Approximations to Multivariate Normal Rectangle Probabilities Based on Conditional Expectations

Harry JOE*

Two new approximations for multivariate normal probabilities for rectangular regions, based on conditional expectations and regression with binary variables, are proposed. One is a second-order approximation that is much more accurate but also more numerically time-consuming than the first-order approximation. A third approximation, based on the moment-generating function of a truncated multivariate normal distribution, is proposed for orthant probabilities only. Its accuracy is between the first- and second-order approximations when the dimension is less than seven and the correlations are not large. All of the approximations get worse as correlations get larger. These new approximations offer substantial improvements on previous approximations. They also compare favorably with the methods of Genz for numerical evaluation of the multivariate normal integral. The approximation methods should be especially useful within a quasi-Newton routine for parameter estimation in discrete models that involve the multivariate normal distribution.

KEY WORDS: Computational complexity; Moment-generating function; Multidimensional numerical quadrature.

1. INTRODUCTION

Rectangle and orthant probabilities from the multivariate normal distribution have many applications in statistics. These include the multivariate probit model (e.g., Ochi and Prentice 1984), the multivariate ordinal response model (e.g., Anderson and Pemberton 1985), multivariate paired comparisons (e.g., Böckenholt 1992a) and Thurstonian models for rankings (e.g., Böckenholt 1992b). In general these probabilities require multidimensional integrals, which can be evaluated by multidimensional numerical quadrature or Monte Carlo simulation; alternatively, approximations can be used if they are good enough. If estimation of parameters in one of the foregoing models is obtained using a quasi-Newton routine applied to the log-likelihood, then the use of Monte Carlo simulation to three or four decimal place accuracy for evaluations of integrals works poorly, because numerical derivatives of the log-likelihood with respect to parameters are not smooth. Until the recent work of Genz (1992), the use of numerical quadrature was too time-consuming for dimensions greater than five. Numerical quadrature has computational time (and memory requirements) exponentially increasing in the dimension of the integral, so faster approximations are useful for parameter estimation (in the foregoing models). For example, a faster approximation method can be used to get preliminary estimates of the parameters; then, if necessary, one can use a better approximation method or numerical quadrature, taking the preliminary estimate as a starting point, when computing the likelihood more accurately.

Recent approximations for special cases have been given by Solow (1990) and Dresner (1990). Earlier approximations are mentioned in the book by Tong (1990), including one by Mendell and Elston (1974) that is mentioned later in this article. A decomposition into a product of conditional probabilities [see (2)] is used for approximations in some of these cited articles. We obtain some new (and better) approximations of the conditional probabilities. In Section 2 we first follow up on and improve on some ideas of Solow (1990) to get approximations for probabilities of rectangular regions. The main ideas are quite simple and depend on conditional expectations and regression for binary variables; the form of the multivariate normal distribution is not used, so that the approximations are also usable for other multivariate normal distributions. First-order and second-order approximations are obtained. The approximations are better when the correlations are smaller. The first-order approximation makes use of all of the univariate and bivariate marginal probabilities, and the second-order approximation makes use of trivariate and four-variate marginal probabilities as well. The second-order approximation generally has four or almost four digits of accuracy up to 12 dimensions if the correlations are not too large and three digits of accuracy if the correlations are large. The first-order approximation is much faster but has one decimal place less accuracy compared to the second-order approximation.

Also in Section 2, an approximation to conditional probabilities based on the moment-generating function of a truncated multivariate normal distribution is given. This approximation is for orthant probabilities only. The accuracy is worse than the second-order approximation mentioned previously and better than the first-order approximation if the correlations are not too large and the dimension m is less than 8. The computational complexity for this method increases exponentially with dimension, so that it is reasonable to use only for dimensions up to seven.

Numerical examples and comparisons of accuracy and computational complexity for all three approximations are given in Section 3, along with comparisons with some previously proposed approximations and the evaluation methods of Genz (1992).

2. APPROXIMATIONS

2.1 Binary Variables Approach

We start with some notation. Let $m \geq 3$ be the dimension and let $(X_1, \ldots, X_m)$ be a multivariate normal random vector

* Harry Joe is Professor, Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2. This research has been supported by an NSERC Canada grant. The author is grateful to the referee and associate editor for their comments and for the reference to the work of Genz.

with zero mean vector, variances of 1, and correlation matrix $\mathbf{R} = (r_{ij})$. We are interested in the probability

$$\Pr(w_1 < X_1 \le x_1, \ldots, w_m < X_m \le x_m), \quad (1)$$

which can be decomposed as the product of conditional probabilities,

$$\Pr(w_1 < X_1 \le x_1, w_2 < X_2 \le x_2) \prod_{k=3}^{m} \Pr(w_k < X_k \le x_k | w_1 < X_1 \le x_1, \ldots, w_{k-1} < X_{k-1} \le x_{k-1}). \quad (2)$$

Let $I_i = I(w_i < X_i \le x_i)$, $i = 1, \ldots, m$ where $I(A)$ denotes the indicator of the event A. Note that $E(I_i) = \Phi(x_i) - \Phi(w_i)$, where Φ is the univariate normal cumulative distribution function.

The first step is an approximation of

$$\Pr(w_k < X_k \le x_k | w_1 < X_1 \le x_1, \ldots, w_{k-1} < X_{k-1} \le x_{k-1}) = E(I_k | I_1 = 1, \ldots, I_{k-1} = 1) \quad (3)$$

by

$$E(I_k) + \Omega_{21}\Omega_{11}^{-1}(1 - E(I_1), \ldots, 1 - E(I_{k-1}))^T, \quad (4)$$

where Ω_{21} is a row vector consisting of the entries $\operatorname{cov}(I_k, I_i) = E(I_k I_i) - E(I_k)E(I_i)$, $i = 1, \ldots, k-1$, and Ω_{11} is a $(k-1) \times (k-1)$ matrix with (i, j) element $\operatorname{cov}(I_i, I_j) = E(I_i I_j) - E(I_i)E(I_j)$, $1 \le i, j \le k-1$. Note that $E(I_i I_j) = \Pr(w_i < X_i \le x_i, w_j < X_j \le x_j)$ is a bivariate marginal probability and that it is easily verified that (3) and (4) are identical if $k = 2$. The use of (4) as an approximation to (3) is analogous to the formula

$$E(\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$$

for a multivariate normal random vector $(\mathbf{Y}_1, \mathbf{Y}_2)$ with mean vector $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and covariance matrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Expression (4) can be substituted into (2) to get one approximation to (1). But the decomposition into conditional probabilities is not unique, and different decompositions lead to different approximations in general. That is, (1) is also equivalent to

$$\Pr(w_{i_1} < X_{i_1} \le x_{i_1}, w_{i_2} < X_{i_2} \le x_{i_2}) \prod_{k=3}^{m} \Pr(w_{i_k} < X_{i_k} \le x_{i_k} | w_{i_1} < X_{i_1} \le x_{i_1}, \ldots, w_{i_{k-1}} < X_{i_{k-1}} \le x_{i_{k-1}}), \quad (5)$$

where $(i_1, \ldots, i_m)$ is a permutation of $(1, \ldots, m)$ with $i_1 < i_2$. There are $m!/2$ permutations that could be considered. For example, for $m = 3$,

$$\begin{aligned}
&\Pr(w_1 < X_1 \le x_1, w_2 < X_2 \le x_2, w_3 < X_3 \le x_3) \\
&\quad = \Pr(w_1 < X_1 \le x_1, w_2 < X_2 \le x_2) \\
&\qquad \times \Pr(w_3 < X_3 \le x_3 | w_1 < X_1 \le x_1, w_2 < X_2 \le x_2) \\
&\quad = \Pr(w_1 < X_1 \le x_1, w_3 < X_3 \le x_3) \\
&\qquad \times \Pr(w_2 < X_2 \le x_2 | w_1 < X_1 \le x_1, w_3 < X_3 \le x_3) \\
&\quad = \Pr(w_2 < X_2 \le x_2, w_3 < X_3 \le x_3) \\
&\qquad \times \Pr(w_1 < X_1 \le x_1 | w_2 < X_2 \le x_2, w_3 < X_3 \le x_3).
\end{aligned}$$

For each permutation, an approximation of the form (4) can be used for each conditional probability. An overall approximation, denoted by P_1, for (1) is the average of the $m!/2$ approximations; the standard deviation of the $m!/2$ approximations gives a rough measure of precision of the overall approximation. For the average P_1, if (4) happens to exceed 1 or be less than zero, it is replaced by 1 and zero. Actually, it is not necessary to take all permutations if m exceeds 6 or 7; 10^2 to 10^4 randomly selected permutations can control the standard error associated with the average of permutations.

Solow (1990) used (1)–(4) for the case $w_1 = \cdots = w_m = -\infty$, $x_1 = \cdots = x_m = x$ and did not consider the last step of averaging over permutations. According to Solow (1990), "experiments with approximating the multivariate normal distribution function for arbitrary arguments have given mixed results." The problems of Solow have been overcome here using the averaging over permutations, especially with the second-order approximation given next.

A second-order improved approximation, which makes use of trivariate and four-variate marginal probabilities, comes from writing (3) as an expectation conditioning on second-order interaction terms; that is,

$$E(I_k | I_1 = 1, \ldots, I_{k-1} = 1) = E(I_k | I_i = 1, i = 1, \ldots, k-1; I_{ij} = 1, 1 \le i < j \le k-1), \quad (6)$$

where $I_{ij} = I_i I_j$, and then regressing. The extension of (4) is then

$$E(I_k) + \Omega_{21}^*(\Omega_{11}^*)^{-1}(1 - E(I_1), \ldots, 1 - E(I_{k-1}), 1 - E(I_{12}), \ldots, 1 - E(I_{k-2,k-1}))^T, \quad (7)$$

where $\Omega_{21}^* = (\Omega_{21}, \mathbf{A})$,

$$\Omega_{11}^* = \begin{bmatrix} \Omega_{11} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix},$$

$\mathbf{A}$ is a row vector with entries $E(I_k I_{ij}) - E(I_k)E(I_{ij})$, $i < j < k$; $\mathbf{B}$ is a $(k-1) \times (k-1)(k-2)/2$-dimensional matrix with entries of the form $E(I_l I_{ij}) - E(I_l)E(I_{ij})$, $1 \le l \le k-1$, $1 \le i < j \le k-1$; and $\mathbf{C}$ is a $(k-1)(k-2)/2 \times (k-1)(k-2)/2$-dimensional matrix with entries of the form $E(I_{ij}I_{i'j'}) - E(I_{ij})E(I_{i'j'})$, $1 \le i < j \le k-1$, $1 \le i' < j' \le k-1$. Note that $E(I_l I_{ij}) = \Pr(w_l < X_l \le x_l, w_i < X_i \le x_i, w_j < X_j \le x_j)$, if l, i, j are distinct, and that it is a bivariate probability if $l = i$ or $l = j$ and $i \ne j$. Also $E(I_{ij}I_{i'j'})$ is a four-variate probability if i, j, i', j' are distinct; a trivariate probability if there are three distinct integers among i, j, i', j'; and a bivariate probability if $i = i'$, $j = j'$, $i \ne j$. In the case $k = 4$ there are no four-variate probabilities involved. It has been shown using a symbolic manipulation software that (6) is identical to (7) if $k = 3$.

We now decompose (1) as

$$\Pr(w_{i_1} < X_{i_1} \le x_{i_1}, \ldots, w_{i_n} < X_{i_n} \le x_{i_n}) \times \prod_{k=n+1}^{m} \Pr(w_{i_k} < X_{i_k} \le x_{i_k} | w_{i_1} < X_{i_1} \le x_{i_1}, \ldots, w_{i_{k-1}} < X_{i_{k-1}} \le x_{i_{k-1}}), \quad (8)$$

where $(i_1, \ldots, i_m)$ is a permutation of $(1, \ldots, m)$ with $i_1 < \cdots < i_n$, and $n = 3$ for $m = 4$, $n = 4$ for $m > 4$. For each of the $m!/n!$ permutations, each conditional probability is approximated by something of the form (7). The overall approximation, denoted by P_2, is the average of the $m!/n!$ approximations. This approximation uses all three- and four-dimensional marginal probabilities for $m > 4$ and all three-dimensional marginal probabilities for $m = 4$. Similar to before, rather than using all permutations, a random set of permutations suffices for m greater than 7 or 8.

2.2 Moment-Generating Function Approach

When the rectangular region is an orthant (i.e., upper limits of ∞ for all margins or lower limits of $-\infty$ for all margins), then an approximation based on the moment-generating function of a truncated multivariate normal distribution can be used. To follow the choice of previous authors, we take upper limits of ∞; that is, we consider (3) with $x_i = \infty$ for all i to get

$$\Pr(X_k > w_k | X_1 > w_1, \ldots, X_{k-1} > w_{k-1}) \tag{9}$$

for $k \geq 2$.

An approximation to (9), which is an alternative to those in Section 2.1, is based on a normal approximation to the distribution of X_k conditional on $X_i > w_i$, $i = 1, \ldots, k-1$. The mean and variance of this conditional distribution can be obtained from the moment- or cumulant-generating function.

Suppose that $(X_1, \ldots, X_{k-1}, X_k)$ has a multivariate normal distribution with means zero, variances 1, and correlation matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12}^T & 1 \end{bmatrix} = (\rho_{ij}),$$

where $\mathbf{R}_{11}$ is a $(k-1) \times (k-1)$ matrix. Let $\bar{\Phi}(\cdot; \mathbf{R})$ denote the survival function for a multivariate normal distribution with correlation matrix $\mathbf{R}$. As a special case of results of Tallis (1961), Finney (1962), and McGill (1992), the moment-generating function for X_k, conditional on $X_i > w_i$, $i = 1, \ldots, k-1$, is

$$\psi(t) = \exp\{t^2/2\}\bar{\Phi}(\mathbf{w} - t\mathbf{R}_{12}^T; \mathbf{R}_{11})/\bar{\Phi}(\mathbf{w}; \mathbf{R}_{11}), \tag{10}$$

where $\mathbf{w} = (w_1, \ldots, w_{k-1})$. Let $d(\mathbf{w}) = \bar{\Phi}(\mathbf{w}; \mathbf{R}_{11})$ and let the first- and second-order derivatives be denoted by d_i, d_{ij}, $i, j = 1, \ldots, k-1$. From the first two derivatives of $\log \psi(t)$ evaluated at zero,

$$\mu = \mu(\mathbf{w}, \mathbf{R}) = E(X_k | X_1 > w_1, \ldots, X_{k-1} > w_{k-1})$$
$$= -[d(\mathbf{w})]^{-1} \sum_{i=1}^{k-1} \rho_{ik} d_i(\mathbf{w}), \tag{11}$$

$$\sigma^2 = \sigma^2(\mathbf{w}, \mathbf{R}) = \operatorname{var}(X_k | X_1 > w_1, \ldots, X_{k-1} > w_{k-1})$$
$$= 1 - \mu^2 + [d(\mathbf{w})]^{-1} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \rho_{ik}\rho_{jk} d_{ij}(\mathbf{w}). \tag{12}$$

A normal approximation to (9) is now

$$1 - \Phi_1((w_k - \mu)/\sigma), \tag{13}$$

where Φ_1 is the univariate standard normal distribution function. In (13), the denominator term $\bar{\Phi}(\mathbf{w}; \mathbf{R}_{11})$ of μ and σ is approximated by recursion if the dimension of $\mathbf{R}_{11}$ is bigger than n for some n; see Equation (17).

Let $\phi_1(\cdot)$ denote the univariate standard normal density and let $\phi_2(\cdot; \rho)$ denote the bivariate standard normal density with correlation ρ. For $k \geq 3$, the d_{k-1} can be simplified into

$$d_{k-1}(\mathbf{w}) = -\phi_1(w_{k-1}) \times \bar{\Phi}((w_1 - \rho_{1,k-1}w_{k-1})/\sqrt{1 - \rho_{1,k-1}^2}, \ldots, (w_{k-2} - \rho_{k-2,k-1}w_{k-1})/\sqrt{1 - \rho_{k-2,k-1}^2}; \mathbf{Q}), \tag{14}$$

where $\mathbf{Q} = (q_{ij})$ is the $(k-2) \times (k-2)$ correlation matrix of partial correlations with $q_{ij} = (\rho_{ij} - \rho_{i,k-1}\rho_{j,k-1})/\sqrt{(1 - \rho_{i,k-1}^2)(1 - \rho_{j,k-1}^2)}$, $i \neq j$. For $i \neq k-1$, d_i can be obtained by interchanging subscripts in (14). Similarly, if $k = 3$, then $d_{k-2,k-1}(\mathbf{w}) = \phi_2(w_{k-2}, w_{k-1}; \rho_{k-2,k-1})$, and if $k > 3$, then

$$d_{k-2,k-1}(\mathbf{w}) = \phi_2(w_{k-2}, w_{k-1}; \rho_{k-2,k-1}) \times \bar{\Phi}(y_1, \ldots, y_{k-3}; \mathbf{T}), \tag{15}$$

where $\mathbf{T} = (\tau_{ij})$ is a $(k-3) \times (k-3)$ correlation matrix with

$$\tau_{ij} = (\rho_{ij} - [1 - \rho_{k-2,k-1}]^{-1}[\rho_{i,k-2}\rho_{j,k-2} + \rho_{i,k-1}\rho_{j,k-1} - \rho_{k-2,k-1}(\rho_{i,k-2}\rho_{j,k-1} + \rho_{i,k-1}\rho_{j,k-2})])/\sqrt{a_{ii}a_{jj}},$$

$$a_{ii} = 1 - [1 - \rho_{k-2,k-1}]^{-1} \times [\rho_{i,k-2}^2 + \rho_{i,k-1}^2 - 2\rho_{k-2,k-1}\rho_{i,k-2}\rho_{i,k-1}],$$

and

$$y_i = (w_i - [1 - \rho_{k-2,k-1}]^{-1}[(\rho_{i,k-2} - \rho_{i,k-1}\rho_{k-2,k-1})w_{k-2} + (\rho_{i,k-1} - \rho_{i,k-2}\rho_{k-2,k-1})w_{k-1}])/\sqrt{a_{ii}}.$$

For $(i, j) \neq (k-2, k-1)$ and $i \neq j$, d_{ij} can be obtained from (15) by interchanging subscripts. Finally, it is straightforward (but tedious) to show that

$$d_{ii}(\mathbf{w}) = -w_i d_i(\mathbf{w}) - \sum_{j=1, j\neq i}^{k-1} \rho_{ij} d_{ij}(\mathbf{w}).$$

There are enough details given for the formulas in (11) and (12) to be coded in a computer routine.

Now the orthant probability

$$\Pr(X_1 > w_1, \ldots, X_m > w_m) = \bar{\Phi}(w_1, \ldots, w_m; \mathbf{R}) \tag{16}$$

can be written as the product of conditional probabilities,

$$\Pr(X_1 > w_1, \ldots, X_n > w_n) \prod_{k=n+1}^{m} \Pr(X_k > w_k | X_i > w_i, i = 1, \ldots, k-1), \tag{17}$$

where n can be equal to 2, 3, or 4. The first unconditional probability in (17) is computed (exactly) and in general the kth conditional probability in (17) is approximated by (13), where the orthant probabilities of μ, σ in (13) are based on approximations to probabilities of the form (16) with m replaced by a smaller integer m' if $m' > n$. (The probabilities

Table 1. Comparison of Accuracy of Approximations in the Equicorrelated Case

m	ρ	w	Exact	M-E	P_1	P_2	P_3
5	.1	0	.05286	.05286	.05287	.05287	.05287
		−.2	.09574	.09576	.09588	.09574	.09575
		−.4	.15877	.15881	.15907	.15878	.15879
		−.6	.24261	.24268	.24302	.24262	.24264
		−.8	.34393	.34401	.34439	.34393	.34397
5	.4	0	.13419	.13542	.13423	.13423	.13429
		−.2	.19603	.19789	.19697	.19606	.19621
		−.4	.27216	.27457	.27405	.27221	.27263
		−.6	.36017	.36285	.36283	.36024	.36100
		−.8	.45575	.45826	.45888	.45583	.45679
9	.1	0	.00955	.00953	.00957	.00956	.00955
		−.2	.02365	.02363	.02385	.02367	.02366
		−.4	.05157	.05156	.05210	.05160	.05159
		−.6	.09979	.09984	.10075	.09983	.09985
		−.8	.17289	.17301	.17424	.17294	.17301
9	.4	0	.06876	.06947	.06888	.06888	.06878
		−.2	.11100	.11274	.11274	.11120	.11119
		−.4	.16887	.17195	.17269	.16922	.16984
		−.6	.24277	.24719	.24878	.24330	.24517
		−.8	.33080	.33612	.33867	.33152	.33478

in μ, σ are computed exactly if $m' \leq n$.) Hence (16) involves a recursive approximation. The approximation is better and takes longer to compute as n increases. We choose $n = 3$ for the numerical study in Section 3, because substantially more computational time is needed for $n = 4$.

Similar to the approximations in Section 2.1, the approximation to (16) is different depending on the order of the indices or conditional probabilities; compare (5) and (8). Again, an overall approximation can be made by averaging over permutations of indices, and the standard deviation over different permutations gives an assessment of the accuracy. Let the overall approximation by the average be denoted by P_3.

3. COMPARISONS AND COMPUTATIONAL EXPERIENCE

In this section we make comparisons of the three approximations given in Section 2 with each other, with other previously proposed approximations, and with the evaluation methods of Genz (1992). Accuracy comparisons are made followed by comparisons of computing time and computational complexity.

We briefly mention other previous approximations defined for orthant probabilities only. The approach of Mendell and Elston (974) and Rice, Reich, Cloninger, and Wette (1979) has a normal approximation like (13) that uses μ and σ from an approximation not based on the moment-generating function. (Mendell and Elston and Rice et al. realized that different approximations come from different permutations but did not take the next step of averaging over the permutations.) For $k = m = 3$, the resulting Mendell–Elston approximation to (1) can sometimes be better than that based on (13) when the w_i's are negative and the correlations are large and positive; but generally, it is worse and can be quite bad if there are some negative correlations. As the Mendell–Elston approximation does not depend on multivariate normal orthant probabilities, it is computationally faster. It is compared for accuracy with the approximations of Section 2 in Table 1. This table shows that the Mendell–Elston approximation is generally worse than both P_2 and P_3.

The cases in Table 1 are a subset of those in a table by Ochi and Prentice (1984); these are equicorrelated upper orthant probabilities of the form $\Pr(X_i > w, i = 1, \ldots, m)$ for dimensions $m = 5, 9$ and correlations $\rho = .1, .4$. Actually, in the positive equicorrelated case the probability can be obtained through a one-dimensional integral, and the approximations in this article are not needed (see Johnson and Kotz 1972, p. 48, and Tong 1990). But Table 1 does provide a summary that is indicative of what happens in the case where the ρ_{ij}'s are not all the same and where the w_i's, x_i's are not all the same.

Approximations for orthant probabilities were also given by Drezner (1990) and Clark (1961). For the examples given in table 5 of Drezner's paper, P_2 and P_3 are better. As Drezner's approximation is only for orthant probabilities and by its construction is good only for small correlations, it will not be considered further. Clark's method is based on approximating the distribution of maxima of jointly multivariate normal random variables by a normal distribution. Some accuracies for this method are given in section 7.9 of Lerman and Manski (1982); there is not even two-decimal-place accuracy in many cases for $m = 3$ or 5, so this approximation method is worse than the ones proposed in this article.

The accuracy and computational time of the approximations depend on the dimension m and the degree of dependence in the correlation matrix. Correlation matrices used for our comparisons are summarized in Table 2. For space considerations, we use patterned matrices, mainly matrices of stationary autoregressive (AR) time series in which the correlation depends on the lag. In Table 2, AR(1) means the correlations have the form $\rho_{ij} = \rho^{|i-j|}$ for the parameter ρ given. For AR(r) with $r > 1$, the correlations are given for

Table 2. Cases of Correlation Matrices Used for Comparisons

Case	m	Type	Correlations
5.1	5	AR (4)	.2, .2, .2, .1
5.2	5	AR (1)	.5
5.3	5	AR (1)	.8
5.4	5	AR (4)	.8, .7, .6, .6
5.5	5	exch	−.15
5.6	5	PM	.1, .55
6.1	6	AR (5)	.2, .2, .2, .1, .1
6.2	6	AR (1)	.5
6.3	6	AR (1)	.8
6.4	6	AR (5)	.8, .7, .6, .6, .6
6.5	6	exch	−.15
6.6	6	PM	.1, .5
7.1	7	AR (6)	.2, .2, .2, .1, .1, .1
7.2	7	AR (1)	.5
7.3	7	AR (1)	.8
7.4	7	AR (6)	.8, .7, .7, .6, .6, .6
7.5	7	exch	−.15
7.6	7	PM	.1, .48
10.1	10	AR (9)	.2, .2, .2, .2, .1, .1. .1, .1, .1
10.2	10	AR (1)	.5
10.3	10	AR (1)	.8
10.4	10	AR (9)	.8, .7, .7, .7, .7, .6, .6, .6, .6
10.5	10	exch	−.10
10.6	10	PM	.1, .42
12.1	12	AR (11)	.2, .2, .2, .2, .2, .1, .1, .1, .1, .1, .1
12.2	12	AR (1)	.5
12.3	12	AR (1)	.8
12.4	12	AR (11)	.8, .8, .7, .7, .7, .7, .6, .6, .6, .6, .6
12.5	12	exch	−.085
12.6	12	PM	.1, .41

lags 1, 2, . . . , r. For a given dimension, cases 1 and 4 are low and high correlations, and cases 2 and 3 are AR(1) with medium and high correlations. Cases 5 and 6 for a given dimension are matrices that are close to being singular; case 5 is a matrix with a constant negative correlation (exchangeable), and case 6 is a patterned matrix (PM) of the form

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_1 & \rho_2 \\ \rho_1 & 1 & \cdots & \rho_1 & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_1 & \rho_1 & \cdots & 1 & \rho_2 \\ \rho_2 & \rho_2 & \cdots & \rho_2 & 1 \end{bmatrix}.$$

The nearly singular case is chosen to see how the approximations behave near the boundary of the range of correlation matrices.

Tables of accuracy and computational times are given in Tables 3–6; Tables 3 and 4 for rectangle probabilities and Tables 5 and 6 for orthant probabilities. For accuracy, the computation used for the "exact" probability is (a) for $m \le 5$, Schervish's (1984) routine with an error bound of $\varepsilon = 10^{-5}$; (b) for $m \ge 6$, Genz's (1992) transformation of the multivariate normal integrand with the double precision version of the code of Berntsen, Espelid, and Genz (1991) using an error bound of $\varepsilon = 10^{-4}$; and (c) Genz's (1992) transformation with Monte Carlo evaluation of the integral to an accuracy of $\varepsilon = 10^{-4}$, if numerical quadrature in (b) fails. For a given correlation matrix, the limits are simulated. For the rectangle probabilities in Tables 3 and 4, the random choices were $x_i \sim N(1.9, .9^2)$ and $x_i - w_i \sim N(3.5, .7^2)$ independently for each $i = 1, \ldots, m$; for the (upper) orthant probabilities in Tables 5 and 6, the random choices were $w_i \sim N(-1.5, .5^2)$, independently for each $i = 1, \ldots, m$. Tables 3 and 5 summarize the average and maximum absolute error over the random limits that were used; approximations P_1, P_2 appear in both tables, and P_3 appears only in Table 5.

Table 3 shows that the first-order approximation P_1 typically has an error in the third decimal place, with the worst cases (of large correlations) having an error in the second decimal place. The second-order approximation P_2 typically has one extra decimal place of accuracy, both in the average and maximum error. Table 5 shows that the approximation P_3 has accuracy in between that of P_1 and P_2 if the correlations are small, but that it becomes worse than P_1 in the case of large correlations. In fact, in this latter situation, P_3 often has an error in the second decimal place. A comparison of Tables 3 and 5 for P_1 and P_2 shows that the error for these methods is on average smaller for orthant probabilities, and significantly so in the cases of larger correlations.

Next we discuss computational times for the approximations. A couple of factors can affect the times. One is the desired accuracy for the lower-dimensional marginal probabilities; the other is the number of permutations used in obtaining the average over permutations of indices. For P_1, the standard deviation (SD) is usually of the order of 10^{-3} and is closer to the order of 10^{-2} in the cases of larger correlations. For P_2, the SD is about $\frac{1}{10}$ that for P_1. For P_3, the

Table 3. Average and Maximum Absolute Errors for Rectangle Probabilities

Case	# sim.	P_1: avg.	Errors max.	P_2: avg.	Errors max.
5.1	50	.001	.004	.0001	.0002
5.2	50	.002	.005	.0001	.0008
5.3	50	.005	.015	.0003	.0010
5.4	50	.004	.013	.0003	.0009
5.5	50	.001	.004	.0001	.0003
5.6	50	.003	.014	.0003	.0023
6.1	50	.001	.003	.0001	.0003
6.2	50	.002	.010	.0001	.0005
6.3	50	.004	.019	.0004	.0018
6.4	50	.006	.022	.0005	.0020
6.5	50	.002	.006	.0003	.0010
6.6	50	.003	.011	.0003	.0014
7.1	20	.001	.005	.0001	.0003
7.2	20	.002	.006	.0002	.0006
7.3	20	.004	.013	.0005	.0015
7.4	20	.007	.031	.0005	.0028
7.5	20	.003	.008	.0005	.0014
7.6	20	.003	.012	.0004	.0016
10.1	20	.001	.003	.0002	.0005
10.2	20	.001	.004	.0001	.0003
10.3	20	.005	.020	.0006	.0019
10.4	20	.006	.025	.0004	.0020
10.5	20	.002	.005	.0003	.0008
10.6	20	.001	.002	.0002	.0009
12.1	10	.002	.006	.0001	.0006
12.2	10	.001	.003	.0001	.0003
12.3	10	.004	.013	.0003	.0012
12.4	10	.008	.028	.0005	.0032
12.5	10	.001	.003	.0002	.0007
12.6	10	.001	.004	.0003	.0008

Table 4. Timings in Seconds (From SPARC 10) for Rectangle Probabilities

Case	# sim.	P_1 (10^{-5})	P_2 (10^{-5})	P_2 (10^{-4})	INT (10^{-3})	INT (10^{-4})	MC (10^{-3})
5.1	50	.05 (0)	3.6 (0.9)	1.8 (.4)	0.6 (1.0)	7 (10)	3.4 (.5)
5.2	50	.05 (0)	4.9 (1.1)	2.8 (.7)	3.5 (6.2)	13 (20)	4.4 (1.9)
5.3	50	.05 (0)	9.3 (2.2)	5.7 (1.3)	2.9 (4.4)	14 (24)	12 (11)
5.4	50	.05 (0)	9.4 (2.6)	5.8 (1.6)	3.5 (3.9)	42 (48)	15 (13)
5.5	50	.05 (0)	3.8 (1.0)	1.9 (.5)	1.3 (1.5)	11 (10)	3.2 (.4)
5.6	50	.05 (0)	5.4 (1.3)	3.2 (.8)	7.2 (7.5)	41 (32)	9.8 (9.1)
6.1	50	.3 (0)	11 (2)	5.4 (1.0)	2.0 (2.9)	35 (43)	5.1 (.4)
6.2	50	.3 (0)	13 (3)	6.9 (1.9)	9 (15)	55 (81)	6.2 (2.4)
6.3	50	3 (0)	25 (6)	15 (3.5)	14 (21)	80 (130)	16 (13)
6.4	50	3 (0)	26 (6)	16 (3.8)	19 (29)	310 (200)	17 (14)
6.5	50	.3 (0)	11 (2)	5.1 (1.0)	14 (22)	140 (130)	5.1 (1.3)
6.6	50	3 (0)	12 (3)	6.6 (1.5)	32 (31)	310 (200)	9.7 (8.2)
7.1	20	2.8 (.1)	25 (4)	13 (2)	2.1 (2.3)	90 (140)	6.2 (.4)
7.2	20	2.8 (.1)	30 (5)	17 (3)	15 (20)	170 (180)	7.0 (1.7)
7.3	20	2.8 (.1)	55 (10)	34 (6)	26 (33)	180 (210)	16 (12)
7.4	20	2.8 (.1)	62 (11)	38 (7)	80 (120)	1,600 (1,900)	16 (14)
7.5	20	2.8 (.1)	26 (5)	13 (2)	80 (160)	770 (880)	4.2 (1.3)
7.6	20	2.8 (.1)	29 (4)	15 (2)	24 (33)	680 (700)	5.4 (2.4)
10.1	20	4.9 (.1)	320 (20)	240 (10)	6.6 (7.9)	>1,100 (>1,900)	6.2 (.6)
10.2	20	4.9 (.1)	340 (30)	260 (10)	60 (160)	>1,600 (>2,300)	6.8 (1.1)
10.3	20	4.8 (.1)	460 (60)	350 (40)	430 (780)	>4,500 (>3,000)	13 (9)
10.4	20	4.8 (.1)	500 (50)	370 (30)	500 (1,500)	>3,900 (>3,200)	13 (10)
10.5	20	4.9 (.1)	310 (20)	240 (10)	200 (600)	>3,800 (>2,700)	6.2 (.6)
10.6	20	4.9 (.1)	320 (20)	240 (10)	13 (25)	>2,600 (>2,800)	6.4 (.5)
12.1	10	8.2 (.1)	930 (50)	760 (20)	16 (10)	>3,000 (>3,200)	7.6 (.6)
12.2	10	8.2 (.1)	970 (40)	790 (20)	230 (380)	>6,200 (>2,500)	8.0 (.5)
12.3	10	8.2 (.1)	1,200 (70)	970 (50)	1,400 (2,600)	>5,500 (>4,200)	11 (2)
12.4	10	8.2 (.1)	1,400 (140)	1,100 (100)	>2,300 (>3,700)	>4,400 (>4,600)	21 (24)
12.5	10	8.2 (.1)	960 (50)	770 (20)	880 (1,200)	>3,900 (>3,500)	7.8 (1.1)
12.6	10	8.3 (.1)	970 (50)	770 (30)	24 (20)	>4,500 (>3,000)	9.1 (2.0)

SD varies from 10^{-2} to 10^{-4}, depending on the degree of correlation. See Table 7 for two examples. This, combined with the accuracy figures reported earlier, suggests that the maximum number of permutations needed is in the range of 10^2 to 10^4. For the summaries in the tables, all permutations were used in cases where $m!/n!$ is less than 2,000 (where $n = 2, 4, 3$ for P_1, P_2, P_3), and otherwise 2,000 random permutations were used. Another factor that can affect the method P_2 is the integration method for the four-variate marginal probabilities; we use the double-precision version of Schervish's (1984) code, because for $m = 4$, numerical integration is faster with this than with the code of Berntsen et al. (1991).

For method P_1, we used an accuracy of 10^{-5} for the bivariate marginal probabilities. For method P_2, timings were obtained for an accuracy of 10^{-5} for bivariate marginal probabilities and either 10^{-5} or 10^{-4} for the trivariate and four-variate marginal probabilities. The accuracy of the multivariate probabilities was essentially the same; thus there is only one column for P_2 in Table 3 but two in Table 4. (Only the times for 10^{-4} are given in Table 6.) For method P_3, we used an accuracy of 10^{-4} for evaluation of the trivariate probabilities and then approximated four-variate and higher-order conditional probabilities. For comparison, timings are also given for the methods of evaluation of Genz (1992). For evaluations using the code of Berntsen et al. (1991), accuracies of 10^{-3} and 10^{-4} were used in the timings. For evaluations using Monte Carlo simulation applied to Genz's transformed integrand, an accuracy of 10^{-3}, together with a minimum of 10^4 simulations, was used. For Monte Carlo evaluation to an accuracy of 10^{-4}, these latter times can be multiplied by 100. (As an aside, another comparison is the Monte Carlo evaluation with Genz's transformed integrand is about 10–20 times faster compared to the simpler, Monte Carlo evaluation of the untransformed integrand.) The tables summarize average times as well as SD's of the times, because the computational times are highly variable for numerical quadrature (with larger probabilities tending to require longer computational time).

The times, in seconds, given in Tables 4 and 6 are from Sun SPARC 10 workstations. Some observations from Tables 4 and 6 are as follows. P_1 is very fast, and it is not until $m = 12$ that Genz's method with Monte Carlo evaluation and an accuracy of 10^{-3} is sometimes faster. (See also the discussion of computational complexity given later.) Overall, P_2 is comparable to Genz's method with numerical quadrature to an accuracy of 10^{-3} when considering both average time and the SD of time. Numerical quadrature takes more time for correlation matrices that have larger correlations or those that are nearly singular. But the accuracy of P_2 is often better than 10^{-3}, and P_2 is much faster on average than Genz's method with numerical quadrature to an accuracy of 10^{-4} (because the extra digit of accuracy for quadrature typically increases the computational time by a factor of 3–10 or more). P_3 is faster than P_2 for dimensions $m = 4$ and 5 but becomes slower for dimensions $m \geq 6$.

Genz's method with numerical quadrature occasionally runs into memory requirements for $m = 7$, and the frequency

Table 5. Average and Maximum Absolute Errors for Orthant Probabilities

Case	# sim.	P_1: avg.	Errors max.	P_2: avg.	Errors max.	P_3: avg.	Errors max.
5.1	50	.0005	.0010	0	.0001	.0001	.0002
5.2	50	.0007	.0014	0	.0001	.0008	.0016
5.3	50	.001	.002	.0001	.0002	.004	.008
5.4	50	.001	.004	.0001	.0003	.003	.007
5.5	50	.0004	.0010	.0000	.0001	.0001	.0001
5.6	50	.001	.004	.0001	.0006	.002	.003
6.1	50	.0008	.0016	0	.0001	.0001	.0002
6.2	50	.001	.002	0	.0001	.001	.002
6.3	50	.001	.003	.0001	.0004	.005	.010
6.4	50	.002	.006	.0002	.0005	.004	.010
6.5	50	.001	.003	.0001	.0001	.0001	.0005
6.6	50	.001	.004	.0002	.0007	.001	.004
7.1	20	.001	.002	0	.0001	.0001	.0002
7.2	20	.001	.002	0	.0001	.002	.003
7.3	20	.001	.005	.0002	.0004	.006	.012
7.4	20	.002	.007	.0004	.0010	.004	.010
7.5	20	.001	.003	.0001	.0002	.0002	.0005
7.6	20	.0012	.004	.0003	.0009	.001	.003

of failure increases as m increases and as more accuracy is requested. (For the comparisons, the amount of memory requested was that possible without the program crashing.) Any Monte Carlo evaluation of multivariate normal probabilities does not perform well with a quasi-Newton routine when estimation of multivariate normal parameters is needed. Hence approximations P_1 and P_2 compare favorably for applications (see Sec. 4 for more discussion).

We end this section with comparisons of computational complexity rates. Numerical integration increases exponentially in time (and memory requirements) as the dimension m increases, whereas Monte Carlo simulation to achieve a fixed precision has a time of $O(m)$. By bounding the number of (random) permutations used, the approximation P_1 is $O(m^2)$, because the time-consuming part is the evaluation of $m(m-1)/2$ bivariate marginal probabilities. Similarly, the approximation P_2 is $O(m^4)$ coming from the evaluation of $\binom{m}{4}$ four-variate marginal probabilities. The approximation P_3 is exponentially increasing in time to compute as m increases, because recursively the number of trivariate normal orthant probabilities e_m satisfies the recursion relation $e_m = e_{m-1} + (m-1)e_{m-2} + m(m-1)e_{m-3}/2$ for $m \geq 5$, with $e_2 = 0$, $e_3 = e_4 = 1$. Tables 4 and 6 give a good idea as to the meaning of these rates.

4. DISCUSSION

Based on the comparisons in Section 3, the second-order approximation P_2 is recommended for up to dimensions of $m = 12$ and larger. It often has close to four decimal places of accuracy, is faster than numerical quadrature with four decimal places of accuracy, and is faster than Monte Carlo simulation with four decimal places of accuracy for dimensions of less than 12. Approximation P_1 may be suitable, especially if correlations are not large, if two to three decimal places of precision is enough; its main advantage is its speed, even for dimensions exceeding $m = 20$. For dimensions less

Table 6. Timings in Seconds (From SPARC 10) for Orthant Probabilities

Case	# sim.	P_1 (10^{-5})	P_2 (10^{-4})	P_3 (10^{-5})	INT (10^{-3})	INT (10^{-4})	MC (10^{-3})
5.1	50	.05 (0)	1.4 (.2)	.8 (.1)	.2 (.2)	4.7 (4.6)	2.3 (.2)
5.2	50	.05 (0)	1.9 (.3)	.7 (.1)	.3 (.3)	5.5 (8.1)	2.7 (.3)
5.3	50	.05 (0)	4.1 (.4)	.8 (.1)	.8 (.7)	2.9 (3.3)	4.4 (1.6)
5.4	50	.05 (0)	4.0 (.4)	.9 (.1)	.7 (.9)	4.7 (5.3)	4.9 (2.1)
5.5	50	.05 (0)	1.3 (.2)	1.0 (.1)	16 (6)	120 (30)	2.4 (.3)
5.6	50	.05 (0)	2.4 (.3)	1.1 (.1)	3.8 (4.1)	39 (27)	4.3 (2.2)
6.1	50	.3 (0)	3.7 (.5)	18 (1)	.6 (.6)	24 (25)	2.8 (.2)
6.2	50	.3 (0)	5.0 (.6)	14 (1)	.8 (.7)	18 (30)	3.5 (.5)
6.3	50	.3 (0)	11 (1)	17 (1)	4.2 (7.1)	26 (47)	6.0 (2.2)
6.4	50	.3 (0)	11 (1)	18 (1)	2.5 (3.1)	23 (33)	7.2 (3.1)
6.5	50	.3 (0)	3.5 (.6)	26 (4)	190 (60)	280 (10)	3.6 (1.0)
6.6	50	.3 (0)	5.2 (.6)	21 (1)	30 (50)	370 (370)	5.7 (3.7)
7.1	20	2.9 (.1)	9 (1)	390 (30)	1.3 (1.1)	80 (110)	3.4 (.2)
7.2	20	3.0 (.1)	12 (1)	300 (20)	2.5 (2.1)	70 (120)	4.7 (1.0)
7.3	20	2.8 (.1)	23 (3)	350 (20)	11 (13)	110 (140)	8.0 (4.0)
7.4	20	2.8 (.1)	25 (2)	370 (20)	13 (15)	300 (390)	10 (5)
7.5	20	3.0 (.1)	9 (1)	620 (70)	>2,300 (>800)	>3,200 (>100)	7.6 (2.4)
7.6	20	2.9 (.1)	12 (1)	440 (30)	80 (190)	>2,600 (>1,400)	6.3 (3.2)

Table 7. Standard Deviations over Permutations for Some Cases

Case	w	x	Exact	P_1 (SD)	P_2 (SD)	P_3 (SD)
5.2	−.1, −.2, −.3, −.8, −.9	∞	.2436	.2444 (.0015)	.2436 (.0002)	.2438 (.0004)
5.4	−.1, −.2, −.3, −.8, −.9	∞	.3923	.3954 (.0071)	.3921 (.0006)	.3900 (.0021)

than 12, it is faster than Monte Carlo simulation with three decimal places of accuracy. Approximation P_3 compares favorably, for orthant probabilities, in accuracy and time only for $m = 5$ and maybe $m = 6$, and only then if the correlations are not too large. When considering accuracy, computational time, and range of applicability of the three approximations (i.e., range of allowable correlation matrices and w_i's and x_i's in (1)), these new approximations, especially P_2, offer substantial improvements on previous approximations, because some previous approximations are for orthant probabilities only and others are for the case of special correlation matrices and so on.

All of the approximations get worse as the correlations increase. Research could be conducted on improvements for the case of larger correlations. Also, using a good method to choose the order of indices (or the permutation) in (5) or (8) will speed up the computational time, especially for P_3. For the moment-generating function approach, more terms in the Edgeworth expansion or Gram–Charlier expansion for the truncated conditional multivariate normal distribution may be possible, albeit extremely tedious in analytic derivations and coding. (See Henery 1981 for a special case.)

Approximations P_1 and P_2 should be especially useful for estimation of parameters in models involving the multivariate normal distributions (see Sec. 1). Lerman and Manski (1982) concluded that the Clark (1961) approximation might be satisfactory for estimation of parameters in a probit model, in which case the approximation P_1 would be more than adequate. Because of its speed, the approximation P_1 could be used to get preliminary estimates of parameters, with an improved method of computing the likelihood used at a second stage if necessary. How approximation methods and numerical quadrature work with quasi-Newton routines will be studied further in future research, as will possible modification of quasi-Newton routines to handle an objective function evaluated via Monte Carlo simulation.

The codes for the approximations P_1 and P_2, both in C and Fortran, are available from the author. The coding of P_1 is easier; it mainly relies on the routine of Donnelly (1973).

The methodology of Section 2 should be more generally usable. The approximation based on binary variables should be applicable to other multivariate distributions that do not have closed-form cumulative distribution functions and that have bivariate, trivariate, and four-variate (marginal) distributions that can be computed easily. Also, the method can be applied to probabilities involving order statistics for order statistics ranking models.

[Received July 1993. Revised June 1994.]

REFERENCES

Anderson, J. A., and Pemberton, J. D. (1985), "The Grouped Continuous Model for Multivariate Orders Categorical Variables and Covariate Adjustment," *Biometrics*, 41, 875–885.

Berntsen, J., Espelid, T. O., and Genz, A. (1991), "Algorithm 698: SCUHRE—An Adaptive Multidimensional Integration Routine for a Vector of Integrals," *Association for Computing Machinery Transactions on Mathematical Software*, 17, 452–456.

Böckenholt, U. (1992a), "Thurstonian Representation for Partial Ranking Data," *British Journal of Mathematical and Statistical Psychology*, 45, 31–49.

——— (1992b), "Applications of Thurstonian Models to Ranking Data," in *Probability Models and Statistical Analyses for Ranking Data*, Lecture Notes in Statistics 80, eds. M. A. Fligner and J. S. Verducci, New York: Springer-Verlag, pp. 157–172.

Clark, C. (1961), "The Greatest of a Finite Set of Random Variables," *Operations Research*, 9, 145–162.

Donnelly, T. G. (1973), "Algorithm 462: Bivariate Normal Distribution," *Communications of the Association for Computing Machinery*, 16, 638.

Drezner, Z. (1990), "Approximations to the Multivariate Normal Integral," *Communications in Statistics—Simulation and Computation*, 19, 527–534.

Finney, D. J. (1962), "Cumulants of Truncated Multinormal Distributions," *Journal of the Royal Statistical Society*, Ser. B, 24, 535–536.

Genz, A. (1992), "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics*, 1, 141–149.

Henery, R. J. (1981), "An Approximation to Certain Multivariate Normal Probabilities," *Journal of the Royal Statistical Society*, Ser. B, 43, 81–85.

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

Lerman, S. R., and Manski, C. F. (1981), "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data With Econometric Applications*, eds. C. F. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 305–319.

McGill, J. I. (1992), "The Multivariate Hazard Gradient and Moments of the Truncated Multinormal Distribution," *Communications in Statistics—Theory and Methods*, 21, 3053–3060.

Mendell, N. R., and Elston, R. C. (1974), "Multifactorial Qualitative Traits: Genetic Analysis and Prediction of Recurrence Risks," *Biometrics*, 30, 41–57.

Ochi, Y., and Prentice, R. L. (1984), "Likelihood Inference in a Correlated Probit Regression Model," *Biometrika*, 71, 531–543.

Rice, J., Reich, T., Cloninger, C. R., and Wette, R. (1979), "An Approximation to the Multivariate Normal Integral: Its Application to Multifactorial Qualitative Traits," *Biometrics*, 35, 451–459.

Schervish, M. J. (1984), "Multivariate Normal Probabilities With Error Bound," *Applied Statistics*, 33, 81–94.

Solow, A. R. (1990), "A Method for Approximating Multivariate Normal Orthant Probabilities," *Journal of Statistical Computation and Simulation*, 37, 225–229.

Tallis, G. M. (1961), "The Moment-Generating Function of the Truncated Multinormal Distribution," *Journal of the Royal Statistical Society*, Ser. B, 23, 223–229.

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