J. R. Statist. Soc. B (2003)
65, *Part* 1, *pp.* 223–234

The evaluation of general non-centred orthant probabilities

Tetsuhisa Miwa,

National Institute for Agro-Environmental Sciences, Tsukuba, Japan

A. J. Hayter

Georgia Institute of Technology, Atlanta, USA

and Satoshi Kuriki

Institute of Statistical Mathematics, Tokyo, Japan

[Received October 2000. Revised July 2002]

Summary. The evaluation of the cumulative distribution function of a multivariate normal distribution is considered. The multivariate normal distribution can have any positive definite correlation matrix and any mean vector. The approach taken has two stages. In the first stage, it is shown how non-centred orthoscheme probabilities can be evaluated by using a recursive integration method. In the second stage, some ideas of Schläfli and Abrahamson are extended to show that any non-centred orthant probability can be expressed as differences between at most $(m-1)!$ non-centred orthoscheme probabilities. This approach allows an accurate evaluation of many multivariate normal probabilities which have important applications in statistical practice.

Keywords: Cumulative distribution function; Orthant probability; Orthoscheme probability; Polyhedral cone; Recursive integration

1. Introduction

Suppose that the random variables $x_1, \ldots, x_m$ have any non-singular multivariate normal distribution and consider the non-centred orthant probability

$$P_m(\boldsymbol{\mu}, \mathbf{R}) = \Pr\{x_i \geqslant 0, 1 \leqslant i \leqslant m\} = \int_0^\infty \cdots \int_0^\infty \phi_m(\mathbf{x}; \boldsymbol{\mu}, \mathbf{R})\, \mathrm{d}x_1 \ldots \mathrm{d}x_m \qquad (1)$$

where $\mathbf{x} = (x_1, \ldots, x_m)' \sim N_m(\boldsymbol{\mu}, \mathbf{R})$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)'$ and a positive definite correlation matrix $\mathbf{R} = \{\rho_{ij}\}$ with density function $\phi_m(\mathbf{x}; \boldsymbol{\mu}, \mathbf{R})$. Note that the correlation matrix $\mathbf{R}$ is not changed if $\mathbf{x}$ is replaced by $-\mathbf{x} = (-x_1, \ldots, -x_m)'$ and that the multivariate normal distribution function can be expressed as

$$\begin{aligned} F_m(\mathbf{c}) &= \Pr\{x_i \leqslant c_i, 1 \leqslant i \leqslant m\} \\ &= \Pr\{-x_i \geqslant -c_i, 1 \leqslant i \leqslant m\} = P_m(-\boldsymbol{\mu} + \mathbf{c}, \mathbf{R}) \end{aligned}$$

for any vector $\mathbf{c} = (c_1, \ldots, c_m)'$.

Since the multivariate normal distribution plays a fundamental role in statistical applications there is an enormous literature on the computation of the integral expression (1). Unless

Address for correspondence: Tetsuhisa Miwa, National Institute for Agro-Environmental Sciences, 3-1-3 Kannondai, Tsukuba 305-8604, Japan.
E-mail: miwa@niaes.affrc.go.jp

 1369–7412/03/65223

the dimension m is very small any direct approach to the evaluation of expression (1) through numerical integration techniques is unmanageable. Alternative approximation methods have included infinite series expansions, dimensional reductions in numerical integration and Monte Carlo simulations. For general discussions of this problem see, for example, Johnson and Kotz (1972), chapter 35, Stuart and Ord (1994), chapter 15, and Tong (1990), chapter 8.

If $\boldsymbol{\mu} = \mathbf{0} = (0, \ldots, 0)'$ then $P_m(\mathbf{0}, \mathbf{R})$ is a centred orthant probability, and furthermore, for a tridiagonal correlation matrix $\mathbf{R}$ satisfying $\rho_{ij} = 0$ for $|i - j| > 1$, $P_m(\boldsymbol{\mu}, \mathbf{R})$ is called a non-centred orthoscheme probability and $P_m(\mathbf{0}, \mathbf{R})$ is a centred orthoscheme probability. There are some interesting relationships between orthant probabilities and orthoscheme probabilities. Schläfli (1858) proved that a centred orthant probability can be written as a function of $m!$ centred orthoscheme probabilities. Abrahamson (1964) showed that with $m = 4$ a quadrivariate centred orthant probability can be expressed as a linear combination of no more than six (instead of $4! = 24$) centred orthoscheme probabilities.

In this paper these results are extended and it is shown how any non-centred orthant probability can be expressed as differences between at most $(m-1)!$ non-centred orthoscheme probabilities. Additionally, a quick and accurate recursive integration algorithm is presented to evaluate non-centred orthoscheme probabilities. Together, these results provide a method to evaluate expression (1) accurately, and consequently any multivariate distribution function $F_m(\mathbf{c})$.

First, in Section 2, the recursive integration method is provided to evaluate non-centred orthoscheme probabilities. Then, as described in Section 3, the required non-centred orthant probability is expressed in terms of non-centred orthoscheme probabilities. The paper concludes with Section 4 where some numerical illustrations of this method are presented.

2. The evaluation of non-centred orthoscheme probabilities

In this section attention is directed towards an $m \times m$ tridiagonal correlation matrix $\mathbf{R} = \{\rho_{ij}\}$ for which $\rho_{ij} = 0$ for $|i - j| > 1$. Miwa *et al.* (2000) gave a procedure to calculate centred orthoscheme probabilities $P_m(\mathbf{0}, \mathbf{R})$. However, in this section an improved procedure is presented for evaluating any non-centred orthoscheme probability $P_m(\boldsymbol{\mu}, \mathbf{R})$.

2.1. Recursive integration formulae

The positive definite tridiagonal matrix $\mathbf{R}$ can be decomposed as $\mathbf{R} = \mathbf{B}\mathbf{B}'$ where the only non-zero elements of $\mathbf{B} = \{b_{ij}\}$ are the diagonal elements b_{ii}, $1 \leqslant i \leqslant m$, and the elements below the diagonal $b_{i,i-1}$, $2 \leqslant i \leqslant m$. If we define

$$D_i = \begin{vmatrix} 1 & \rho_{21} & 0 & \cdots & 0 \\ \rho_{21} & 1 & \rho_{32} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho_{i-1,i-2} & 1 & \rho_{i,i-1} \\ 0 & \cdots & 0 & \rho_{i,i-1} & 1 \end{vmatrix}, \qquad 1 \leqslant i \leqslant m,$$

then $D_i > 0$, $1 \leqslant i \leqslant m$, since $\mathbf{R}$ is positive definite. Moreover, these determinants are related by the equations

$$D_i = D_{i-1} - \rho_{i,i-1}^2 D_{i-2}, \qquad 2 \leqslant i \leqslant m,$$

with $D_0 = D_1 = 1$ and can therefore be evaluated in a sequential manner. Furthermore, the elements of $\mathbf{B}$ can then be successively obtained by equating the corresponding elements in $\mathbf{R} = \mathbf{B}\mathbf{B}'$ as

$$b_{11} = 1 = \sqrt{(D_1/D_0)},$$
$$b_{i,i-1} = \rho_{i,i-1}/b_{i-1,i-1} = \rho_{i,i-1}\sqrt{(D_{i-2}/D_{i-1})},$$
$$b_{ii} = \sqrt{(1 - b_{i,i-1}^2)} = \sqrt{(D_i/D_{i-1})}, \qquad 2 \leqslant i \leqslant m.$$

Next, if $\mathbf{x} = (x_1, \ldots, x_m)' \sim N_m(\boldsymbol{\mu}, \mathbf{R})$, then it follows from the decomposition $\mathbf{R} = \mathbf{B}\mathbf{B}'$ that $\mathbf{z} = (z_1, \ldots, z_m)' = \mathbf{B}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim N_m(\mathbf{0}, \mathbf{I}_m)$ and

$$x_1 = z_1 + \mu_1,$$
$$x_i = b_{i,i-1}z_{i-1} + b_{ii}z_i + \mu_i, \qquad 2 \leqslant i \leqslant m.$$

Then

$$\begin{aligned} P_m(\boldsymbol{\mu}, \mathbf{R}) &= \Pr\{x_i \geqslant 0, 1 \leqslant i \leqslant m\} \\ &= \Pr\{z_1 + \mu_1 \geqslant 0, b_{i,i-1}z_{i-1} + b_{ii}z_i + \mu_i \geqslant 0, 2 \leqslant i \leqslant m\}. \end{aligned}$$

Now define

$$\begin{aligned} f_{m-1}(z) &= \int_{(-\mu_m - b_{m,m-1}z)/b_{mm}}^{\infty} \phi(t)\, \mathrm{d}t, \\ f_{i-1}(z) &= \int_{(-\mu_i - b_{i,i-1}z)/b_{ii}}^{\infty} f_i(t)\, \phi(t)\, \mathrm{d}t, \qquad 2 \leqslant i \leqslant m-1, \end{aligned} \tag{2}$$

where $\phi(t)$ is the standard normal probability density function, so that the required probability is given by

$$P_m(\boldsymbol{\mu}, \mathbf{R}) = \int_{-\mu_1}^{\infty} f_1(z)\, \phi(z)\, \mathrm{d}z. \tag{3}$$

These integral expressions can be evaluated by using a recursive computational approach as follows. First, the function $f_{m-1}(z)$ is evaluated over a fine grid of points and the values are stored as an array. Then using equation (2) the values of $f_{m-2}(z)$ can be calculated over a similar grid of points and also stored as an array. At this point the original array can be deleted. Continuing in this fashion, the functions $f_i(z)$ are recursively calculated and eventually the required non-centred orthoscheme probability is obtained through the evaluation of equation (3). The computational intensity increases only linearly in m, and therefore sufficient accuracy can be achieved for any practical value of m by incorporating a sufficient number of grid points.

This recursive computational approach has been employed in Hayter and Liu (1996) and Miwa *et al.* (2000). It should also be pointed out that Moran (1985) derived similar recursive formulae for the special case when the $\rho_{i,i-1}$ are all equal to certain values by expressing $x_1, \ldots, x_m$ in terms of $m + 1$ independent normal random variables.

2.2. Comparisons with exact values

When $\rho_{i,i-1} = \rho, 2 \leqslant i \leqslant m$, and $\rho_{ij} = 0, |i - j| > 1$, the exact values of $P_m(\mathbf{0}, \mathbf{R})$ are known for some special values of ρ. For example, if $\rho = -\frac{1}{2}$ then $P_m(\mathbf{0}, \mathbf{R}) = 1/(m + 1)!$. Also, when

$\rho = \frac{1}{2}$ Moran (1983) proved that

$$P_{2n}(\mathbf{0}, \mathbf{R}) = \frac{(-1)^n \times 2^{2n+2}(2^{2n+2} - 1)B_{2n+2}}{(2n+2)!},$$

$$P_{2n+1}(\mathbf{0}, \mathbf{R}) = \frac{(-1)^{n+1} E_{2n+2}}{(2n+2)!},$$

where B_n and E_n are the Bernoulli and Euler numbers.

These results allow the accuracy of the proposed recursive integration procedure to be checked. In the implementation of the recursive integration (2), grid points are allocated within $[-8, 8]$ where the lengths of the grid intervals are proportional to $\phi(z)^{-1/4}$, and cubic polynomial approximations are applied to the functions $f_i(z)$. This choice of grid points and the cubic polynomial approximation are discussed in detail in Miwa *et al.* (2000). The results for $m = 5$ and $m = 10$ are given in Table 1 where G is the number of grid points employed. With $G = 128$ grid points the numerical results agree with the known exact values up to eight decimal places except for the case $m = 10$ and $\rho = -\frac{1}{2}$. However, the same accuracy can be achieved with $G = 512$ grid points for the case $m = 10$ and $\rho = -\frac{1}{2}$, where the probability value is actually very small. For all cases the computational time was negligible.

It should be remembered that no computational difficulty is added to the recursive integration procedure for non-centred orthoscheme probabilities $P_m(\boldsymbol{\mu}, \mathbf{R})$, although no closed form expressions are available for the probabilities in this case. The accuracy of the overall method to evaluate multivariate normal probabilities is discussed in Section 4.

3. Expressing orthant probabilities as differences of orthoscheme probabilities

In this section it is shown how a non-centred orthant probability $P_m(\boldsymbol{\mu}, \mathbf{R})$ can be expressed as differences between at most $(m-1)!$ non-centred orthoscheme probabilities. This extends the work of Schläfli (1858) and is based on the geometrical consideration employed by Abrahamson (1964) with $m = 4$.

To proceed it is useful to have the following definition concerning the form of a symmetric matrix $\boldsymbol{\Sigma}$.

Table 1. Numerical calculations of orthoscheme probabilities†

G	*Probabilities for the following values of m and ρ:*			
	$m = 5$		$m = 10$	
	$\rho = \frac{1}{2}$	$\rho = -\frac{1}{2}$	$\rho = \frac{1}{2}$	$\rho = -\frac{1}{2}$
32	0.0847221	0.001388885	0.00886321	0.2502×10^{-7}
64	0.084722219	0.0013888888	0.0088632345	0.25051×10^{-7}
128	0.084722222	0.0013888889	0.0088632355	0.250520×10^{-7}
256				$0.25052105 \times 10^{-7}$
512				$0.25052108 \times 10^{-7}$
Exact	0.084722222	0.0013888889	0.0088632355	$0.25052108 \times 10^{-7}$

†G is the number of grid points.

Definition 1. For $0 \leqslant r \leqslant m-3$ an $m \times m$ symmetric matrix $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ is said to have 'orthoscheme order r' if $\sigma_{ij} = 0$ for $1 \leqslant i \leqslant r, i+2 \leqslant j \leqslant m$ and $\sigma_{r+1,j} \neq 0$ for some j satisfying $r+3 \leqslant j \leqslant m$. A tridiagonal matrix is defined to have orthoscheme order $m-2$.

According to this definition, any symmetric matrix with $\sigma_{1j} \neq 0$ for some j satisfying $3 \leqslant j \leqslant m$ has orthoscheme order 0. However, if the normal random variables $x_1, \ldots, x_m$ have a covariance matrix $\boldsymbol{\Sigma}$, then the covariance matrix corresponding to some permutation of these random variables may have an orthoscheme order that is different from that of $\boldsymbol{\Sigma}$.

The main result of this section is to show how a non-centred orthant probability defined in terms of a correlation matrix $\mathbf{R}$ of orthoscheme order r can be expressed in terms of the differences of at most $m-r-1$ non-centred orthant probabilities defined in terms of correlation matrices whose orthoscheme order is each strictly larger than r.

The positive definite correlation matrix $\mathbf{R}$ can be written in the form $\mathbf{R} = \mathbf{A}'\mathbf{A}$ where the matrix $\mathbf{A}$ is composed of the linearly independent column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$, so that $\mathbf{a}_i'\mathbf{a}_i = 1$ and $\mathbf{a}_i'\mathbf{a}_j = \rho_{ij}$ for $i \neq j$. Let the polyhedral cone Q, with vertex at the origin, be defined by

$$Q = \{\mathbf{z}: \mathbf{a}_i'\mathbf{z} \geqslant 0, 1 \leqslant i \leqslant m\}.$$

Then if $\mathbf{x} \sim N_m(\boldsymbol{\mu}, \mathbf{R})$ so that $\mathbf{z} = \mathbf{A}'^{-1}\mathbf{x} \sim N_m(\mathbf{A}'^{-1}\boldsymbol{\mu}, \mathbf{I}_m)$, it follows that

$$\begin{aligned} P_m(\boldsymbol{\mu}, \mathbf{R}) &= \Pr\{x_i \geqslant 0, 1 \leqslant i \leqslant m\} \\ &= \Pr\{\mathbf{a}_i'\mathbf{z} \geqslant 0, 1 \leqslant i \leqslant m\} \\ &= \Pr\{\mathbf{z} \in Q\}. \end{aligned}$$

The polyhedral cone Q is bounded by m hyperplanes

$$H_i = \{\mathbf{z}: \mathbf{a}_i'\mathbf{z} = 0\}, \qquad 1 \leqslant i \leqslant m,$$

where $\mathbf{a}_i$ is the normal vector to the hyperplane H_i. Also, it is useful to know the edge vectors of the cone. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are the column vectors of $\mathbf{V} = \mathbf{A}'^{-1}$ so that $\mathbf{a}_i'\mathbf{v}_i = 1$ and $\mathbf{a}_i'\mathbf{v}_j = 0$ for $i \neq j$. The vector $\mathbf{v}_k$ simultaneously belongs to the hyperplanes H_i, $1 \leqslant i \leqslant m, i \neq k$, and so it can be viewed as an edge vector of the polyhedral cone Q. The polyhedral cone Q can then be expressed in terms of the linearly independent edge vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ as

$$Q = \{\mathbf{z}: \mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_m\mathbf{v}_m,\ \lambda_i \geqslant 0\}.$$

This representation is not changed if we consider $c_i\mathbf{v}_i$ instead of $\mathbf{v}_i$ for any positive constants c_i or if the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are rearranged in a different order.

Now suppose that the correlation matrix $\mathbf{R} = \{\rho_{ij}\}$ has orthoscheme order r with $0 \leqslant r \leqslant m-3$. Let a new edge vector $\mathbf{p}$ be defined by

$$\mathbf{p} = \gamma_{r+2}\mathbf{v}_{r+2} + \ldots + \gamma_m\mathbf{v}_m,$$

where the coefficients $\gamma_s, r+2 \leqslant s \leqslant m$, are given by

$$\gamma_s = \begin{cases} -\rho_{r+1,s}, & \text{if all } \rho_{r+1,i} \leqslant 0 \text{ for } r+2 \leqslant i \leqslant m, \\ \rho_{r+1,s}, & \text{otherwise,} \end{cases} \tag{4}$$

so that at least one γ_s is strictly positive. Next, for each $\gamma_s \neq 0, r+2 \leqslant s \leqslant m$, consider a cone $Q^{(s)}$ which is defined by the linearly independent edge vectors

$$\mathbf{v}_1, \ldots, \mathbf{v}_{s-1}, \mathbf{p}, \mathbf{v}_{s+1}, \ldots, \mathbf{v}_m. \tag{5}$$

Then we can establish the following lemma. The proof is provided in Appendix A.

Lemma 1.

(a) $Q \cup (\cup_{\gamma_s<0} Q^{(s)}) = \cup_{\gamma_s>0} Q^{(s)}$.
(b) The intersection $Q^{(s)} \cap Q^{(t)}$, $s \neq t$, is contained within a subspace of dimension $m-1$ if γ_s and γ_t have the same sign.
(c) The intersection $Q \cap Q^{(s)}$ for $\gamma_s < 0$ is contained within a subspace of dimension $m-1$.

Since each subspace of dimension $m-1$ has probability 0, it follows directly from lemma 1 that

$$\Pr\{\mathbf{z} \in Q\} + \sum_{\gamma_s<0} \Pr\{\mathbf{z} \in Q^{(s)}\} = \sum_{\gamma_s>0} \Pr\{\mathbf{z} \in Q^{(s)}\}.$$

Furthermore, by rearranging the edge vectors (5) in the order

$$\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}, \mathbf{p}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_{s-1}, \mathbf{v}_{s+1}, \ldots, \mathbf{v}_m,$$

the corresponding normal vectors $\mathbf{a}_i^{(s)}$, $1 \leqslant i \leqslant m$, are seen to be

$$\mathbf{a}_i^{(s)} = \mathbf{a}_i, \qquad 1 \leqslant i \leqslant r+1,$$

$$\mathbf{a}_{r+2}^{(s)} = \operatorname{sgn}(\gamma_s)\mathbf{a}_s,$$

$$\mathbf{a}_i^{(s)} = c_{i-1}^{(s)}\left(\mathbf{a}_{i-1} - \frac{\rho_{r+1,i-1}}{\rho_{r+1,s}}\mathbf{a}_s\right), \qquad r+3 \leqslant i \leqslant s,$$

$$\mathbf{a}_i^{(s)} = c_i^{(s)}\left(\mathbf{a}_i - \frac{\rho_{r+1,i}}{\rho_{r+1,s}}\mathbf{a}_s\right), \qquad s+1 \leqslant i \leqslant m,$$

with

$$c_i^{(s)} = \left(1 - \frac{2\rho_{r+1,i}\rho_{is}}{\rho_{r+1,s}} + \frac{\rho_{r+1,i}^2}{\rho_{r+1,s}^2}\right)^{-1/2}.$$

Note that $\mathbf{a}_i^{(s)\prime}\mathbf{a}_j^{(s)} = 0$ for $1 \leqslant i \leqslant r+1, i+2 \leqslant j \leqslant m$. Then defining $\mathbf{R}^{(s)} = \{\rho_{ij}^{(s)}\}$ and $\boldsymbol{\mu}^{(s)} = (\mu_1^{(s)}, \ldots, \mu_m^{(s)})'$ by

$$\begin{aligned} \rho_{ij}^{(s)} &= \mathbf{a}_i^{(s)\prime}\mathbf{a}_j^{(s)}, \\ \mu_i^{(s)} &= \mathbf{a}_i^{(s)\prime}\mathbf{A}'^{-1}\boldsymbol{\mu} \end{aligned} \tag{6}$$

so that

$$\begin{aligned} \Pr\{\mathbf{z} \in Q^{(s)}\} &= \Pr\{\mathbf{a}_i^{(s)\prime}\mathbf{z} \geqslant 0, 1 \leqslant i \leqslant m\} \\ &= P_m(\boldsymbol{\mu}^{(s)}, \mathbf{R}^{(s)}), \end{aligned}$$

the result of this section is summarized in the following theorem.

Theorem 1. Suppose that the positive definite correlation matrix $\mathbf{R} = \{\rho_{ij}\}$ has orthoscheme order r with $0 \leqslant r \leqslant m-3$. Then, for any vector $\boldsymbol{\mu}$,

$$P_m(\boldsymbol{\mu}, \mathbf{R}) = \sum_{\gamma_s>0} P_m(\boldsymbol{\mu}^{(s)}, \mathbf{R}^{(s)}) - \sum_{\gamma_s<0} P_m(\boldsymbol{\mu}^{(s)}, \mathbf{R}^{(s)}), \tag{7}$$

where γ_s, $r+2 \leqslant s \leqslant m$, are given by expression (4) and $\mathbf{R}^{(s)}$ and $\boldsymbol{\mu}^{(s)}$ by expression (6).

The important point of theorem 1 is that the correlation matrices $\mathbf{R}^{(s)}$ each have an orthoscheme order that is strictly larger than the orthoscheme order of the correlation matrix $\mathbf{R}$.

Specifically, if $\mathbf{R}$ has orthoscheme order $m-3$ then the matrices $\mathbf{R}^{(s)}$ are all tridiagonal matrices. Since there are at most $m-r-1$ non-zero elements $\gamma_s, r+2 \leqslant s \leqslant m$, it follows that with the repeated application of theorem 1 any non-centred orthant probability $P_m(\boldsymbol{\mu}, \mathbf{R})$ can eventually be dissected into differences of at most $(m-1)!$ non-centred orthoscheme probabilities. It should also be pointed out that the formulae for $\mathbf{R}^{(s)}$ and $\boldsymbol{\mu}^{(s)}$ given in theorem 1 provide a practical algorithm to perform such a dissection.

The recursive integration method described in Section 2 allows the quick and accurate evaluation of orthoscheme probabilities for any value of m. Taken in conjunction with the dissection involving at most $(m-1)!$ orthoscheme terms presented in this section, the procedure proposed therefore allows the accurate evaluation of the non-centred orthant probability $P_m(\boldsymbol{\mu}, \mathbf{R})$ for any correlation matrix $\mathbf{R}$.

4. Numerical results

Since the dissection given in Section 3 for expressing a non-centred orthant probability $P_m(\boldsymbol{\mu}, \mathbf{R})$ as differences between non-centred orthoscheme probabilities is exact, the numerical accuracy of this procedure for evaluating multivariate normal probabilities depends only on the numerical accuracy of the evaluations of the orthoscheme probabilities. To investigate the overall numerical accuracy of the procedure proposed some computations are presented in this section for situations in which the exact value of the orthant probability can be evaluated by other means. The computations were implemented in the C language on a personal computer with an Intel® Pentium® 4 processor at 1.5 GHz.

4.1. Centred orthant probabilities

First consider a correlation matrix $\mathbf{R}$ with $\rho_{ij} = \frac{1}{2}, i \neq j$. In this case it is known that $P_m(\mathbf{0}, \mathbf{R}) = 1/(m+1)$. Additionally, consider a covariance matrix $\boldsymbol{\Sigma}$ whose inverse $\boldsymbol{\Sigma}^{-1} = \{\sigma^{ij}\}$ has elements

$$\begin{aligned} \sigma^{ii} &= 1, \\ \sigma^{i,i+1} = \sigma^{i+1,i} &= -\tfrac{1}{2}, \\ \sigma^{ij} = 0, \qquad & |i-j| > 1; \end{aligned}$$

in this case Anis and Lloyd (1953) proved that the corresponding centred orthant probability $P_m(\mathbf{0}, \mathbf{R})$ is also equal to $1/(m+1)$. The correlation matrix corresponding to $\boldsymbol{\Sigma}$ has all entries non-zero, although when the dissection in Section 3 is applied the number of orthoscheme probabilities required is considerably less than $(m-1)!$. For example, the number of orthoscheme probabilities in the dissection is 323 for $m = 9$ compared with $(9-1)! = 40320$. To investigate the effect of accumulating orthoscheme probabilities on the overall accuracy, these two centred orthant probabilities were calculated by the procedure proposed for the case $m = 9$ where the known exact probability is 0.1. The results are given in Table 2. It can be seen that in both cases the values obtained agree very closely with the exact value.

4.2. Non-centred orthant probabilities

Additional calculations of non-centred orthant probabilities $P_m(\boldsymbol{\mu}, \mathbf{R})$ were performed for equicorrelation cases with $\rho_{ij} = \rho, i \neq j$, and $\mu_i = \mu, 1 \leqslant i \leqslant m$. When the dissection in Section 3 is applied it can be shown that $(m-1)!$ orthoscheme probabilities are required, but that they are all identical. These orthant probabilities can also be calculated by a one-dimensional

Table 2. Numerical calculations of orthant probabilities†

G	*Probabilities for the following values of* ρ *or* σ*:*	
	$\rho_{ij} = \frac{1}{2}$	$\sigma^{ii} = 1,\ \sigma^{i,i+1} = -\frac{1}{2}$
32	0.0999751100 (12.61)	0.1000142803 (0.08)
64	0.0999990004 (23.01)	0.1000006159 (0.16)
128	0.0999999489 (44.02)	0.1000000321 (0.30)
256	0.0999999968 (87.49)	0.1000000020 (0.59)
512	0.0999999998 (172.12)	0.1000000001 (1.18)
Exact	0.1000000000	0.1000000000

†G is the number of grid points. The computational time in seconds is given in parentheses.

integration of terms involving the standard normal cumulative distribution function. In Tong (1990) tables can be found of the orthant probabilities to five decimal places for $m = 2, 3, \ldots, 10$, $\mu = -2.0, -1.0, \ldots, 4.0$ and $\rho = 0.1, 0.2, \ldots, 0.9$ (as well as many other cases). Each of these orthant probabilities was calculated by the procedure proposed with $G = 128$ grid points and comparisons were made with Tong's tables. Almost all the $9 \times 7 \times 9 = 567$ entries agree completely, with the discrepancies being 11 entries which disagree by 1 unit at the fifth decimal place and two entries which disagree by 2 units at the fifth decimal place. However, these last two entries agree with values tabulated by Gupta (1963) (Table 3).

Overall, these computations indicate that the numerical accuracy of the procedure proposed in this paper is satisfactory for both centred and non-centred orthant probabilities.

4.3. A note on computational time

The computational time for calculating an orthoscheme probability is proportional to $m \times G$ where G is the number of grid points in the recursive method described in Section 2. Consequently, if an orthant probability $P_m(\mu, \mathbf{R})$ is dissected into $(m-1)!$ orthoscheme probabilities, the total computational time is proportional to $m! \times G$. Table 2 confirms that the computational time increases linearly in G. It can also be observed from Tables 1 and 2 that an accuracy of at least one more decimal place is obtained by doubling the number of grid points.

The computational time was investigated for various values of m in the equicorrelated case $\rho_{ij} = \frac{1}{2}$ where $(m-1)!$ orthoscheme probabilities are needed (Table 4). It can be seen that the computational time increases at a rate of $m!$ and so the computations will become prohibitive for large m. However, if the correlation matrix has a special structure so that the dissection

Table 3. Comparison of orthant probabilities

Parameters	*Probabilities from the following sources:*		
	Proposed procedure	*Tong (1990)*	*Gupta (1963)*
$m = 9,\ \rho = 0.9,\ \mu = 0.0$	0.31380	0.31382	0.31380
$m = 10,\ \rho = 0.9,\ \mu = 0.0$	0.30747	0.30749	0.30747

Table 4. Comparisons of computational time for known exact probabilities†

m	*Results for our procedure and the following values of G:*				*Results for Genz's procedure and the following values of ε:*			
	$G = 16$		$G = 128$		$\varepsilon = 0.001$		$\varepsilon = 0.0001$	
	Time (s)	*Absolute error*	*Time (s)*	*Absolute error*	*Time (s)*	*Absolute error*	*Time (s)*	*Absolute error*
5	0.00	2.1×10^{-6}	0.01	1.4×10^{-9}	0.68	2.1×10^{-4}	59.67	3.5×10^{-5}
6	0.01	3.5×10^{-6}	0.08	2.0×10^{-10}	0.74	2.7×10^{-4}	97.86	2.2×10^{-5}
7	0.07	9.2×10^{-6}	0.58	5.8×10^{-9}	1.14	3.5×10^{-4}	100.22	2.7×10^{-5}
8	0.58	5.7×10^{-5}	4.76	2.1×10^{-8}	1.10	2.5×10^{-4}	97.70	3.3×10^{-5}
9	5.35	1.6×10^{-4}	44.02	5.1×10^{-8}	1.55	2.8×10^{-4}	137.62	3.6×10^{-5}
10	53.99	3.5×10^{-4}	450.46	1.0×10^{-7}	1.39	2.1×10^{-4}	124.86	2.3×10^{-5}
20					1.85	2.8×10^{-4}	198.90	3.3×10^{-5}

†Equicorrelated correlation matrices: $\rho_{ij} \equiv \rho = \frac{1}{2}$, $\mu_i = 0.0$. G is the number of grid points. The value of ε is used by Genz's procedure to stop the repetition (Genz, 1992).

procedure in Section 3 gives rise to a number of orthoscheme probabilities that are smaller than $(m-1)!$, then our procedure may still be applied to cases with large m.

Another possible approach to orthant probability calculations is a Monte Carlo method. Genz (1992) proposed a Monte Carlo procedure combined with transformations of the integral variables. His method was evaluated for the same set of orthant probabilities (Table 4). The results obtained suggest that the computational time of Genz's method is no more than a few seconds for m as large as 20, as long as the absolute error is less than 10^{-3}. However, one major difficulty of the Monte Carlo method is that the computational time increases by about 100 times to achieve an increase of one decimal place in the accuracy. The computational time was also compared for randomly generated correlation matrices and mean vectors where exact probabilities are unknown (Table 5). Our procedure is faster for moderate values of m (e.g. $m \leqslant 7$). Thus we can expect that enough accuracy can be obtained by increasing the number of grid points G, because the computational time is linear in G.

5. Concluding remarks

We have shown how the non-centred orthant probability $P_m(\boldsymbol{\mu}, \mathbf{R}) = \Pr\{x_i \geqslant 0, 1 \leqslant i \leqslant m\}$ of a multivariate normal random variable $\mathbf{x} \sim N_m(\boldsymbol{\mu}, \mathbf{R})$ can be evaluated accurately for any correlation matrix $\mathbf{R}$ and any mean vector $\boldsymbol{\mu}$. This enables us to calculate the multivariate normal distribution function $F_m(\mathbf{c}) = \Pr\{x_i \leqslant c_i, 1 \leqslant i \leqslant m\}$. Furthermore, it is known that the two-sided probability

$$F_m(\mathbf{c}, \mathbf{d}) = \Pr\{c_i \leqslant x_i \leqslant d_i, 1 \leqslant i \leqslant m\}$$

can be expressed in terms of 2^m non-centred orthant probabilities.

The procedure proposed is based on the recursive integration algorithm to evaluate non-centred orthoscheme probabilities, and we have demonstrated that, making use of the tridiagonal nature of the correlation matrix, the computational intensity is only linear in the dimension m.

Table 5. Computational time comparisons for unknown exact probabilities†

m	*Results for our procedure‡, $\varepsilon = 0.001$*				*Results for Genz's procedure, $\varepsilon = 0.001$*			
	Computational time (s)		*Absolute error § ($\times 10^{-4}$)*		*Computational time (s)*		*Absolute error § ($\times 10^{-4}$)*	
	Average §§	*SD §*	*Average §§*	*SD §§*	*Average §§*	*SD §§*	*Average §§*	*SD §§*
5	0.04	0.04	2.4	7.1	5.62	3.80	2.8	1.7
6	0.58	0.26	2.4	3.7	8.15	5.50	3.1	2.1
7	4.54	1.87	2.2	2.8	9.15	5.20	2.9	2.3
8	33.38	9.88	5.0	9.8	12.70	7.98	3.4	2.9

†Random correlation matrices and random mean vectors.
‡Our procedure starts with $G = 16$ grid points. The calculations are repeated by doubling the number of grid points G until the absolute difference of two consecutively calculated probabilities becomes smaller than $\varepsilon/2$.
§The absolute error is obtained against the orthant probability calculated by our procedure with $G = 1024$ grid points.
§§The average and standard deviation SD are calculated over 20 sets of randomly generated correlation matrices and mean vectors.

It is also interesting that the two-sided probability $F_m(\mathbf{c}, \mathbf{d})$ for a tridiagonal matrix $\mathbf{R}$ can also be evaluated directly by a recursive algorithm similar to that described in Section 2.

Finally, computer programs to implement the procedure proposed have been coded in the C language and are available from the first author.

Acknowledgements

The authors gratefully acknowledge the Joint Editor, the Associate Editor and the referees for their helpful comments and suggestions.

Appendix A: Proof of lemma 1

A.1. Proof of part (a)

Consider any $\lambda_i \geqslant 0, 1 \leqslant i \leqslant m$, and suppose that $\min_{\gamma_s>0}(\lambda_s/\gamma_s) = \lambda_t/\gamma_t$ with $\gamma_t > 0$. Then $\mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_m\mathbf{v}_m \in Q$ can be written as

$$\begin{aligned}\mathbf{z} &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_t\mathbf{v}_t + \ldots + \lambda_m\mathbf{v}_m \\ &= \lambda_1\mathbf{v}_1 + \ldots + \frac{\lambda_t}{\gamma_t}\left(\mathbf{p} - \sum_{i=r+2,\, i\neq t}^{m} \gamma_i\mathbf{v}_i\right) + \ldots + \lambda_m\mathbf{v}_m\end{aligned}$$

where the coefficient of $\mathbf{p}$ is $\lambda_t/\gamma_t \geqslant 0$, the coefficients of $\mathbf{v}_i$, $1 \leqslant i \leqslant r+1$, are $\lambda_i \geqslant 0$ and the coefficients of $\mathbf{v}_i$, $r+2 \leqslant i \leqslant m, i \neq t$, are $\lambda_i - \lambda_t\gamma_i/\gamma_t \geqslant 0$. Thus $\mathbf{z} \in Q^{(t)}$ with $\gamma_t > 0$, and this implies that $Q \subseteq \cup_{\gamma_s>0} Q^{(s)}$. Furthermore,

$$\mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \lambda_m\mathbf{v}_m \in Q^{(s)}$$

for $\gamma_s < 0$ can be written as

$$\mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \frac{\lambda_t}{\gamma_t}\left(\mathbf{p} - \sum_{i=r+2,\, i\neq t}^{m} \gamma_i\mathbf{v}_i\right) + \ldots + \lambda_m\mathbf{v}_m$$

where the coefficient of $\mathbf{p}$ is $\lambda_s + \lambda_t/\gamma_t \geqslant 0$, the coefficients of $\mathbf{v}_i$, $1 \leqslant i \leqslant r+1$, are $\lambda_i \geqslant 0$, the coefficient of $\mathbf{v}_s$ is $-\lambda_t\gamma_s/\gamma_t \geqslant 0$ and the coefficients of $\mathbf{v}_i$, $r+2 \leqslant i \leqslant m$, $i \neq s$, $i \neq t$, are $\lambda_i - \lambda_t\gamma_i/\gamma_t \geqslant 0$. Consequently, $\mathbf{z} \in Q^{(t)}$ with $\gamma_t > 0$, and this implies that $\cup_{\gamma_s<0} Q^{(s)} \subseteq \cup_{\gamma_s>0} Q^{(s)}$.

In addition,

$$\mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \lambda_m\mathbf{v}_m \in Q^{(s)}$$

with $\lambda_i \geqslant 0$, $1 \leqslant i \leqslant m$, for $\gamma_s > 0$ can be written as

$$\mathbf{z} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\left(\sum_{i=r+2}^{m} \gamma_i\mathbf{v}_i\right) + \ldots + \lambda_m\mathbf{v}_m.$$

In this expression the coefficients of $\mathbf{v}_i$, $1 \leqslant i \leqslant r+1$, are $\lambda_i \geqslant 0$, and the coefficient of $\mathbf{v}_s$ is $\lambda_s\gamma_s \geqslant 0$. The coefficients of $\mathbf{v}_i$, $r+2 \leqslant i \leqslant m$, $i \neq s$, are $\lambda_i + \lambda_s\gamma_i$. Let W be the set of indices i, $r+2 \leqslant i \leqslant m$, $i \neq s$, for which this coefficient is strictly negative. If $W = \emptyset$ then $\mathbf{z} \in Q$. However, if $W \neq \emptyset$ notice that $\gamma_i < 0$ for $i \in W$ and suppose that $\max_{i\in W}(\lambda_i/\gamma_i) = \lambda_w/\gamma_w$ with $\gamma_w < 0$. Then $\mathbf{z}$ can be expressed as

$$\begin{aligned}\mathbf{z} &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \lambda_w\mathbf{v}_w + \ldots + \lambda_m\mathbf{v}_m \\ &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \frac{\lambda_w}{\gamma_w}\left(\mathbf{p} - \sum_{i=r+2,\, i\neq w}^{m} \gamma_i\mathbf{v}_i\right) + \ldots + \lambda_m\mathbf{v}_m, \qquad (8)\end{aligned}$$

where the coefficient of $\mathbf{p}$ is $\lambda_s + \lambda_w/\gamma_w = (\lambda_w + \lambda_s\gamma_w)/\gamma_w \geqslant 0$, the coefficients of $\mathbf{v}_i$, $1 \leqslant i \leqslant r+1$, are $\lambda_i \geqslant 0$, the coefficient of $\mathbf{v}_s$ is $-\lambda_w\gamma_s/\gamma_w \geqslant 0$ and the coefficients of $\mathbf{v}_i$, $r+2 \leqslant i \leqslant m$, $i \neq s$, $i \neq w$, are $\lambda_i - \lambda_w\gamma_i/\gamma_w$. If $\gamma_i \geqslant 0$ then clearly $\lambda_i - \lambda_w\gamma_i/\gamma_w \geqslant 0$, and if $i \in W$ then the coefficient is similarly non-negative by the definition of w. Finally, if $\gamma_i < 0$ and $i \notin W$ then the coefficient can also be seen to be non-negative by writing it as

$$\lambda_i - \frac{\lambda_w\gamma_i}{\gamma_w} = \lambda_i + \lambda_s\gamma_i - \frac{\gamma_i}{\gamma_w}(\lambda_w + \lambda_s\gamma_w) \geqslant 0.$$

Therefore all the coefficients in expression (8) for $\mathbf{z}$ are non-negative, which implies that if $W \neq \emptyset$ then $\mathbf{z} \in Q^{(w)}$ with $\gamma_w < 0$. Consequently, it has been shown that $\cup_{\gamma_s>0} Q^{(s)} \subseteq Q \cup (\cup_{\gamma_s<0} Q^{(s)})$ and it has also been established that

$$Q \cup \left(\bigcup_{\gamma_s<0} Q^{(s)}\right) = \bigcup_{\gamma_s>0} Q^{(s)}.$$

A.2. Proof of part (b)

Suppose that γ_s and γ_t, $s \neq t$, have the same sign. Then if $\mathbf{z} \in Q^{(s)} \cap Q^{(t)}$ it can be expressed as

$$\begin{aligned}\mathbf{z} &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{p} + \ldots + \lambda_m\mathbf{v}_m \\ &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_s(\gamma_{r+2}\mathbf{v}_{r+2} + \ldots + \gamma_m\mathbf{v}_m) + \ldots + \lambda_m\mathbf{v}_m \\ &= \xi_1\mathbf{v}_1 + \ldots + \xi_t\mathbf{p} + \ldots + \xi_m\mathbf{v}_m \\ &= \xi_1\mathbf{v}_1 + \ldots + \xi_t(\gamma_{r+2}\mathbf{v}_{r+2} + \ldots + \gamma_m\mathbf{v}_m) + \ldots + \xi_m\mathbf{v}_m,\end{aligned}$$

where $\lambda_i \geqslant 0$ and $\xi_i \geqslant 0$ for $1 \leqslant i \leqslant m$. Equating the coefficients of $\mathbf{v}_s$ and $\mathbf{v}_t$ gives

$$\begin{aligned}\lambda_s\gamma_s &= \xi_s + \xi_t\gamma_s, \\ \lambda_t + \lambda_s\gamma_t &= \xi_t\gamma_t,\end{aligned}$$

and hence $\lambda_t\gamma_s = -\xi_s\gamma_t$. Since γ_s and γ_t are non-zero and have the same sign, this requires that $\lambda_t = \xi_s = 0$. Therefore, $Q^{(s)} \cap Q^{(t)}$ is contained within a subspace of dimension $m-1$.

A.3. Proof of part (c)

If $\mathbf{z} \in Q \cap Q^{(s)}$ for $\gamma_s < 0$ it can be expressed as

$$\begin{aligned}\mathbf{z} &= \lambda_1\mathbf{v}_1 + \ldots + \lambda_s\mathbf{v}_s + \ldots + \lambda_m\mathbf{v}_m \\ &= \xi_1\mathbf{v}_1 + \ldots + \xi_s\mathbf{p} + \ldots + \xi_m\mathbf{v}_m \\ &= \xi_1\mathbf{v}_1 + \ldots + \xi_s(\gamma_{r+2}\mathbf{v}_{r+2} + \ldots + \gamma_m\mathbf{v}_m) + \ldots + \xi_m\mathbf{v}_m,\end{aligned}$$

where $\lambda_i \geqslant 0$ and $\xi_i \geqslant 0$ for $1 \leqslant i \leqslant m$. Now equating the coefficients of $\mathbf{v}_s$ gives $\lambda_s = \xi_s \gamma_s$, which since $\gamma_s < 0$ implies that $\lambda_s = \xi_s = 0$, so $Q \cap Q^{(s)}$ is also contained within a subspace of dimension $m - 1$.

References

Abrahamson, I. G. (1964) Orthant probabilities for the quadrivariate normal distribution. *Ann. Math. Statist.*, **35**, 1685–1703.

Anis, A. A. and Lloyd, E. H. (1953) On the range of partial sums of a finite number of independent normal variates. *Biometrika*, **40**, 35–42.

Genz, A. (1992) Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, **1**, 141–149.

Gupta, S. S. (1963) Probability integrals of multivariate normal and multivariate t. *Ann. Math. Statist.*, **34**, 792–828.

Hayter, A. J. and Liu, W. (1996) A note on the calculation of $\Pr\{X_1 < X_2 < \cdots < X_k\}$. *Am. Statistn*, **50**, 365.

Johnson, N. L. and Kotz, S. (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Miwa, T., Hayter, A. J. and Liu, W. (2000) Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models. *Comput. Statist. Data Anal.*, **34**, 17–32.

Moran, P. A. P. (1983) A new expansion for the multivariate normal distribution. *Aust. J. Statist.*, **25**, 339–344.

Moran, P. A. P. (1985) Calculation of multivariate normal probabilities—another special case. *Aust. J. Statist.*, **27**, 60–67.

Schläfli, L. (1858) On the multiple integral $\int^n dxdy\ldots dz$, whose limits are $p_1 = a_1x + b_1y + \ldots + h_1z > 0$, $p_2 > 0, \ldots, p_n > 0$, and $x^2 + y^2 + \ldots + z^2 < 1$. *Q. J. Pure Appl. Math.*, **2**, 269–301.

Stuart, A. and Ord, J. K. (1994) *Kendall's Advanced Theory of Statistics*, vol. 1, 5th edn. London: Arnold.

Tong, Y. L. (1990) *The Multivariate Normal Distribution*. New York: Springer.