2

Special Cases

The problem of evaluating (1.1) or (1.3) can often be simplified by specializing either k, $\mathbf{R}$ or $\mathbf{a}$ and $\mathbf{b}$. In Section 2.1 we focus on the work that has been done on bivariate and trivariate probabilities and not on general MVN and MVT probabilities. In Section 2.2 we consider calculating probabilities over special integration regions, such orthants, ellipsoids, and hyperboloids. Finally, in Section 2.3 we discuss MVN and MVT problems involving special correlation structures. We do not consider the univariate cases, which have been carefully analyzed elsewhere; see Johnson and Kotz (1970a,b) for extensive discussions and references. Highly accurate implementations for $\Phi(x)$, $\Phi^{-1}(x)$, $T(x;\nu)$, and $T^{-1}(x;\nu)$ are available in standard statistical computing environments. We assume the availability of these functions for many of the computational methods that we discuss in this and later chapters.

2.1 Bivariate and Trivariate Probabilities

2.1.1 Bivariate Probabilities

The method developed by Owen (1956) was for a long time the most widely used approach to calculate bivariate normal (BVN) probabilities. Owen showed that

$$\Phi_2(-\infty, \mathbf{b}; \rho) = \frac{\Phi(b_1) + \Phi(b_2)}{2} - E(b_1, \hat{b}_1) - E(b_2, \hat{b}_2) - c,$$

where ρ is the correlation coefficient,

$$c = \begin{cases} 0, & \text{if } b_1 b_2 > 0 \text{ or } b_1 b_2 = 0,\ b_1 + b_2 \geq 0 \\ \frac{1}{2}, & \text{otherwise} \end{cases},$$

$$\hat{b}_1 = \frac{b_2 - b_1\rho}{b_1\sqrt{1-\rho^2}}, \quad \hat{b}_2 = \frac{b_1 - b_2\rho}{b_2\sqrt{1-\rho^2}},$$

A. Genz, *Computation of Multivariate Normal and t Probabilities*
DOI: 10.1007/978-3-642-01689-9_2,

and where the function E (called $T-$*function* by Owen) is defined as

$$E(h,a)=\frac{1}{2\pi}\int_{0}^{a}\frac{e^{-h^2(1+x^2)}}{1+x^2}dx. \tag{2.1}$$

For numerical computations, Owen (1956) evaluated the integral in (2.1) by expanding the exponential into a power series and integrating it term by term. The resultant series expression

$$E(h,a)=\frac{1}{2\pi}\left\{\tan^{-1}(a)-\sum_{j=0}^{\infty}c_j a^{2j+1}\right\}$$

with

$$c_j=\frac{(-1)^j}{2j+1}\left\{1-e^{-h^2/2}\sum_{i=0}^{j}\frac{h^{2i}}{2^i i!}\right\}$$

converges for all h and a, although the convergence may be slow if neither h nor a is small (Wijsman, 1996). Therefore, much effort has been devoted to more efficient computations of $E(h,a)$. Recent developments include the approaches of Wijsman (1996) and Patefield and Tandy (2000). They proposed hybrid methods based on different approaches of evaluating the integral in (2.1). The $(h,a)-$plane is divided into disjoint subsets such that a minimum computing effort is required by selecting an appropriate method for each subset.

Divgi (1979) developed an approximation that avoids the computation of $E(h,a)$. Let

$$R^2=\frac{a_1^2-2a_1a_2\rho+a_2^2}{1-\rho^2},$$

with

$$\pi/2-\theta=\sin^{-1}\left(\frac{a_1}{R}\right) \quad \text{and} \quad \theta-\gamma=\sin^{-1}\left(\frac{a_2}{R}\right).$$

Then,

$$\Phi_2(\mathbf{a},\infty;\rho)=W(R,\pi/2-\theta)+W(R,\theta-\gamma)+c', \tag{2.2}$$

where c' is a constant dependent on $\mathbf{a}$. The function $W(R,\psi)$ was first introduced by Ruben (1961) as the probability content of the sector

$$\{X\geq R\}\cap\{(X-R)\tan(\psi)\geq Y\geq 0\}$$

for two standard normal variates X and Y. Divgi (1979) proposed to approximate the function W with the help of a polynomial expansion of Mill's ratio $(1-\Phi(x))/\phi(x)$. Terza and Welland (1991) compared equation (2.2) with several competing methods, including Owen's original series expansion given above. The study came to the conclusion that the method of Divgi (1979) outperformed the other methods, achieving "*... 14 digits accuracy 10 times faster then its nearest competitor*".

At approximately the same time, Drezner and Wesolowsky (1990) presented a simple method based on a reduction formula by Sheppard (1900),

$$\Phi_2(\mathbf{a}, \infty; \rho) = \frac{1}{2\pi} \int_{\cos^{-1}(\rho)}^{\pi} e^{-\frac{a_1^2 - 2a_1a_2\cos(x) + a_2^2}{2\sin^2(x)}} dx.$$

Differentiating with respect to ρ and integrating from 0 to ρ yields

$$\Phi_2(\mathbf{a}, \infty; \rho) = \Phi(-a_1)\Phi(-a_2) + \frac{1}{2\pi} \int_0^{\rho} \frac{1}{\sqrt{1-x^2}} e^{-\frac{a_1^2 - 2a_1a_2x + a_2^2}{2(1-x^2)}} dx, \qquad (2.3)$$

which Wang and Kennedy (1990) considered "*a competitor*" to the Divgi algorithm. Drezner and Wesolowsky showed that the use of low-order numerical integration methods applied to (2.3) could produce very accurate $\Phi_2(\mathbf{a}, \infty; \rho)$ values. It is worth noting that equation (2.3) already appears in the derivation for $E(h, a)$ in Owen (1956, p. 1078); it can also be determined from the general MVN identity (Plackett, 1954)

$$\frac{\partial \phi_k(\mathbf{x}; \mathbf{R})}{\partial \rho_{ij}} = \frac{\partial^2 \phi_k(\mathbf{x}; \mathbf{R})}{\partial x_i \partial x_j}. \qquad (2.4)$$

Genz (2004) modified (2.3) by additionally substituting $x = \sin(\theta)$ so that (2.3) becomes

$$\Phi_2(\mathbf{a}, \infty; \rho) = \Phi(-a_1)\Phi(-a_2) + \frac{1}{2\pi} \int_0^{\sin^{-1}(\rho)} e^{-\frac{a_1^2 - 2a_1a_2\sin(\theta) + a_2^2}{2\cos^2(\theta)}} d\theta.$$

The resulting modified method is believed to be slightly more accurate.

Other approaches for computing BVN probabilities were presented by Moskowitz and Tsai (1989), Cox and Wermuth (1991) and Maghsoodloo and Huang (1995). Albers and Kallenberg (1994) and Drezner and Wesolowsky (1990) discussed simple approximations to BVN probabilities for large values of ρ.

One of the few direct approaches for computing bivariate t (BVT) probabilities was introduced by Dunnett and Sobel (1954). They succeeded in expressing $T_2(-\infty, \mathbf{b}; \rho, \nu)$ as a weighted sum of incomplete beta functions. In addition, the authors provide asymptotic expressions (in ν) for T_2 and for the inverse problem of finding equi-coordinate quantiles $\mathbf{b} = (b, b)^t$, such that $T_2(-\infty, \mathbf{b}; \rho, \nu) = p$ for a given $0 < p < 1$.

Genz (2004) considered the use of a bivariate generalization of Plackett's formula in the form

$$\frac{\partial T_2(-\infty, \mathbf{b}; \rho, \nu)}{\partial \rho} = \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{b_1^2 + b_2^2 - 2\rho b_1 b_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu}{2}}$$

as the basis for a BVT algorithm. Integration of this equation provides the formula

$$T_2(-\infty, \mathbf{b}; \rho, \nu) = T_2(-\infty, \mathbf{b}; u, \nu) + \frac{1}{2\pi} \int_u^\rho \frac{1}{\sqrt{1-r^2}} \left(1 + \frac{b_1^2 + b_2^2 - 2rb_1b_2}{\nu(1-r^2)}\right)^{-\frac{\nu}{2}} dr,$$

where $u = \text{sign}(\rho)$ and

$$T_2(-\infty, \mathbf{b}; u, \nu) = \begin{cases} T(\min(b_1, b_2); \nu), & \text{if } u = 1, \\ \max(0, T(b_1; \nu) - T(-b_2; \nu)), & \text{if } u = -1. \end{cases}$$

Genz studied the use of this formula with various numerical integration methods, but concluded that an implementation of the Dunnett and Sobel (1954) algorithm was the most efficient.

2.1.2 Trivariate Probabilities

The trivariate integration problem has been addressed less often in the literature. For the triviariate normal (TVN) case, Gupta (1963a) conditioned on the third integration variable and thereby obtained

$$\Phi_3(-\infty, \mathbf{b}; \mathbf{R}\,) = \int_{-\infty}^{b_1} \Phi_2\left(\frac{b_2 - \rho_{21}y}{\sqrt{1-\rho_{21}^2}}, \frac{b_3 - \rho_{31}y}{\sqrt{1-\rho_{31}^2}}; \frac{\rho_{32} - \rho_{21}\rho_{31}}{\sqrt{(1-\rho_{21}^2)(1-\rho_{31}^2)}}\right) \phi(y) dy. \tag{2.5}$$

A different approach is based on Plackett's identity (2.4). Plackett (1954) integrated this identity to show that

$$\Phi_3(-\infty, \mathbf{b}; \mathbf{R}) = \Phi_3(-\infty, \mathbf{b}; \mathbf{R}') + \frac{1}{2\pi} \sum_{i<j} \int_0^1 \frac{\rho_{ij} - \rho'_{ij}}{\sqrt{1 - r_{ij}^2(y)}} e^{-\frac{x_i^2 - 2r_{ij}(y)x_ix_j + x_j^2}{2(1-r_{ij}^2(y))}} \Phi\left(x'_l(y)\right) dy, \tag{2.6}$$

where $r_{ij}(y) = (1-y)\rho'_{ij} + y\rho_{ij}$, l is the third coordinate and

$$x'_l(y) = \frac{(1 - r_{ij}^2(y))x_l - (r_{il}(y) - r_{ij}(y)r_{jl}(y))x_i - (r_{jl}(y) - r_{ij}(y)r_{il}(y))x_j}{\sqrt{1 - r_{ij}^2(y)|\mathbf{R}|}}.$$

The reference matrix $\mathbf{R}' = (\rho'_{ij})$ is chosen so that the associated probability Φ_3 is easily computed. Equation (2.6) thus reduces the computational effort to three univariate integrals. There are several choices for $\mathbf{R}'$. Plackett proved that ρ'_{32} can always be chosen so that the second term in (2.6) consists of one single integral. Other possibilities are $\mathbf{R}' = \mathbf{I}_3$, where $\mathbf{I}_k$ is the $k \times k$ unit matrix, resulting in $\Phi_3(-\infty, \mathbf{b}; \mathbf{R}') = \prod_{i=1}^{3} \Phi(b_i)$, or the use of the product correlation structure (2.16). Drezner (1994) proposed using $\rho'_{21} = \rho'_{31} = 0$ and $\rho'_{32} = \rho_{32}$, in which case $\partial\Phi_3/\partial r_{32}(t) = 0$ and the sum in (2.6) consists of only two integrals instead of three. Detailed discussions on the numerical stability of the various methods are given by Gassmann (2002). Genz (2004) reported a comparison study of the Plackett identity methods and methods based on the numerical evaluation of equation (2.5). The results of these studies indicate that a Plackett identity method which uses Drezner's choice for $\mathbf{R}'$ with numerical integration can provide the most efficient general method for computing TVN probabilities.

Genz (2004) also considered algorithms for efficient and accurate computation of trivariate t (TVT) probabilities. A generalization of Plackett's TVN identity was derived for the TVT case in the form

$$\frac{\partial T_3(-\infty, \mathbf{b}; \mathbf{R}, \nu)}{\partial \rho_{21}} = \frac{(1 + \frac{f_3(\rho_{21})}{\nu})^{-\frac{\nu}{2}}}{2\pi\sqrt{1-\rho_{21}^2}} \cdot T\left(\frac{u_3(\rho_{21})}{(1 + \frac{f_3(\rho_{21})}{\nu})^{\frac{1}{2}}}; \nu\right), \tag{2.7}$$

where

$$f_3(r) = \frac{b_1^2 + b_2^2 - 2rb_1b_2}{(1-r^2)}$$

and

$$u_3(r) = \frac{b_3(1-r^2) - b_1(\rho_{31} - r\rho_{32}) - b_2(\rho_{32} - r\rho_{31})}{((1-r^2)(1-r^2-\rho_{31}^2-\rho_{32}^2+2r\rho_{31}\rho_{32}))^{\frac{1}{2}}}.$$

Integration of this formula can provide formulas for TVT probabilities that combine a reference matrix probability and univariate integrals, but the choice of a reference matrix is more difficult, compared to the TVN case. The preferred $\mathbf{R}'$ for TVN computations (Drezner, 1994) does not have an easily computed TVT value. Genz (2004) recommends a hybrid method that uses an initial reference $\mathbf{R}''$ with $\rho''_{21} = \rho''_{31} = 0$, and $\rho''_{32} = \text{sign}(\rho_{32})$. The singular $T_3(-\infty, \mathbf{b}; \mathbf{R}'', \nu)$ value can be computed using univariate t and BVT values. Numerical integration of equation (2.7) from $\mathbf{R}''$ to $\mathbf{R}'$, followed by integration from $\mathbf{R}'$ to $\mathbf{R}$ provides an efficient and accurate numerical method for TVT probability computations. Some software for the accurate computation of TVN and TVT probabilities will be discussed in Section 5.5.

2.2 Special Integration Regions

2.2.1 Orthants

If the integral (1.1) is defined over the positive orthant $[0, \infty]^k$, the associated MVN integral is called a (centered) *orthant probability* P_k. The evaluation of orthant probabilities is a classical problem whose history and applications are briefly summarized by Owen (1985). Note that $T_k(-\infty, \mathbf{0}; \mathbf{R}, \nu) = \Phi_k(-\infty, \mathbf{0}; \mathbf{R})$ for all ν, as seen from (1.3).

For integrals P_k with general correlation matrices, explicit formulas are only available for small values of k:

$$P_1 = \frac{1}{2},$$

$$P_2 = \frac{1}{4} + \frac{\sin^{-1}(\rho_{12})}{2\pi}$$

and

$$P_3 = \frac{1}{8} + \frac{1}{4\pi}\left\{\sin^{-1}(\rho_{12}) + \sin^{-1}(\rho_{23}) + \sin^{-1}(\rho_{13})\right\}.$$

For general k, the following approach halves the dimensionality of the integration problem. If $k = 2n$, Childs (1967) showed that

$$2^{2n}P_{2n} = 1 + \frac{2}{\pi}\sum_{i<j}^{2n}\sin^{-1}(\rho_{ij}) + \sum_{j=2}^{n}\left(\frac{2}{\pi}\right)^j + \sum_{i_1<\ldots<i_{2j}}^{2n} I_{2j}\left(\mathbf{R}^{i_1,\ldots,i_{2j}}\right), \quad (2.8)$$

where $\mathbf{R}^{i_1,\ldots,i_{2j}}$ denotes the submatrix consisting of the $i_1^{th}, \ldots, i_{2j}^{th}$ rows and columns of $\mathbf{R}$ and

$$I_{2j}(\mathbf{\Lambda}_{2j}) = (-2\pi)^j \int_{\mathbb{R}^{2j}} \exp(-\mathbf{z}^t\mathbf{\Lambda}_{2j}\mathbf{z}) \prod_{i=1}^{2j} z_i^{-1} d\mathbf{z},$$

where $\mathbf{\Lambda}_{2j}$ is a covariance matrix with $2j$ covariates. Childs (1967) also developed a similar formula for $k = 2n+1$, but a result of David (1953) ensures that the computation of any orthant probability of odd order $2n+1$ can be reduced to a sum of integrals of order at most $2n$. Sun (1988a) extended formula (2.8) and obtained the following recursive relationship among the I_{2j}'s,

$$I_{2j}(\mathbf{\Lambda}_{2j}) = \int_0^1 \sum_{i=2}^{2j} \frac{\lambda_{1i}}{\sqrt{\lambda_{11}\lambda_{ii} - \lambda_{1i}^2 x^2}} I_{2j-2}(\mathbf{\Lambda}_{2j-2}^i)dx. \quad (2.9)$$

Therefore, by using (2.8) and the recursive application of (2.9), the computation of orthant probabilities can be reduced to the computation of several multidimensional integrals of order at most $n-1$. In addition, the unbounded integration region over the positive orthant is transformed to an integration over the unit hypercube $[\mathbf{0}, \mathbf{1}]$ and the methods of Section 4.2 can be applied. Sun (1988a,b) established explicit formulas up to $k = 9$. These formulas were extended to $k = 11$ by Sun and Asano (1989) when $\mathbf{R}$ is tridiagonal.

A few other methods shall be reviewed briefly. Evans and Swartz (1988) developed a class of Monte Carlo estimators for the given integration problem. The estimators take the form of a constant multiplied by $||\mathbf{W}\mathbf{z}||^{-k}$, where $\mathbf{z}$ is distributed on a $(k-1)-$dimensional manifold and $\mathbf{W}$ is the decomposition $\mathbf{W} = \mathbf{D}\,\mathrm{diag}(||\mathbf{d}_1||^{-1}, \ldots, ||\mathbf{d}_\mathrm{k}||^{-1})$ with $\mathbf{D} = \mathbf{R}^{-1/2} = (\mathbf{d}_1, \ldots, \mathbf{d}_k)$ and where

$||.||$ denotes the Euclidean norm. The different estimators arise based on different choices of the manifold as the authors try to stabilize the estimator as much as possible. In particular the authors show that earlier results of Moran (1984) arise naturally within the context of their estimators. Both importance sampling and control variate methods are discussed. Another method was developed by Gibbons et al (1987, 1990). They modified the approximation by Clark (1961) to the moments of the maximum of k jointly normal variables, and used the formula $P_k = P(\min\{X_1, \ldots, X_k\} \geq 0)$. Finally, Ni and Kedem (1999) used the Cholesky decomposition of $\mathbf{R}$, followed by a polar coordinate transformation. These transformations are discussed for more general MVN and MVT problems in Section 4.1. Some specific orthant probability problems can be expressed in terms of simplified numerical expressions developed by Ni and Kedem (1999, 2000).

2.2.2 Ellipsoids

General MVN probabilities for *elliptical regions* are defined by

$$\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}) = \frac{1}{\sqrt{|\mathbf{\Sigma}|(2\pi)^k}} \int\limits_{\{(\mathbf{x}-\mathbf{c})^t\mathbf{A}(\mathbf{x}-\mathbf{c})\leq t\}} e^{-\frac{1}{2}\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}} d\mathbf{x},$$

for a positive semidefinite $k \times k$ matrix $\mathbf{A}$, and $t > 0$, so that the integration region is an ellipsoid centered at $\mathbf{c}$. Several statistical applications require $\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma})$, and some of these are surveyed by Ruben (1960). This type of problem can be put into a simpler standard form if we let $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$, where $\mathbf{L}$ is the lower triangular Cholesky factor for $\mathbf{\Sigma}$. If we determine a spectral decomposition for $\mathbf{L}^t\mathbf{A}\mathbf{L} = \mathbf{Q}\mathbf{D}\mathbf{Q}^t$, with $\mathbf{Q}$ an orthogonal matrix and $\mathbf{D}$ a diagonal matrix, then the result of the transformation $\mathbf{x} = \mathbf{L}\mathbf{Q}\mathbf{z}$ is

$$\begin{aligned}\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}) &= \Phi_k(\mathbf{D}, \boldsymbol{\delta}, t; \mathbf{I}_k) \\ &= \frac{1}{\sqrt{(2\pi)^k}} \int\limits_{\{(\mathbf{z}-\boldsymbol{\delta})^t\mathbf{D}(\mathbf{z}-\boldsymbol{\delta})\leq t\}} e^{-\frac{1}{2}\mathbf{z}^t\mathbf{z}} d\mathbf{z}, \end{aligned} \tag{2.10}$$

where $\boldsymbol{\delta} = \mathbf{Q}^t\mathbf{L}^{-1}\mathbf{c}$, because

$$\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x} = \mathbf{z}^t\mathbf{Q}^t\mathbf{L}^t(\mathbf{L}\mathbf{L}^t)^{-1}\mathbf{L}\mathbf{Q}\mathbf{z} = \mathbf{z}^t\mathbf{z}$$

and

$$\begin{aligned}(\mathbf{x}-\mathbf{c})^t\mathbf{A}(\mathbf{x}-\mathbf{c}) &= (\mathbf{L}\mathbf{Q}\mathbf{z}-\mathbf{c})^t\mathbf{A}(\mathbf{L}\mathbf{Q}\mathbf{z}-\mathbf{c}) \\ &= (\mathbf{z}-\mathbf{Q}^t\mathbf{L}^{-1}\mathbf{c})^t\mathbf{D}(\mathbf{z}-\mathbf{Q}^t\mathbf{L}^{-1}\mathbf{c}).\end{aligned}$$

Ruben (1962) derived a series solution for the problem (2.10) in the form

$$\Phi_k(\mathbf{D}, \boldsymbol{\delta}, t; \mathbf{I}_k) = \sum_{j=0}^{\infty} c_j F(k' + 2j, t/\beta).$$

In this formula, $F(l, y)$ is a central χ^2 distribution function with l degrees of freedom, k' is the rank of $\mathbf{D}$ and β is a parameter. If we denote the nonzero diagonal entries in $\mathbf{D}$ by $d_1, d_2, \ldots, d_{k'}$, it follows from Ruben that $0 < \beta < 2\min_i d_i$ is a sufficient condition for uniform convergence of the series. The series coefficients are given by

$$c_0 = Ae^{-\lambda/2} \text{ and } c_j = j^{-1}\sum_{i=0}^{j-1} g_{j-i}c_i \text{ for } j > 0,$$

where

$$A = \prod_{i=1}^{k'} \sqrt{\beta/d_i}, \ \lambda = \sum_{i=0}^{k'} \delta_i^2, \text{ and } g_j = \sum_{i=1}^{k'} \gamma_i^{j-1}(j\delta_i^2(1-\gamma_i)+\gamma_i)/2,$$

with $\gamma_i = 1 - \beta/d_i$. An implementation of this method has been provided by Sheil and O'Muircheartaigh (1977), where the choice $\beta = 29\min_i d_i/32$ is used.

Simulation methods for $\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma})$ based on spherical-radial integration were provided by Lohr (1993) and Somerville (2001) and will be discussed later in Section 4.1.1. Ruben (1960, 1961, 1962) also discussed related problems of determining the probability contents over other geometrical regions (simplices, polyhedral cones, etc.) under spherical normal distributions.

General MVT probabilities for elliptical regions can be defined in a way that is similar to $\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma})$, by

$$T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})\sqrt{|\mathbf{\Sigma}|(\nu\pi)^k}} \int\limits_{\{(\mathbf{x}-\mathbf{c})^t\mathbf{A}(\mathbf{x}-\mathbf{c})\le t\}} \left(1 + \frac{\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+k}{2}} d\mathbf{x}. \tag{2.11}$$

An equivalent definition, in terms of $\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma})$, can be determined if the integral in (2.11) is multiplied by a χ integral term (with value 1), so that

$$T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu) = \frac{2^{1-\frac{k+\nu}{2}}}{\Gamma(\frac{k+\nu}{2})} \int_0^\infty r^{k+\nu-1} e^{-\frac{r^2}{2}} dr,$$

If we then change variables using $r = s\sqrt{1 + \mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}/\nu}$, change the order of integration, cancel the $\Gamma(\frac{\nu+k}{2})$ terms, and separate the exponential terms,

$$T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu) = \frac{2^{1-\frac{k+\nu}{2}}}{\Gamma(\frac{\nu}{2})\sqrt{|\mathbf{\Sigma}|(\nu\pi)^k}} \int_0^\infty s^{k+\nu-1} e^{-\frac{s^2}{2}} \int\limits_{\{(\mathbf{x}-\mathbf{c})^t\mathbf{A}(\mathbf{x}-\mathbf{c})\le t\}} e^{-\frac{s^2}{2\nu}\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}} d\mathbf{x} ds.$$

After a final transformation $\mathbf{x} = \sqrt{\nu}\mathbf{y}/s$, and some further cancelations in the constant terms

$$T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^\infty s^{k+\nu-1} e^{-\frac{s^2}{2}} \frac{1}{\sqrt{|\mathbf{\Sigma}|(2\pi)^k}} \int_{\{(\mathbf{y}-\frac{s}{\sqrt{\nu}}\mathbf{c})^t \mathbf{A}(\mathbf{y}-\frac{s}{\sqrt{\nu}}\mathbf{c}) \le \frac{s^2 t}{\nu}\}} e^{-\frac{1}{2}\mathbf{y}^t \mathbf{\Sigma}^{-1}\mathbf{y}} d\mathbf{y} ds,$$

which can be written in terms of Φ_k as

$$T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^\infty s^{\nu-1} e^{-\frac{s^2}{2}} \Phi_k \left(\mathbf{A}, \frac{s\mathbf{c}}{\sqrt{\nu}}, \frac{s^2 t}{\nu}; \mathbf{\Sigma} \right) ds. \tag{2.12}$$

Simulation methods for $T_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}, \nu)$ are discussed in Sections 4.1.1 and 4.1.2.

2.2.3 Hyperboloids

There are applications in financial mathematics (Albanese and Seco, 2001; Brummelhuis et al, 2002; Sadefo Kamdem, 2005; Sadefo Kamdem and Genz, 2008) where the integration region is determined by a set of the form $\{\mathbf{x} : (\mathbf{x} - \mathbf{c})^t \mathbf{A}(\mathbf{x} - \mathbf{c}) \le t\}$, with $\mathbf{A}$ a symmetric indefinite matrix. Following the notation in the previous section, after Cholesky decomposition of $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$, the spectral decomposition of $\mathbf{L}^t\mathbf{A}\mathbf{L} = \mathbf{Q}\mathbf{D}\mathbf{Q}^t$, and the transformation $\mathbf{x} = \mathbf{L}\mathbf{Q}\mathbf{z}$, we obtain the same (MVN case) equation (2.10)

$$\begin{aligned}\Phi_k(\mathbf{A}, \mathbf{c}, t; \mathbf{\Sigma}) &= \Phi_k(\mathbf{D}, \boldsymbol{\delta}, t; \mathbf{I}_k) \\ &= \frac{1}{\sqrt{(2\pi)^k}} \int_{\{(\mathbf{z}-\boldsymbol{\delta})^t \mathbf{D}(\mathbf{z}-\boldsymbol{\delta}) \le t\}} e^{-\frac{1}{2}\mathbf{z}^t\mathbf{z}} d\mathbf{z},\end{aligned}$$

with $\boldsymbol{\delta} = \mathbf{Q}^t\mathbf{L}^{-1}\mathbf{c}$, but now we assume that the diagonal matrix $\mathbf{D}$ has some negative entries. The variables can be now reordered so that $\mathbf{D} = \text{diag}\{d_1^+, d_2^+, \ldots, d_{k_+}^+, -d_1^-, -d_2^-, \ldots, -d_{k_-}^-\}$ with all $d_i^+ \ge 0$ and all $d_i^- > 0$, and the $\mathbf{z}$ and $\boldsymbol{\delta}$ vectors are partitioned into components associated with the non-negative and negative diagonal entries in $\mathbf{D}$ with $\mathbf{z} = (\mathbf{z}_+, \mathbf{z}_-)$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_+, \boldsymbol{\delta}_-)$. Then the *hyperboloid* integration region can be written in the form

$$R = \{\mathbf{z} : \ \mathbf{z}_+^t \mathbf{D}_+ \mathbf{z}_+ \le \hat{t} + \mathbf{z}_-^t \mathbf{D}_- \mathbf{z}_-\},$$

with $\mathbf{D}_\pm = \text{diag}(d_i^\pm)$ and $\hat{t} = t + \boldsymbol{\delta}_-^t \mathbf{D}_- \boldsymbol{\delta}_- - \boldsymbol{\delta}_+^t \mathbf{D}_+ \boldsymbol{\delta}_+$. Now, $\Phi_k(\mathbf{D}, \boldsymbol{\delta}, t; \mathbf{I}_k)$ can be written as

$$\Phi_k(\mathbf{D}, \boldsymbol{\delta}, t; \mathbf{I}_k) = \int_{\{\hat{t}+\mathbf{z}_-^t\mathbf{D}_-\mathbf{z}_- \ge 0\}} \frac{e^{-\frac{1}{2}\mathbf{z}_-^t\mathbf{z}_-}}{\sqrt{(2\pi)^{k_-}}} \int_{\{\mathbf{z}_+^t\mathbf{D}_+\mathbf{z}_+ \le \hat{t}+\mathbf{z}_-^t\mathbf{D}_-\mathbf{z}_-\}} \frac{e^{-\frac{1}{2}\mathbf{z}_+^t\mathbf{z}_+}}{\sqrt{(2\pi)^{k_+}}} d\mathbf{z},$$

with $d\mathbf{z} = (d\mathbf{z}_+, d\mathbf{z}_-)$ so that the $\mathbf{z}_-$ integral is the outer integral. The method from Ruben (1960) could be used for the numerical evaluation of the inner

$\mathbf{z}_+$ integral, combined with another method for the outer integral. Simulation methods for these integrals will be discussed in Chapter 5. A similar analysis can also be applied to MVT problems over hyperboloid regions, working with either equation (2.11) or (2.12).

2.3 Special Correlation Structures

There are several cases, where a special correlation matrix $\mathbf{R}$ leads to simplified computational problems. In some cases the dimensionality of the integration problem can be reduced, and in other cases a special structure for $\mathbf{R}$ allows a faster algorithm to be used. We consider two main classes of special correlation structures. In Section 2.3.1 we consider problems involving correlation matrices that can be written as the sum of a diagonal matrix and a reduced rank matrix. In Section 2.3.2 we review methods for correlation matrices that have a banded structure.

2.3.1 Diagonal and Reduced Rank Correlation Matrices

In this section we assume that $\mathbf{R}$ can be written as

$$\mathbf{R} = \mathbf{D} + \mathbf{V}\mathbf{V}^t, \tag{2.13}$$

where $\mathbf{D}$ denotes a diagonal matrix with nonzero diagonal entries d_i, and $\mathbf{V}$ is a $k \times l$ matrix with $l \leq k-1$. Marsaglia (1963) showed that for the MVN case

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{R}) = \int_{\mathbb{R}^l} \phi_l(\mathbf{y}; \mathbf{I}_l) \int_{\mathbf{a}-\mathbf{V}\mathbf{y}}^{\mathbf{b}-\mathbf{V}\mathbf{y}} \phi_k(\mathbf{x}; \mathbf{D}) d\mathbf{x} d\mathbf{y}.$$

The inner integral can be written as a product of one-dimensional integrals. After the change of variables $\mathbf{x} = \mathbf{D}^{-1/2}\mathbf{z}$, the previous formula becomes

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{R}) = \tag{2.14}$$
$$\int_{\mathbb{R}^l} \phi_l(\mathbf{y}; \mathbf{I}_l) \prod_{i=1}^{k} \left[\Phi\left(\frac{b_i - \sum_{j=1}^{l} v_{ij} y_j}{\sqrt{d_i}} \right) - \Phi\left(\frac{a_i - \sum_{j=1}^{l} v_{ij} y_j}{\sqrt{d_i}} \right) \right] d\mathbf{y}.$$

Note that any correlation matrix can be written as $\mathbf{R} = e\mathbf{I}_k + \mathbf{V}\mathbf{V}^t$ with $l \leq k-1$, where e denotes the smallest eigenvalue of $\mathbf{R}$.

There is a natural generalization of formula (2.14) for the MVT problem in the form given by equation (1.3), which can be rewritten as

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{R}, \nu) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^{\infty} s^{\nu-1} e^{-\frac{s^2}{2}} \tag{2.15}$$
$$\int_{\mathbb{R}^l} \phi_l(\mathbf{y}; \mathbf{I}) \prod_{i=1}^{k} \left[\Phi\left(\frac{\frac{s b_i}{\sqrt{\nu}} - \sum_{j=1}^{l} v_{ij} y_j}{\sqrt{d_i}} \right) - \Phi\left(\frac{\frac{s a_i}{\sqrt{\nu}} - \sum_{j=1}^{l} v_{ij} y_j}{\sqrt{d_i}} \right) \right] d\mathbf{y} ds.$$

If $l = 1$, the problem is said to have *product correlation structure*. Problems with this form arise in a number of statistical applications (Dunnett, 1989). In this case, $\rho_{ij} = \lambda_i \lambda_j$ for $i \neq j$. If all $|\lambda_i| < 1$, then $\mathbf{R}$ can be written as $\mathbf{R} = \mathbf{D} + \mathbf{v}\mathbf{v}^t$, with $d_i = 1 - \lambda_i^2$ and $v_i = \lambda_i$, and equation (2.14) takes the simplified form

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{R}) = \int_{\mathbb{R}} \phi(y) \prod_{i=1}^{k} \left[\Phi\left(\frac{b_i - \lambda_i y}{\sqrt{1 - \lambda_i^2}} \right) - \Phi\left(\frac{a_i - \lambda_i y}{\sqrt{1 - \lambda_i^2}} \right) \right] dy. \tag{2.16}$$

Expressions similar to equation (2.16) were derived independently by several authors. We refer to Curnow and Dunnett (1962) and Marsaglia (1963) for additional references. If $\lambda_i = 1$ for some i, then the problem becomes a singular problem, see Section 5.2 for further details.

The computation of MVN probabilities in the form (2.16) reduces to the computation of a one-dimensional integral over $\mathbb{R}$ with a Gaussian weight function. Gauss-Hermite integration rules (Davis and Rabinowitz, 1984) can be used to approximate integrals in this form. Another method for this type of integral involves first applying the transformation $y = \Phi^{-1}(t)$ so that

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{R}) = \int_0^1 \prod_{i=1}^{k} \left[\Phi\left(\frac{b_i - \lambda_i \Phi^{-1}(t)}{\sqrt{1 - \lambda_i^2}} \right) - \Phi\left(\frac{a_i - \lambda_i \Phi^{-1}(t)}{\sqrt{1 - \lambda_i^2}} \right) \right] dt, \tag{2.17}$$

and then using a selected one-dimensional integration method for the finite integration interval $[0, 1]$.

In the equicorrelated case, where $\rho_{ij} = \rho$ for all i and j, equation (2.16) is valid for $\rho \geq 0$ with $\lambda_i = \sqrt{\rho}$. Steck and Owen (1962) have shown that (2.16) continues to hold for $\rho > -(k-1)^{-1}$, where the arising complex normal integral with argument $z = x + iy$ is defined by

$$\Phi(z) = \frac{1}{2\pi} e^{-\frac{y^2}{2}} \int_{-\infty}^{x} e^{-ity - \frac{t^2}{2}} dt,$$

with $i^2 = -1$. Extending this result, Nelson (1991) proved that (2.16) remains valid for $\rho_{ij} = -\lambda_i \lambda_j$ in the nonsingular case $\sum_{i=1}^{k} \lambda_i^2 / (1 + \lambda_i^2) < 1$ (negative product correlation structure). Nelson (1991) further tried to prove by induction that (2.16) is also valid in the singular case $\sum_{i=1}^{k} \lambda_i^2 / (1 + \lambda_i^2) = 1$ but only the induction step was completed. The missing analytical proof for $k = 2$ to start the induction was given by Soong and Hsu (1998). The latter authors also provided numerical details particular to the present complex integration problem. Further relationships to evaluate negative product correlated probabilities were given by Kwong (1995) and Kwong and Iglewicz (1996).

Yang and Zhang (1997) extended the above results to quasi-decomposable correlation matrices with $\rho_{ij} = \lambda_i \lambda_i + \tau_{ij}$, where τ_{ij} are nonzero deviations for some i and j. This case is also covered by the general formula (Marsaglia, 1963)

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{A} + \mathbf{B}) = \int_{\mathbb{R}^k} \phi_k(\mathbf{y}; \mathbf{B}) \int_{\mathbf{a}-\mathbf{y}}^{\mathbf{b}-\mathbf{y}} \phi_k(\mathbf{x}; \mathbf{A}) d\mathbf{x} d\mathbf{y}.$$

Curnow and Dunnett (1962) provided a method for reducing the dimension by a factor of two when $\rho_{ij} = \gamma_i/\gamma_j$ with $|\gamma_i| < |\gamma_j|$ for $i < j$.

2.3.2 Banded Correlation Matrices

Banded correlation matrices satisfy the condition $\rho_{ij} = 0$ whenever $|i-j| > l$, for some $l \geq 0$. The simplest nontrivial case is $l = 1$, where $\mathbf{\Sigma}$ is tri-diagonal. In this case, the Φ_k values have been called *orthoscheme probabilities*. Problems in this form have been studied by several authors (Schläfli, 1858; Abrahamson, 1964; Hayter and Liu, 1996; Miwa et al, 2003; Hayter, 2006; Craig, 2008).

If we determine the Cholesky decomposition for $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$, then $\mathbf{L}$ is a lower bi-diagonal matrix. After the transformation $\mathbf{x} = \mathbf{L}\mathbf{y}$, equation (1.1) becomes

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \int_{a_1/l_{11}}^{b_1/l_{11}} \phi(y_1) \int_{(a_2-l_{21}y_1)/l_{22}}^{(b_2-l_{21}y_1)/l_{22}} \phi(y_2) \cdots \int_{(a_k-l_{k,k-1}y_{k-1})/l_{k,k}}^{(b_k-l_{k,k-1}y_{k-1})/l_{k,k}} \phi(y_k) d\mathbf{y}.$$

If we define

$$g_k(y) = \Phi\left(\frac{b_k - l_{k,k-1}y}{l_{k,k}}\right) - \Phi\left(\frac{a_k - l_{k,k-1}y}{l_{k,k}}\right),$$

and

$$g_j(y) = \int_{(a_j-l_{j,j-1}y)/l_{j,j}}^{(b_j-l_{j,j-1}y)/l_{j,j}} \phi(t) g_{j+1}(t) dt,$$

for $j = k-1, k-2, \ldots, 2$, then

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \int_{a_1/l_{11}}^{b_1/l_{11}} \phi(t) g_2(t) dt.$$

If the $g_j(y)$ functions are successively computed for $j = k, k-1, \ldots, 2$ at selected y values using an appropriately chosen one-dimensional integration method, then the total computational work can be significantly reduced, compared to methods for the general MVN problem. If, for example, the integration method for each $g_j(y)$ value requires m integrand values, then the time complexity for an MVN computation is $O(km^2)$. An application of this method with cubic polynomial integration for the one-dimensional integrals is given in Miwa et al (2000). Craig (2008) has described further refinements of this recursive integration method with implementations using the fast Fourier transform to reduce the time complexity to $O(km\log(m))$. Miwa et al (2000) and Craig (2008) also show that any MVN cdf probability can be written as a combination of at most $(k-1)!$ orthoscheme probabilities. We discuss these methods in more detail in Section 4.1.4. Similar techniques are possible for MVT probabilities if the separation-of-variables method discussed in Section 4.1.2 is used.

If we consider the $l = 2$ case, then $\mathbf{\Sigma}$ is a quin-diagonal matrix, and the Cholesky factor $\mathbf{L}$ is lower tri-diagonal. Thus,

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int\limits_{a_1/l_{11}}^{b_1/l_{11}} \phi(y_1) \int\limits_{(a_2-l_{21}y_1)/l_{22}}^{(b_2-l_{21}y_1)/l_{22}} \phi(y_2) \int\limits_{(a_3-l_{31}y_1-l_{32}y_2)/l_{33}}^{(b_3-l_{31}y_1-l_{32}y_2)/l_{33}} \phi(y_3)$$

$$\cdots \int\limits_{(a_k-l_{k,k-2}y_{k-2}-l_{k,k-1}y_{k-1})/l_{k,k}}^{(b_k-l_{k,k-2}y_{k-2}-l_{k,k-1}y_{k-1})/l_{k,k}} \phi(y_n)d\mathbf{y}.$$

If we define

$$h_k(x,y) = \Phi\left(\frac{b_k - l_{k,n-2}x - l_{k,k-1}y}{l_{k,k}}\right) - \Phi\left(\frac{a_k - l_{k,k-2}x - l_{k,k-1}y}{l_{k,k}}\right),$$

$$h_j(x,y) = \int\limits_{(a_j-l_{j,j-2}x-l_{j,j-1}y)/l_{j,j}}^{(b_j-l_{j,j-2}x-l_{j,j-1}y)/l_{j,j}} \phi(t)h_{j+1}(y,t)dt,$$

for $j = k-1, k-2, \ldots 3$, and

$$h_2(y) = \int\limits_{(a_2-l_{21}y)/l_{22}}^{(b_2-l_{21}y)/l_{22}} \phi(t)h_3(y,t)dt,$$

then

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int\limits_{a_1/l_{11}}^{b_1/l_{11}} \phi(t)h_2(t)dt.$$

If tables of the $h_j(x,y)$ values are computed and the integration method for each $h_j(x,y)$ value requires m integrand values, then the time complexity for a MVN computation is $O(km^3)$. A similar technique is possible for MVT probabilities if the separation-of-variables method discussed in Section 4.1.2 is used. When k is large, a similar analysis shows that if $\mathbf{\Sigma}$ is a $(2l+1)$-diagonal matrix then an $O(km^{l+1})$ time complexity method can be constructed for the computation of MVN and MVT probabilities.

A related class of problems, where $\mathbf{\Sigma}^{-1}$ is banded, has also been studied. If $\mathbf{\Sigma}^{-1}$ is tridiagonal, then

$$\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x} = r_{11}x_1^2 + 2r_{21}x_1x_2 + r_{22}x_2^2 + \ldots + 2r_{k,k-1}x_{k-1}x_k + r_{kk}x_k^2,$$

where $\mathbf{\Sigma}^{-1} = (r_{ij})$, and

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \frac{1}{\sqrt{|\mathbf{\Sigma}|}} \int_{a_1}^{b_1} \phi(r_{11}x_1^2) \int_{a_2}^{b_2} \phi(2r_{21}x_1x_2 + r_{22}x_2^2) \cdots \int_{a_k}^{b_k} \phi(2r_{k,k-1}x_{k-1}x_k + r_{kk}x_k^2) d\mathbf{x}.$$

This class of MVN and MVT probabilities can also be computed using a sequence of iterated one-dimensional integrals, and the result is an $O(km^2)$ method if an m-point one-dimensional numerical integration method is used. Problems in this form have been studied by Genz and Kahaner (1986) and Craig (2008).

http://www.springer.com/978-3-642-01688-2

Computation of Multivariate Normal and t Probabilities

Genz, A.; Bretz, F.

2009, VIII, 126 p., Softcover
ISBN: 978-3-642-01688-2