y) so that the function $u(y)$ will be the minimum of a set of linear functions of y, and the function $l(y)$ will be the maximum of a set of linear functions of y.

Consequently, there is actually no need to decompose T_2 terms into T_1 terms. For the purpose of attaining an evaluation of T_{r+1} through a series of steps each with the computational intensity of a one-dimensional integral calculation, it is sufficient to express T_{r+1} with T_2 terms. However, an important consideration is the number of terms in the decomposition of T_{r+1}, since this is related to the number of these steps that is required.

Suppose that there are no redundant inequalities in the definition of the set R in equation (4.2). Then if $I_1 \cup I_2 \cup I_3 = \emptyset$, equation (4.11) shows that T_{r+1} is expressed as one T_r term. If $I_2 \cup I_3 = \emptyset$ but $I_1 \neq \emptyset$, then equation (4.5) shows that T_{r+1} is expressed as $|I_1|$ terms T_r, and if $I_1 \cup I_3 = \emptyset$ but $I_2 \neq \emptyset$, then equation (4.6) shows that T_{r+1} is expressed as $|I_2|$ terms T_r. Finally, if $I_1 \cup I_3 \neq \emptyset$ and $I_2 \cup I_3 \neq \emptyset$, then either equation (4.8) or (4.9) shows that T_{r+1} is expressed as $|I_1| + |I_2| + 2|I_3|$ terms T_r (unless $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$, in which case T_{r+1} is expressed with one term T_r).

The number of T_{r-1} terms in the decomposition of each of the T_r terms depends upon the values of I_1^*, I_2^* and I_3^* in the respective sets R^* shown in equation (4.19). Notice that if $I_1 \cup I_2 \cup I_3 = \emptyset$ so that equation (4.11) is used, then the set R^* has the same number of one-sided and two-sided inequalities as the set Q, although each of the inequalities may or may not involve z.

Suppose that the set Q is defined by m_1 one-sided inequalities and m_2 two-sided inequalities, so that the set R is defined by $m_1 + |I_1| + |I_2|$ one-sided inequalities and $m_2 + |I_3|$ two-sided inequalities. If $I_1 \cup I_2 \cup I_3 \neq \emptyset$ the number of inequalities needed to define a set R^* is the number of inequalities needed to define a set $A_i(b_i)$ for $i \in I_1 \cup I_3$, or a set $A_i(a_i)$ for $i \in I_2 \cup I_3$ (or the set Q when $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$). While the definition of the set $A_i(\delta)$ in equation (4.3) ostensibly involves $m_1 + |I_1| + |I_2|$ one-sided inequalities and $m_2 + |I_3|$ two-sided inequalities, some of these inequalities may be redundant.

In particular, if $i \in I_1$ then the inequality with $j_1 = i$ is irrelevant in the definition of $A_i(b_i)$, and if $i \in I_2$ then the inequality with $j_2 = i$ is irrelevant in the definition of $A_i(a_i)$, so that in these cases the sets $A_i(b_i)$ and $A_i(a_i)$ involve at most $m_1 + |I_1| + |I_2| - 1$ one-sided inequalities. Furthermore, if $i \in I_3$ then both the inequalities with $j_1 = i$ and $j_2 = i$ are irrelevant in the definitions of $A_i(b_i)$ and $A_i(a_i)$, so that in these cases the sets $A_i(b_i)$ and $A_i(a_i)$ involve at most $m_2 + |I_3| - 1$ two-sided inequalities. If $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$ then R^* has m_1 one-sided inequalities and m_2 two-sided inequalities.

Therefore, if $I_1 \cup I_2 \cup I_3 \neq \emptyset$ it can be seen that the number of inequalities needed to define a set R^* is always strictly smaller than the number of inequalities needed to define the set R.

However, it should be remembered that the number of terms in the decomposition of T_{r+1} depends upon the number of one-sided and two-sided inequalities in R that depend upon y, while the number of terms in the decomposition of a T_r term depends upon the number of one-sided and two-sided inequalities in R^* that depend upon z. Consequently, it does not necessarily follow that the number of terms in the decomposition of T_{r+1} is strictly larger than the number of terms in each of the subsequent decompositions of the T_r terms.

There are three essential points which enable the repeated dimensional reduction of $T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ with steps which have the computational intensity of a one-dimensional integral calculation. Firstly the integrand of $T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ factors into separate functions $h(\boldsymbol{x})$ and $g(y)$, secondly the integration region R can be decomposed into regions where the variable y is only restricted by one function $f_i(\boldsymbol{x})$ of $\boldsymbol{x}$ (in this case $f_i(\boldsymbol{x})$ is a linear combination of $\boldsymbol{x}$), and thirdly for each of these functions a change of variables can be made from $\boldsymbol{x}$ to $(\boldsymbol{w},z)$ with $z=f_i(\boldsymbol{x})$, such that $h(\boldsymbol{x})$ factors into separate functions for $\boldsymbol{w}$ and z. Each term in this decomposition of $T_{r+1}(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma})$ is intrinsically evaluated by the recursive integration method described in section 3 for the tree structure given in equation (3.6).

5. The Evaluation of multivariate normal probabilities

Consider the probability defined by a set of inequalities of linear combinations of the multivariate normal random variable $\boldsymbol{X}\sim N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$ given in equation (1.1). This probability can be evaluated using the repeated dimensional reduction method described in section 4 because it can be put into the form $T_k(R,g,\boldsymbol{\mu},\boldsymbol{\Sigma})$. A linear transformation can be made from the random variables $\boldsymbol{X}$ to $\boldsymbol{Y}$ such that Y_k is independent of Y_i, $1\le i\le k-1$. The variable y_k can then be the y variable in expression (4.1), and the function $g(y)$ will be the probability density function of Y_k. Since the integration region R is defined by inequalities of linear combinations of $\boldsymbol{Y}$, the required probability T_k can be decomposed all the way to T_2 terms or T_1 terms.

It is helpful to write the probability in equation (1.1) as

$$P(\boldsymbol{c}_i\boldsymbol{X}\le u_i, 1\le i\le n_1;\ l_i\le \boldsymbol{c}_i\boldsymbol{X}\le u_i, n_1+1\le i\le n_1+n_2) \tag{5.1}$$

for $\boldsymbol{c}_i\in\Re^k$, $1\le i\le n_1+n_2$, and (finite) constants u_i and l_i, so that it can be seen to involve n_1 one-sided inequalities and n_2 two-sided inequalities. In equation (5.1) it can be stipulated that no $\boldsymbol{c}_i$ is zero, and that $\boldsymbol{c}_i\ne a\boldsymbol{c}_j$ for $i\ne j$ and $a\in\Re$. It can also be assumed that the space spanned by $\boldsymbol{c}_i$, $1\le i\le n_1+n_2$, has dimension k, since otherwise the probability can be expressed in terms of a multivariate normal distribution with dimension strictly smaller than k. Therefore, it is necessarily the case that $k\le n_1+n_2$.

Suppose that $\boldsymbol{c}_i$, $i\in U$, are contained within a $k-1$ dimensional subspace of $\Re^k$, and such that $\boldsymbol{c}_i$ is not a member of this subspace for $i\notin U$. Also, let $\boldsymbol{e}_i\in\Re^k$, $1\le i\le k-1$, span this subspace and let $\boldsymbol{e}_k\in\Re^k$ be orthogonal to $\boldsymbol{e}_i\boldsymbol{\Sigma}$, $1\le i\le k-1$. If the transformation from the random variables $\boldsymbol{X}$ to the random variables $\boldsymbol{Y}$ is made by defining $Y_i=\boldsymbol{e}_i\boldsymbol{X}$, $1\le i\le k$, it follows that Y_k is independent of Y_i, $1\le i\le k-1$ as required, and also that $|I_1|+|I_2|+|I_3|=n_1+n_2-|U|$.

In order to minimize the number of T_{k-1} terms in the decomposition of T_k, it is therefore generally advantageous to have $|U|$ as large as possible. Specifically, if $\boldsymbol{c}_i$, $i\in U$, consist of

u_1 one-sided inequalities and u_2 two-sided inequalities, the number of T_{k-1} terms is

$$|I_1| + |I_2| + 2|I_3| = n_1 - u_1 + 2n_2 - 2u_2$$

(or one if $n_1 = u_1$ and $n_2 - u_2 = 1$) so that it is generally advantageous to maximize $u_1 + 2u_2$. Of course, in certain cases it is not possible to have $|U| = u_1 + u_2$ larger than $k-1$.

The decompositions of the T_{k-1} terms will in general not afford the luxury of being able to choose the z variable in the representation (4.19) of the integration region. In section 4.2 two cases for integrating the y variable were considered, and in the second case the random variable Z is determined by the integration region for y. However, in the first case there is some latitude in the choice of Z, and the consideration of the number of terms in the subsequent decomposition can be a guide for this choice, in a similar manner to the determination of the optimum choice of the random variables for the initial decomposition of T_k. Nevertheless, in general the choice of the random variables $\boldsymbol{Y}$ which are used for T_k decides the total number of T_2 terms in the complete decomposition.

Suppose that there are only one-sided inequalities in equation (5.1) so that $n_2 = 0$. As a worst case it may not be possible to have $|U|$ larger than $k-1$, so that there will be $n_1 - (k-1)$ terms T_{k-1}. Furthermore, each of the integration regions R^* for the T_{k-1} terms may be defined by $n_1 - 1$ one-sided inequalities which may each involve the variable z, so that each of the T_{k-1} terms may be decomposed into as many as $n_1 - 1$ terms T_{k-2}. Continuing this argument it follows that the maximum number of T_2 terms in the decomposition of a probability defined by n_1 one-sided inequalities of linear combinations of a k-dimensional multivariate normal random variable is

$$(n_1 - (k-1))\,(n_1 - 1)(n_1 - 2)\ldots(n_1 - k + 3) = (n_1 - (k-1))\,\frac{(n_1 - 1)!}{(n_1 - k + 2)!}. \quad (5.2)$$

The maximum number of T_2 terms in the decomposition of an orthant probability for a k-dimensional multivariate normal random variable is obtained by putting $n_1 = k$ in equation (5.2) which gives $(k-1)!/2$.

If there are only two-sided inequalities in equation (5.1) so that $n_1 = 0$, then all of the integration regions throughout the decomposition process only involve two-sided inequalities. Again, as a worst case it may not be possible to have $|U|$ larger than $k-1$, so that there will be $2(n_2 - (k-1))$ terms T_{k-1}. Furthermore, each of the integration regions R^* for the T_{k-1} terms may be defined by $n_2 - 1$ two-sided inequalities which may each involve the variable z, so that each of the T_{k-1} terms may be decomposed into as many as $2(n_2 - 1)$ terms T_{k-2}. It follows from the continuation of this argument that the maximum number of T_2 terms in the decomposition of a probability defined by n_2 two-sided inequalities of linear combinations of a k-dimensional multivariate normal random variable is

$$(2(n_2 - (k-1)))\ 2(n_2 - 1)\ 2(n_1 2 - 2)\ \ldots\ 2(n_2 - k + 3)$$

$$= 2^{k-2}\,(n_2 - (k-1))\,\frac{(n_2 - 1)!}{(n_2 - k + 2)!}. \quad (5.3)$$

If $n_1 = 0$ and $n_2 = k$ in equation (5.1), then a rectangular box (two-sided orthant) probability for a k-dimensional multivariate normal random variable is obtained. The maximum number

of T_2 terms in the decomposition of such a probability is $2^{k-4}(k-1)!$. This is one half of the value obtained by putting $n_2 = k$ in equation (5.3), because the special case $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$ can be applied so that T_k can be expressed with only one T_{k-1} term.

Notice that when $k = 3$ the rectangular box probability (and the orthant probability) only require one T_2 term. This is one of the examples discussed in section 6. If neither of n_1 and n_2 are zero in equation (5.1), then the integration regions in the decomposition process can consist of one-sided and two-sided inequalities. In general it is difficult to calculate how many T_2 terms there will be in the full decomposition. However, an upper bound on the number can be obtained from equation (5.3) by pretending that all $n_1 + n_2$ inequalities are two-sided.

When the probability (1.1) is decomposed into T_2 terms, each T_2 term is of the form shown in equation (3.6) consisting of a chain of independent normal random variables. However, the second variable is chosen as the root variable, so that a T_2 term is more properly expressed as

$$P(Z_1 \in I_1(Z_2), Z_2 \in I_2, Z_3 \in I_3(Z_2), \ldots, Z_k \in I_k(Z_{k-1})) \tag{5.4}$$

where Z_i, $1 \le i \le k$, are independent normal random variables. In fact, Z_k is the variable y in the initial representation of T_k according to expression (4.1), and Z_{k-1} is the variable z in the set R^* given in expression (4.19) which is the integration region for a T_{k-1} term, and so on. The intervals $I_i(x)$, $3 \le i \le k$, are either $\Re$, $(-\infty, \delta_i)$, (δ_i, ∞), $(\delta_{i1}, \delta_{i2})$, $(-\infty, x)$, (x, ∞) or $(x+\delta_{i1}, x+\delta_{i2})$, depending upon which of the seven expressions in equation (4.18) is used. The variables Z_1 and Z_2 correspond to the variables x and y respectively in equation (4.20), with $I_1(y) = (l(y), u(y))$ and $I_2 = (l, u)$.

If $W_1 = Z_1$, $W_2 = Z_2$, and $W_i = Z_i - Z_{i-1}$, for any i, $3 \le i \le k$, for which $I_i(x)$ is either $(-\infty, x)$, (x, ∞) or $(x+\delta_{i1}, x+\delta_{i2})$, while $W_i = Z_i$, for any i, $3 \le i \le k$, for which $I_i(x)$ is either $\Re$, $(-\infty, \delta_i)$, (δ_i, ∞) or $(\delta_{i1}, \delta_{i2})$, it follows that expression (5.4) can be considered to be

$$P((W_1, W_2) \in J_{12}, W_3 \in J_3, \ldots, W_k \in J_k) \tag{5.5}$$

with $J_{12} \subseteq \Re^2$ and $J_i \subseteq \Re$, $3 \le i \le k$, where the random variables $\boldsymbol{W}$ have a multivariate normal distribution with a tri-diagonal covariance matrix.

If the T_2 term in equation (5.4) is decomposed into T_1 terms, then each T_1 term will be of the form

$$P(Z_1 \in I_1, Z_2 \in I_2(Z_1), Z_3 \in I_3(Z_2), \ldots, Z_k \in I_k(Z_{k-1})) \tag{5.6}$$

so that the variable Z_1 is now the root variable. The random variables Z_i, $2 \le i \le k$ and the intervals $I_i(x)$, $3 \le i \le k$ are the same in equations (5.4) and (5.6). The interval $I_2(x)$ will also be either $\Re$, $(-\infty, \delta_2)$, (δ_2, ∞), $(\delta_{21}, \delta_{22})$, $(-\infty, x)$, (x, ∞) or $(x+\delta_{21}, x+\delta_{22})$. Expression (5.6) can be put in the form of expression (5.5), although $(W_1, W_2) \in J_{12}$ can be simplified to $W_1 \in J_1$ and $W_2 \in J_2$. Nevertheless, with regards to computational ease it should be remembered that it is unnecessary and generally disadvantageous to decompose the T_2 terms into T_1 terms.

A special case of a decomposition to T_1 terms is addressed in Miwa, Hayter and Kuriki (2003) where it is shown that a k-dimensional orthant probability for a general multivari-

ate normal distribution can be expressed as no more than $(k-1)!$ orthant probabilities for multivariate normal distributions with a tri-diagonal covariance matrix. In that paper a geometrical approach was employed to decompose the orthant region. The paper also provides some examples of the computation times required to calculate the orthant probabilities using that decomposition method.

In summary, the approach suggested in this article for the evaluation of the probability in equation (1.1) in based upon expressing it in terms of probabilities of the form (5.4), which are equivalent to probabilities of the form (5.5) . Each term of the form (5.4) is evaluated using the recursive integration method described in section 3. The decomposition of equation (1.1) is exact, and so the accuracy of the computation depends only on the accuracy of the one-dimensional integrations and the function storages employed in the recursive integration of (5.4).

Once the standard normal cumulative distribution function $\Phi(x)$ has been stored, the recursive integration of the probability in expression (5.4) generally requires $k-2$ steps, where each step has the computationally intensity of a one-dimensional integral calculation. However, many of these steps are common for different terms of the decomposition. This is because when expressed in the form (5.4), different T_2 terms will have some random variables Z_i and intervals I_i in common. Specifically, if two T_2 terms are derived from a common T_j term, $3 \le j \le k-1$, then the random variables Z_i, $j \le i \le k$, and the intervals I_i, $j+1 \le i \le k$, will be the same. Consequently, the computational intensity of this method is actually less than if the T_2 terms in the decomposition were unrelated. The number of distinct steps is actually equal to the total number of T_i terms in the decomposition with $2 \le i \le k-1$.

A rough estimate of the overall computational intensity of this method can then be made by using the upper bounds (5.2) and (5.3) on the number of terms in the decomposition. For example, consider the integration of a six-dimensional multivariate normal distribution over a region defined by two-sided inequalities for ten different linear combinations of the random variables. Equation (5.3) with $k=6$ and $n_2=10$ gives a value of at most 40,320 T_2 terms, and additionally there are actually at most 10 T_5 terms, at most 180 T_4 terms and at most 2,880 T_3 terms. Therefore, the computational intensity of this probability is no more than the equivalent of $10+180+2,880+40,320=43,390$ one-dimensional numerical integrations. This is clearly computationally feasible, and it is an immense improvement over calculating this probability by the direct numerical integration of a six-dimensional integral. If each dimension of a six-dimensional integral is specified with N grid points, then in general the direct numerical integration of the six-dimensional integral requires N^5 one-dimensional numerical integrations.

6. Examples

This section contains some examples of the evaluation of the probability (1.1) for $k=3$ and $k=4$.

6.1. Example with $k=3$

Consider the rectangular box probability

$$A = P(l_i \le X_i \le u_i; 1 \le i \le 3)$$

where the random variables $\boldsymbol{X} = (X_1, X_2, X_3)$ have a multivariate normal distribution with mean $\mathbf{0}$, and a positive-definite covariance matrix $\boldsymbol{\Sigma}$ with variances one, and covariance terms ρ_{ij}, $1 \leq i < j \leq 3$. A rectangular box probability for random variables with any mean and covariance matrix can be put in this form with a suitable scaling of the random variables and with adjustments of the boundary values l_i and u_i. Also, some of the boundary values may be infinite, and specifically an orthant probability is obtained by taking $l_i = -\infty$, $1 \leq i \leq 3$.

In accordance with the arguments presented in this paper, the required probability will be reduced to a one-dimensional integral expression by choosing appropriate linear transformations of the variables that isolate one variable as being independent of the others, so that its integration can be performed and the dimension reduced. First X_3 will be replaced by Y_3 which is integrated out, and then the remaining two variables are transformed so that they are independent, and one of them can then be integrated out.

The probability A is defined by the linear combinations $\boldsymbol{c}_i\boldsymbol{X}$, with $\boldsymbol{c}_1 = (1,0,0)$, $\boldsymbol{c}_2 = (0,1,0)$ and $\boldsymbol{c}_3 = (0,0,1)$. With $U = \{1,2\}$, $\boldsymbol{e}_1 = \boldsymbol{c}_1$ and $\boldsymbol{e}_2 = \boldsymbol{c}_2$, it follows that $\boldsymbol{e}_3$ can be chosen to be

$$\boldsymbol{e}_3 = (\rho_{12}\rho_{23} - \rho_{13}, \rho_{12}\rho_{13} - \rho_{23}, 1 - \rho_{12}^2)$$

since it is orthogonal to both $\boldsymbol{e}_1\boldsymbol{\Sigma} = (1, \rho_{12}, \rho_{13})$ and $\boldsymbol{e}_2\boldsymbol{\Sigma} = (\rho_{12}, 1, \rho_{23})$. The transformation $Y_i = \boldsymbol{e}_i\boldsymbol{X}$, $1 \leq i \leq 3$, therefore gives

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\mathbf{0}, \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}\right)$$

and

$$Y_3 = (\rho_{12}\rho_{23} - \rho_{13})X_1 + (\rho_{12}\rho_{13} - \rho_{23})X_2 + (1 - \rho_{12}^2)X_3 \sim N(0, \tau_3^2)$$

where

$$\tau_3^2 = (\rho_{12}\rho_{23} - \rho_{13})^2 + (\rho_{12}\rho_{13} - \rho_{23})^2 + (1 - \rho_{12}^2)^2 + 2\rho_{12}(\rho_{12}\rho_{23} - \rho_{13})(\rho_{12}\rho_{13} - \rho_{23})$$
$$+ 2\rho_{13}(\rho_{12}\rho_{23} - \rho_{13})(1 - \rho_{12}^2) + 2\rho_{23}(\rho_{12}\rho_{13} - \rho_{23})(1 - \rho_{12}^2).$$

and with Y_3 independent of (Y_1, Y_2). Consequently, the integration over the variable y_3 can be performed to give

$$A = \int_{(y_1,y_2) \in Q} \int \phi_2\left(y_1, y_2; \mathbf{0}, \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}\right)$$

$$\left[\Phi\left(\frac{\boldsymbol{c}^*\boldsymbol{y} + (1 - \rho_{12}^2)u_3}{\tau_3}\right) - \Phi\left(\frac{\boldsymbol{c}^*\boldsymbol{y} + (1 - \rho_{12}^2)l_3}{\tau_3}\right)\right] dy_1 dy_2 \tag{6.1}$$

where $Q = \{(y_1, y_2) : l_1 \leq y_1 \leq u_1, l_2 \leq y_2 \leq u_2\}$ and $\boldsymbol{c}^*\boldsymbol{y} = (\rho_{12}\rho_{23} - \rho_{13})y_1 + (\rho_{12}\rho_{13} - \rho_{23})y_2$. If both l_3 and u_3 are finite, this uses the special case where $I_1 \cup I_2 = \emptyset$ and $|I_3| = 1$.

If $\rho_{13} = \rho_{23} = 0$ then $\boldsymbol{c}^*\boldsymbol{y} = 0$ and expression (6.1) simplifies to

$$A = P(l_i \le X_i \le u_i; 1 \le i \le 2)\ P(l_3 \le X_3 \le u_3)$$
$$= P(l_i \le X_i \le u_i; 1 \le i \le 2)\ [\Phi(u_3) - \Phi(l_3)]. \tag{6.2}$$

The probability for the random variables X_1 and X_2 is then a two-dimensional integral expression which can be evaluated using the methods described in section 2.

Suppose that either $\rho_{13} \neq 0$ or $\rho_{23} \neq 0$. Then it is possible that $\rho_{12}\rho_{13} - \rho_{23} = 0$, although in this case $\rho_{13} \neq 0$. Similarly, it is possible that $\rho_{12}\rho_{23} - \rho_{13} = 0$, although in this case $\rho_{23} \neq 0$. It is not possible for both $\rho_{12}\rho_{13} - \rho_{23} = 0$ and $\rho_{12}\rho_{23} - \rho_{13} = 0$, since this would imply that $|\rho_{12}| = 1$ which violates the condition that $\mathbf{\Sigma}$ is positive-definite. Therefore, define

$$W_2 = \boldsymbol{c}^*\boldsymbol{Y} = (\rho_{12}\rho_{23} - \rho_{13})Y_1 + (\rho_{12}\rho_{13} - \rho_{23})Y_2 \sim N(0, \tau_2^2)$$

where

$$\tau_2^2 = (\rho_{12}\rho_{23} - \rho_{13})^2 + (\rho_{12}\rho_{13} - \rho_{23})^2 + 2\rho_{12}(\rho_{12}\rho_{23} - \rho_{13})(\rho_{12}\rho_{13} - \rho_{23})$$

and set

$$W_1 = \rho_{23}Y_1 - \rho_{13}Y_2 \sim N(0, \tau_1^2)$$

with

$$\tau_1^2 = \rho_{23}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}$$

so that W_1 is independent of W_2.

The required probability can now be expressed as

$$A = \int_{(w_1, w_2) \in Q^*} \int \phi_1(w_1; 0, \tau_1^2)\ \phi_1(w_2; 0, \tau_2^2)$$

$$\left[\Phi\left(\frac{w_2 + (1-\rho_{12}^2)u_3}{\tau_3}\right) - \Phi\left(\frac{w_2 + (1-\rho_{12}^2)l_3}{\tau_3}\right)\right] dw_1\ dw_2 \tag{6.3}$$

where the set Q^* is defined by the inequalities

$$\tau_1^2 l_1 \le -(\rho_{12}\rho_{13} - \rho_{23})w_1 - \rho_{13}w_2 \le \tau_1^2 u_1$$

and

$$\tau_1^2 l_2 \le (\rho_{12}\rho_{23} - \rho_{13})w_1 - \rho_{23}w_2 \le \tau_1^2 u_2.$$

If $\rho_{12}\rho_{13} - \rho_{23} = 0$ define $I^{(1)} = (-\tau_1^2 u_1/\rho_{13}, -\tau_1^2 l_1/\rho_{13})$ for $\rho_{13} > 0$, and define $I^{(1)} = (-\tau_1^2 l_1/\rho_{13}, -\tau_1^2 u_1/\rho_{13})$ for $\rho_{13} < 0$. Also, if $\rho_{12}\rho_{23} - \rho_{13} = 0$ define $I^{(2)} = (-\tau_1^2 u_2/\rho_{23}, -\tau_1^2 l_2/\rho_{23})$ for $\rho_{23} > 0$, and define $I^{(2)} = (-\tau_1^2 l_2/\rho_{23}, -\tau_1^2 u_2/\rho_{23})$ for $\rho_{23} < 0$. In addition, if $\rho_{12}\rho_{13} - \rho_{23} \neq 0$ let

$$b_1(w,h) = \frac{-\tau_1^2 h - \rho_{13}w}{\rho_{12}\rho_{13} - \rho_{23}}$$

and define $l_1(w) = b_1(w, u_1)$, $u_1(w) = b_1(w, l_1)$, when $\rho_{12}\rho_{13} - \rho_{23} > 0$, and define $l_1(w) =$

$b_1(w,l_1)$, $u_1(w) = b_1(w,u_1)$, when $\rho_{12}\rho_{13} - \rho_{23} < 0$. Similarly, if $\rho_{12}\rho_{23} - \rho_{13} \neq 0$ let

$$b_2(w,h) = \frac{\tau_1^2 h + \rho_{23} w}{\rho_{12}\rho_{23} - \rho_{13}}$$

and define $l_2(w) = b_2(w,l_2)$, $u_2(w) = b_2(w,u_2)$, when $\rho_{12}\rho_{23} - \rho_{13} > 0$, and define $l_2(w) = b_2(w,u_2)$, $u_2(w) = b_2(w,l_2)$, when $\rho_{12}\rho_{23} - \rho_{13} < 0$.

Consequently, if $\rho_{12}\rho_{13} - \rho_{23} = 0$ then

$$A = \int_{w \in I^{(1)}} \phi_1(w;0,\tau_2^2) \left[\Phi\left(\frac{u_2(w)}{\tau_1}\right) - \Phi\left(\frac{l_2(w)}{\tau_1}\right)\right] \left[\Phi\left(\frac{w + (1-\rho_{12}^2)u_3}{\tau_3}\right) - \Phi\left(\frac{w + (1-\rho_{12}^2)l_3}{\tau_3}\right)\right] dw \tag{6.4}$$

if $\rho_{12}\rho_{23} - \rho_{13} = 0$ then

$$A = \int_{w \in I^{(2)}} \phi_1(w;0,\tau_2^2) \left[\Phi\left(\frac{u_1(w)}{\tau_1}\right) - \Phi\left(\frac{l_1(w)}{\tau_1}\right)\right] \left[\Phi\left(\frac{w + (1-\rho_{12}^2)u_3}{\tau_3}\right) - \Phi\left(\frac{w + (1-\rho_{12}^2)l_3}{\tau_3}\right)\right] dw \tag{6.5}$$

and otherwise

$$A = \int_{w \in I^{(3)}} \phi_1(w;0,\tau_2^2) \left[\Phi\left(\frac{\min\{u_1(w),u_2(w)\}}{\tau_1}\right) - \Phi\left(\frac{\max\{l_1(w),l_2(w)\}}{\tau_1}\right)\right] \left[\Phi\left(\frac{w + (1-\rho_{12}^2)u_3}{\tau_3}\right) - \Phi\left(\frac{w + (1-\rho_{12}^2)l_3}{\tau_3}\right)\right] dw \tag{6.6}$$

where the interval $I^{(3)}$ is defined to be the values of w for which $\max\{l_1(w),l_2(w)\} \leq \min\{u_1(w),u_2(w)\}$.

As explained in section 5, the three dimensional rectangular box probability A can be evaluated with just one term, using expression (6.2), (6.4), (6.5) or (6.6). The evaluation only requires knowledge of $\Phi(x)$ and a one-dimensional integral calculation. Notice that expression (6.6), for example, can be written as

$$A = P\left(Z_1 \in I_1(Z_2), Z_2 \in I^{(3)}, Z_3 \in I_3(Z_2)\right)$$

where $I_1(x) = (\max\{l_1(x),l_2(x)\}, \min\{u_1(x),u_2(x)\})$, $I_3(x) = (x + (1-\rho_{12}^2)l_3, x + (1-\rho_{12}^2)l_3)$ and the random variables $Z_i \sim N(0,\tau_i^2)$, $1 \leq i \leq 3$, are independent, which is in the form of equation (5.4) with $k = 3$.

Finally, suppose that there are more than three linear combinations of $\boldsymbol{X}$ which define the probability of interest. Specifically, suppose that the required probability is

$$A = P(l_i \le X_i \le u_i, 1 \le i \le 3;\ l_i \le c_{i1}X_1 + c_{i2}X_2 \le u_i, 4 \le i \le 3+m_1;$$

$$l_i \le c_{i1}X_1 + c_{i2}X_2 + c_{i3}X_3 \le u_i, 4+m_1 \le i \le 3+m_1+m_2)$$

where $c_{i3} \neq 0$, $4+m_1 \le i \le 3+m_1+m_2$.

If $m_2 = 0$, so that the additional linear combinations $c_{i1}X_1 + c_{i2}X_2$, $4 \le i \le 3+m_1$, are all of the form $\boldsymbol{c}_i\boldsymbol{X}$ where $\boldsymbol{c}_i$ is in the space spanned by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, the probability A can still be expressed with just one term. The expressions (6.1) and (6.3) are both still valid, although the sets Q and Q^* must incorporate the additional inequalities. The expression for A can be derived from (6.3), with an integration being performed over the w_2 variable, and the bounds on w_1 being derived from the set Q^* and incorporated into the arguments of $\Phi(x)$.

However, if $m_2 \neq 0$ then the probability A needs to be expressed with more than one term. This is because the random variable Y_3 is a component of more than one of the linear combinations of the random variables $\boldsymbol{Y}$ which define the probability A. In the decomposition of A there will be one term for each value l_3, u_3, and l_i, u_i, $4+m_1 \le i \le 3+m_1+m_2$, which is finite.

6.2. Example with $k = 4$

In this section it is shown how the decomposition process can be applied to an example concerning a four-dimensional multivariate normal distribution where the probability is specified by five one-sided inequalities. Specifically, let $\boldsymbol{X} = (X_1, X_2, X_3, X_4) \sim N_4(\boldsymbol{0}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & \rho \\ 0 & 1 & 0 & \rho \\ 0 & 0 & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

and consider the probability

$$A = P(X_i \le u_i, 1 \le i \le 4;\ 2X_1 + X_2 + X_3 + X_4 \le u_5)$$

with $2u_1 + u_2 + u_3 + u_4 > u_5$ so that none of the five inequalities in the definition of A is redundant.

As in the previous example, the required probability will be reduced to a series of one-dimensional integral expressions by choosing appropriate linear transformations of the variables that isolate one variable as being independent of the others, so that its integration can be performed and the dimension reduced. This is done three times in order to reduce to one-dimensional expressions, which in this case also include functions in the integrand ($g_1(x)$ and $g_2(x)$) that need to be evaluated from one-dimensional integral calculations. The increase in the number of terms, and the integrands that also require a one-dimensional integral calculation, are typical aspects of the methodology proposed in this paper. Similar derivations for a four-dimensional box probability are presented in Hayter and Lin (2011), where programming and practical implementation issues are discussed.

Let $\boldsymbol{e}_1 = (1,0,0,0)$, $\boldsymbol{e}_2 = (0,1,0,0)$ and $\boldsymbol{e}_3 = (0,0,1,0)$, and then set $\boldsymbol{e}_4 = (-\rho, -\rho, -\rho, 1)$ which is orthogonal to $\boldsymbol{e}_i\boldsymbol{\Sigma}$, $1 \le i \le 3$. Then the change of variables $Y_i = \boldsymbol{e}_i\boldsymbol{X}$, $1 \le i \le 4$, allows the probability to be written

$$A = \int \cdots \int_{\boldsymbol{y} \in R} \phi_3(y_1, y_2, y_3; \mathbf{0}, \boldsymbol{I}_3)\ \phi_1(y_4; 0, 1-3\rho^2)\ d\boldsymbol{y}$$

with

$$R = \{\boldsymbol{y} :\ y_1 \le u_1,\ y_2 \le u_2,\ y_3 \le u_3,$$

$$y_4 \le -\rho(y_1+y_2+y_3)+u_4,\ y_4 \le -(2+\rho)y_1-(1+\rho)(y_2+y_3)+u_5\}.$$

However, $A = B_1 + B_2$ where

$$B_i = \int \cdots \int_{\boldsymbol{y} \in R_i} \phi_3(y_1, y_2, y_3; \mathbf{0}, \boldsymbol{I}_3)\ \phi_1(y_4; 0, 1-3\rho^2)\ d\boldsymbol{y}$$

for $i = 1, 2$ with

$$R_1 = \{\boldsymbol{y} :\ y_1 \le u_1,\ y_2 \le u_2,\ y_3 \le u_3,$$

$$-\rho(y_1+y_2+y_3)+u_4 \le -(2+\rho)y_1-(1+\rho)(y_2+y_3)+u_5,\ y_4 \le -\rho(y_1+y_2+y_3)+u_4\}$$

and

$$R_2 = \{\boldsymbol{y} :\ y_1 \le u_1,\ y_2 \le u_2,\ y_3 \le u_3,$$
$$-(2+\rho)y_1-(1+\rho)(y_2+y_3)+u_5 \le -\rho(y_1+y_2+y_3)+u_4,$$
$$y_4 \le -(2+\rho)y_1-(1+\rho)(y_2+y_3)+u_5\}.$$

The integration over the y_4 variable can then be performed for B_1 and B_2 to give

$$B_1 = \int \cdots \int_{(y_1,y_2,y_3) \in R_1^*} \phi_3(y_1, y_2, y_3; \mathbf{0}, \boldsymbol{I}_3)\ \Phi\left(\frac{-\rho(y_1+y_2+y_3)+u_4}{\sqrt{1-3\rho^2}}\right) dy_1\ dy_2\ dy_3$$

where

$$R_1^* = \{(y_1, y_2, y_3) :\ y_1 \le u_1,\ y_2 \le u_2,\ y_3 \le u_3,\ 2y_1+y_2+y_3 \le u_5-u_4\}$$

and

$$B_2 = \int \cdots \int_{(y_1,y_2,y_3) \in R_2^*} \phi_3(y_1, y_2, y_3; \mathbf{0}, \boldsymbol{I}_3)$$
$$\Phi\left(\frac{-(2+\rho)y_1-(1+\rho)(y_2+y_3)+u_5}{\sqrt{1-3\rho^2}}\right) dy_1\ dy_2\ dy_3$$

where

$$R_2^* = \{(y_1, y_2, y_3) :\ y_1 \le u_1,\ y_2 \le u_2,\ y_3 \le u_3,\ -2y_1-y_2-y_3 \le u_4-u_5\}.$$

A change of variables is now required for B_1 and B_2. For B_1 the new variables W_1, W_2 and $W_3 = Y_1 + Y_2 + Y_3 \sim N(0, 3)$ can be used where W_1 and W_2 are both chosen to

be independent of W_3, and for B_2 the new variables W_1', W_2' and $W_3' = (2+\rho)Y_1 + (1+\rho)(Y_2+Y_3) \sim N(0, 3\rho^2+8\rho+6)$ can be used where again W_1' and W_2' are both chosen to be independent of W_3'. After these changes, the integration regions R_1^* and R_2^* can be partitioned in order to isolate the dependence on W_3 and W_3' respectively.

For example, let $W_1 = Y_1 - Y_3$ and $W_2 = Y_2 - Y_3$ so that

$$B_1 = \int \cdots \int_{(w_1,w_2,w_3) \in S_1} \phi_2\left(w_1, w_2; \mathbf{0}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right) \phi_1(w_3; 0, 3) \, \Phi\left(\frac{-\rho w_3 + u_4}{\sqrt{1-3\rho^2}}\right) dw_1 \, dw_2 \, dw_3$$

where

$$S_1 = \{(w_1, w_2, w_3) : \; w_3 \le c_i(w_1, w_2), 1 \le i \le 4\}$$

with $c_1(w_1,w_2) = -2w_1 + w_2 + 3u_1$, $c_2(w_1,w_2) = w_1 - 2w_2 + 3u_2$, $c_3(w_1,w_2) = w_1 + w_2 + 3u_3$ and $c_4(w_1,w_2) = (-2w_1 + w_2 + 3(u_5 - u_4))/4$. The set S_1 can then be partitioned into S_{11}, S_{12}, S_{13} and S_{14} where

$$S_{1i} = \{(w_1, w_2, w_3) : \; w_3 \le c_i(w_1, w_2); c_i(w_1, w_2) \le c_j(w_1, w_2), 1 \le j \le 4, j \ne i \}$$

for $1 \le i \le 4$, so that $B_1 = C_{11} + C_{12} + C_{13} + C_{14}$ with

$$C_{1i} = \int \cdots \int_{(w_1,w_2,w_3) \in S_{1i}} \phi_2\left(w_1, w_2; \mathbf{0}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right) \phi_1(w_3; 0, 3) \, \Phi\left(\frac{-\rho w_3 + u_4}{\sqrt{1-3\rho^2}}\right) dw_1 \, dw_2 \, dw_3$$

for $1 \le i \le 4$. The integration over the w_3 variable can be performed for each of C_{11}, C_{12}, C_{13} and C_{14}. For example, if

$$g_1(x) = \int_{w_3=-\infty}^{x} \phi_1(w_3; 0, 3) \, \Phi\left(\frac{-\rho w_3 + u_4}{\sqrt{1-3\rho^2}}\right) dw_3$$

then

$$C_{11} = \int \cdots \int_{(w_1,w_2) \in S_{11}^*} \phi_2\left(w_1, w_2; \mathbf{0}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right) g_1(-2w_1 + w_2 + 3u_1) \, dw_1 \, dw_2$$

with

$$S_{11}^* = \{(w_1, w_2) : \; -w_1 + w_2 \le u_2 - u_1, \; -w_1 \le u_3 - u_1, \; -2w_1 + w_2 \le u_5 - u_4 - 4u_1\}.$$

A final change of variables $Z_2 = -2W_1 + W_2 \sim N(0, 6)$ can be made for C_{11}, with $Z_1 =$

$W_2 \sim N(0,2)$ being chosen to be independent of Z_2. The integration region S_{11}^* provides the inequalities $z_1 + z_2 \leq 2(u_2 - u_1)$, $-z_1 + z_2 \leq 2(u_3 - u_1)$ and $z_2 \leq u_5 - u_4 - 4u_1$ which can be expressed as

$$z_2 \leq u_5 - u_4 - 4u_1, \quad z_2 + 2u_1 - 2u_3 \leq z_1 \leq -z_2 + 2u_2 - 2u_1.$$

Notice that $2u_1 + u_2 + u_3 + u_4 > u_5$ implies that $z_2 + 2u_1 - 2u_3 \leq -z_2 + 2u_2 - 2u_1$ for $z_2 \leq u_5 - u_4 - 4u_1$. Consequently,

$$C_{11} = \int_{z_2=-\infty}^{u_5-u_4-4u_1} \phi_1(z_2;0,6)\; g_1(z_2+3u_1) \left[\Phi\left(\frac{-z_2+2u_2-2u_1}{\sqrt{2}}\right) - \Phi\left(\frac{z_2+2u_1-2u_3}{\sqrt{2}}\right)\right] dz_2.$$

The final expressions for C_{12}, C_{13} and C_{14} will also be one-dimensional integrals with the integrand involving $\Phi(x)$ and the function $g_1(x)$.

The term B_2 is also decomposed into $B_2 = C_{21} + C_{22} + C_{23} + C_{24}$, where each of the terms C_{21}, C_{22}, C_{23} and C_{24} can be evaluated from a one-dimensional integral expression with the integrands involving $\Phi(x)$ and the function

$$g_2(x) = \int_{w_3'=-\infty}^{x} \phi_1(w_3';0,3\rho^2+8\rho+6)\; \Phi\left(\frac{-w_3'+u_5}{\sqrt{1-3\rho^2}}\right) dw_3.$$

Therefore, the required probability is

$$A = C_{11} + C_{12} + C_{13} + C_{14} + C_{21} + C_{22} + C_{23} + C_{24}$$

so that the decomposition produces eight terms, which is consistent with equation (5.2) with $k = 4$ and $n_1 = 5$. Notice that A can be calculated from $\Phi(x)$, $g_1(x)$, $g_2(x)$ and eight one-dimensional integrations. However, the computational intensity of calculating and storing the functions $g_1(x)$ and $g_2(x)$ (and similarly $\Phi(x)$ if necessary) is each equivalent to a one-dimensional integration. Therefore, the total computational intensity of calculating the probability A is no more than the equivalent of 10 one-dimensional integral calculations (or 11 one-dimensional integral calculations if $\Phi(x)$ needs to be calculated). In fact, this bound on the computational intensity is valid for any probability defined by five one-sided inequalities of linear combinations of $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$ where $\boldsymbol{X}$ can have any covariance matrix. The probability T_4 is decomposed into at most two T_3 terms (B_i) and at most eight T_2 terms (C_{ij}).

Finally, notice that for this example C_{11} can be written as

$$C_{11} = P(Z_1 \in I_1(Z_2), Z_2 \in I_2, Z_3 \in I_3(Z_2), Z_4 \in I_4(Z_3))$$

where $I_1(x) = (x + 2u_1 - 2u_3, -x + 2u_2 - 2u_1)$, $I_2 = (-\infty, u_5 - u_4 - 4u_1)$, $I_3(x) = (-\infty, x + 3u_1)$ and $I_4(x) = (-\infty, -\rho x + u_4)$, and the random variables $Z_1 \sim N(0,2)$, $Z_2 \sim N(0,6)$, $Z_3 \sim N(0,3)$ and $Z_4 \sim N(0, 1-3\rho^2)$ are independent. This is in the form of equation (5.4) with $k = 4$. Similar representations can be made for the other C_{ij} terms.

References

Armitage, P., McPherson, C.K., Rowe, B.C., 1969. Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society, Series A*, 132, 235–244.

Bretz, F., Genz, A., Hothorn, L.A., 2001. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43, 645–656.

Genz, A., Bretz, F., 1999. Numerical computation of multivariate *t* probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63, 361–378.

Genz, A., Bretz, F., 2002. Methods for the computation of multivariate *t*-probabilities. *Journal of Computational and Graphical Statistics*, 11, 950–971.

Genz, A., Bretz, F., 2009. *Computation of Multivariate Normal and t Probabilities*. Springer.

Hayter, A.J., 2006. Recursive integration methodologies with statistical applications. *Journal of Statistical Planning and Inference*, 136, 2284–2296.

Hayter, A.J., Liu, W., 1996a. Exact calculations for the one-sided studentized range test for testing against a simple ordered alternative. *Computational Statistics and Data Analysis*, 22, 17–25.

Hayter, A.J., Liu, W., 1996b. A note on the calculation of $P(X_1 < X_2 < \ldots < X_k)$. *The American Statistician*, 50(4), 365.

Hayter, A.J., Lin, Y., 2011. The evaluation of two-sided orthant probabilities for a quadrivariate normal distribution. To appear in *Computational Statistics,* 2012.

Jennison, C., Turnbull, B.W., 1991. Exact calculations for sequential t, χ^2 and F tests. *Biometrika*, 78, 133–141.

Miwa, T., Hayter, A.J., Kuriki, S., 2003. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society, Series B*, 65, 223–234.

Somerville, P.N., 1997. Multiple testing and simultaneous confidence intervals: calculation of constants. *Computational Statistics and Data Analysis*, 25, 217–233.

Somerville, P.N., 1998. Numerical computation of multivariate normal and multivariate *t* probabilities over convex regions. *J. Computational and Graphical Statistics*, 7, 529–544.

Tong, Y.L., 1990. The Multivariate Normal Distribution. Springer-Verlag.