195

Lecture Notes in Statistics

Alan Genz

Frank Bretz

Computation of Multivariate Normal and *t* Probabilities

Springer

Lecture Notes in Statistics **195**

Edited by P. Bickel, P.J. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

Alan Genz • Frank Bretz

Computation of Multivariate Normal and *t* Probabilities

Springer

Alan Genz
Department of Mathematics
Washington State University
Pullman, WA 99164-3113
USA
alangenz@wsu.edu

Frank Bretz
Hannover Medical School
and Novartis Pharma AG
WSJ-27.1.005
4002 Basel
Switzerland
frank.bretz@novartis.com

ISSN 0930-0325
ISBN 978-3-642-01688-2 e-ISBN 978-3-642-01689-9
DOI 10.1007/978-3-642-01689-9
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929712

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In this monograph we review numerical computation methods for multivariate normal and t probabilities while focusing on recent computer intensive integration methods. This monograph has two main objectives. First, we aim to convince the reader that these type of integration problems are computationally feasible for moderate accuracies and dimensions. Numerical results suggest that typical computation problems require only a few seconds of workstation time. Generalizations are available to singular problems, to domains other than the classical rectangular regions and to related integrals having multivariate normal and t weights. Special cases such as lower dimensional integrals or specific correlations structures allow quick and accurate computations for most practical purposes. Second, this monograph gives the reader a glimpse into current multidimensional integration techniques. This work might therefore also serve as a guide for other integration problems.

The monograph is organized as follows. In Chapter 1 we motivate the problem and offer a historical perspective to show how methodological approaches have evolved over time. We then introduce the integration problems and some related notation. We conclude this chapter with several numerical examples, which are used later to illustrate the methods. Chapter 2 is devoted to special cases, for which the integration problem can be simplified. We first consider the work done on bivariate and trivariate normal and t probabilities. We then consider calculating probabilities over special integrations regions, such as orthants, ellipsoids, and hyperboloids. Finally, we review multivariate normal and t probability problems for special correlation structures. In Chapter 3 we describe inexact methods for approximating general multivariate normal and t probabilities by one or more integration problems which have easier solutions than the original problem. This includes, for example, methods that are based on Boole's formula, which uses combinations of lower dimensional problems to approximate the original problem. In Chapter 4 we describe approximation methods for the general integration problems that can lead to exact results given sufficient computational resources. These methods are based on reparameterizing the integrals of the original problems, thus en-

abling the use of efficient numerical integration methods, including stochastic and deterministic methods. In Chapter 5, some related and application specific topics are considered, such as singular distributions, integration problems having an application specific expectation function with multivariate normal or t weight, and a review of numerical test results. A description of current software implementations in MATLAB and R is also given. Finally, in Chapter 6 we illustrate the theoretical results from the previous chapters with numerical examples from different applications, including multiple comparison procedures, Bayesian statistics and computational finance.

Acknowledgments. We are very grateful to Torsten Hothorn for implementing and maintaining the R package `mvtnorm`. We are also grateful to Klaus Strassburger for his valuable comments on an earlier version of this monograph.

Pullman, Basel, *Alan Genz*
March 2009 *Frank Bretz*

Contents

1

Introduction

1.1 A Historical Perspective

The numerical availability of distribution functions is important for many statistical applications. The numerical computation of multivariate distribution functions is considerably more complex than the numerical evaluation of univariate distributions due to the "curse" of dimensionality. This is particularly true for the multivariate normal (MVN) and t (MVT) distributions. These distributions occur frequently in many different applications, including multiple comparison procedures (Hochberg and Tamhane, 1987; Hsu, 1996), integrated log-likelihood problems (Lin and Breslow, 1996; Hartzel et al, 2001), applications of the multivariate probit model (Ochi and Prentice, 1984), Bayesian computations (Shaw, 1988; Evans and Swartz, 1995), stochastic programming (Prékopa, 1995), structural system reliability (Pandey, 1998) as well as many applications in engineering and finance. For a general overview of the MVN and the MVT distributions we refer to the books of Tong (1990), Kotz et al (2000) and Kotz and Nadarajah (2004).

Although considerable research has been devoted to the computation of MVN, MVT and related probabilities, reliable and highly accurate numerical algorithms are not yet available for many MVN and MVT problems. Initial work on the computation of MVN probabilities was done as early as 1858, when Schläfli (1858) studied orthant probabilities using differential equations similar to those used by Plackett (1954) one hundred years later. Mehler (1866) derived a series expansion for the bivariate normal case, which later became known as the Kibble series for the general MVN case. Similarly, Pearson (1901) introduced the tetrachoric series expansion in the correlation coefficient ρ for the bivariate normal distribution function. Sheppard (1900) obtained exact bivariate probabilities over positive orthants. But broad research on the MVN problem began only in the mid-20th century with the work of Kendall (1941), Kibble (1945), David (1953), van der Vaart (1953) and Plackett (1954) among others. At the same time, the MVT distribution was introduced independently by Cornish (1954) and Dunnett and Sobel (1954).

A. Genz, *Computation of Multivariate Normal and t Probabilities*
DOI: 10.1007/978-3-642-01689-9_1, © Springer-Verlag Berlin Heidelberg 2009

The year 1980 was key in the history of MVN and MVT computations. Previously, most research focused on asymptotic density expansions and related approaches. With the negative results published by Harris and Soms (1980), later research on MVN computations changed direction drastically. In the same year, Deák (1980) published his article on three-digits accuracy by Monte Carlo techniques. Although applications of Monte Carlo methods on MVN problems had been studied much earlier (Escoufier, 1967), these methods became more popular with the rise of modern computer technology in the early eighties. Modern methods for MVN and MVT computations often use Monte Carlo and modified Monte Carlo methods. There has been significant progress since the last major review of MVN (and MVT) computational methods (Martynov, 1981). This monograph summarizes some of the recent developments in this quickly evolving area.

1.2 Problem Setting

We consider integrals of the general form

$$\int_{\mathbf{A}} f(\mathbf{x})g(\mathbf{x})d\mathbf{x},$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_k)^t \in \mathbb{R}^k$, $\mathbf{A} \subseteq \mathbb{R}^k$, $f(\mathbf{x})$ is either the MVN or the MVT (central or non-central) density function and $g(\mathbf{x})$ is an application specific expectation function. If $g(\mathbf{x}) = 1$, the problem reduces to the ordinary MVN and MVT integrals. The integration region $\mathbf{A}$ is typically a hyperrectangle

$$[\mathbf{a}, \mathbf{b}], -\infty \leq a_i < b_i \leq \infty, i = 1, \ldots, k.$$

Many applications, however, involve more general integration regions, such as intersections of half-spaces or elliptically contoured regions.

The major focus of this monograph is on $g(\mathbf{x}) = 1$ and $\mathbf{A}$ defined as above. Let $X_1, \ldots, X_k$ be jointly MVN or MVT distributed random variables. We are then interested in computing the probability

$$P(\mathbf{A}) = P(a_1 \leq X_1 \leq b_1, \ldots, a_k \leq X_k \leq b_k).$$

The corresponding *MVN integral* is given by

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \frac{1}{\sqrt{|\mathbf{\Sigma}|(2\pi)^k}} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \ldots \int_{a_k}^{b_k} e^{-\frac{1}{2}\mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x}} dx_k \ldots dx_1, \tag{1.1}$$

where $\mathbf{\Sigma}$ is a symmetric, positive definite $k \times k$ covariance matrix. In this monograph, we do not distinguish whether $\mathbf{\Sigma} = (\sigma_{ij})$ or its standardized form as a correlation matrix $\mathbf{R} = (\rho_{ij})$ is used. If $\mathbf{\Sigma}$ is positive semidefinite and

consequently $|\mathbf{\Sigma}| = 0$, the integral is called singular and will be discussed separately in Section 5.2.

The *non-central MVN integral* is given by

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \boldsymbol{\delta}) = \frac{1}{\sqrt{|\mathbf{\Sigma}|(2\pi)^k}} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\delta})^t \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\delta})} dx_k \dots dx_1,$$

where $\boldsymbol{\delta}$ denotes the $k \times 1$ non-centrality vector with $-\infty < \delta_i < \infty$. Because

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \boldsymbol{\delta}) = \Phi_k(\mathbf{a} - \boldsymbol{\delta}, \mathbf{b} - \boldsymbol{\delta}; \mathbf{\Sigma}),$$

any non-central MVN integral can be determined from a central MVN integral with shifted integration limits. We also note that a general non-central MVN probability $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \boldsymbol{\delta})$ can be transformed to a standardized MVN probability. If $\mathbf{D}$ denotes the diagonal matrix which has the square roots of the diagonal entries for $\mathbf{\Sigma}$ on its diagonal, the correlation matrix $\mathbf{R}$ is defined by $\mathbf{\Sigma} = \mathbf{DRD}$. Then the transformation $\mathbf{x} = \mathbf{Dy} + \boldsymbol{\delta}$ reduces the general non-central MVN probability to

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \boldsymbol{\delta}) = \Phi_k(\mathbf{D}^{-1}(\mathbf{a} - \boldsymbol{\delta}), \mathbf{D}^{-1}(\mathbf{b} - \boldsymbol{\delta}); \mathbf{R}).$$

We will sometimes refer to the associated probability density function (pdf) in (1.1) as $\phi_k(\mathbf{x}; \mathbf{\Sigma})$ or $\phi_{k,\mathbf{\Sigma}}$. In the case where $\mathbf{a} = -\infty$ (i.e., $a_i = -\infty, i = 1, \dots, k$), equation (1.1) defines a multivariate normal cumulative distribution function (cdf). A related special case with $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = \infty$ is sometimes called a MVN *orthant probability*. The univariate standard cdf and pdf are denoted by $\Phi(x)$ and $\phi(x)$, respectively.

The *central MVT integral* is given by

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})\sqrt{|\mathbf{\Sigma}|(\nu\pi)^k}} \int_{a_1}^{b_1} \dots \int_{a_k}^{b_k} \left(1 + \frac{\mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+k}{2}} d\mathbf{x} \quad (1.2)$$

$$= \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^\infty s^{\nu-1} e^{-\frac{s^2}{2}} \Phi_k\left(\frac{s\mathbf{a}}{\sqrt{\nu}}, \frac{s\mathbf{b}}{\sqrt{\nu}}; \mathbf{\Sigma}\right) ds, \quad (1.3)$$

where $\nu \in \mathbb{R}^+$ is a scale (or degrees-of-freedom) parameter. The second form (1.3) allows any method for the MVN problem to be applied to the MVT problem at the cost of an additional integration; see Cornish (1954) for the original definition and a derivation of the equivalence between the two forms.

We will also sometimes refer to the associated pdf in (1.2) as $t_k(\mathbf{x}; \mathbf{\Sigma}, \nu)$. The univariate standard t cdf and pdf, with scale parameter ν, are denoted by $T(x; \nu)$ and $t(x; \nu)$, respectively, with

$$T(x;\nu) = \int_{-\infty}^{x} t(s;\nu)ds$$
$$= \int_{-\infty}^{x} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1+\frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds.$$

One possible definition for a *non-central MVT integral* is

$$T_k(\mathbf{a},\mathbf{b};\boldsymbol{\Sigma},\nu,\boldsymbol{\delta}) =$$
$$\frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})\sqrt{|\boldsymbol{\Sigma}|(\nu\pi)^k}} \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \left(1+\frac{(\mathbf{x}-\boldsymbol{\delta})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\delta})}{\nu}\right)^{-\frac{\nu+k}{2}} d\mathbf{x}.$$

Any non-central MVT integral in this form can be determined from a central MVT integral with shifted limits. An alternate definition of the non-central MVT integral is based on equation (1.3) and given by

$$T_k(\mathbf{a},\mathbf{b};\boldsymbol{\Sigma},\nu,\boldsymbol{\delta}) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^{\infty} s^{\nu-1}e^{-\frac{s^2}{2}} \Phi_k\left(\frac{s\mathbf{a}}{\sqrt{\nu}}-\boldsymbol{\delta}, \frac{s\mathbf{b}}{\sqrt{\nu}}-\boldsymbol{\delta};\boldsymbol{\Sigma}\right) ds; \quad (1.4)$$

see Owen (1965) for the bivariate case.

Note that throughout this monograph we consider the most common form of the MVT distribution, which is defined as dividing a MVN distributed vector by a common χ_ν variable. Other forms of the central and non-central MVT distribution are discussed in Johnson and Kotz (1972) and Kotz and Nadarajah (2004).

1.3 Some Examples

Throughout the rest of the monograph we will use a series of numerical examples to illustrate the different methods and types of problems that we discuss. This section provides a collection of integration problems that we will refer to later.

1.3.1 Simple Three-Dimensional Examples

To start with, let (X_1, X_2, X_3) be trivariate normal distributed with correlation matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \frac{3}{5} & \frac{1}{3} \\ \frac{3}{5} & 1 & \frac{11}{15} \\ \frac{1}{3} & \frac{11}{15} & 1 \end{bmatrix}$$

and non-centrality vector $\boldsymbol{\delta} = (0, 0, 0)^t$. Assume that we are interested in calculating the probability

$$P(-\infty < X_1 \le 1, -\infty < X_2 \le 4, -\infty < X_3 \le 2).$$

The trivariate normal probability is then

$$\begin{aligned} \Phi_{3ex} &= \frac{1}{\sqrt{|\boldsymbol{\Sigma}|(2\pi)^3}} \int_{-\infty}^{1} \int_{-\infty}^{4} \int_{-\infty}^{2} e^{-\frac{1}{2}\mathbf{x}^t \boldsymbol{\Sigma}^{-1}\mathbf{x}} dx_3 dx_2 dx_1 \qquad (1.5) \\ &\approx 0.827984897456834. \end{aligned}$$

A related MVT example, with $\nu = 5$, is

$$\begin{aligned} T_{3ex} &= \frac{\Gamma(4)}{\Gamma(\frac{5}{2})\sqrt{|\boldsymbol{\Sigma}|(5\pi)^3}} \int_{-\infty}^{1} \int_{-\infty}^{4} \int_{-\infty}^{2} \left(1 + \frac{\mathbf{x}^t \boldsymbol{\Sigma}^{-1}\mathbf{x}}{5}\right)^{-4} dx_3 dx_2 dx_1 \qquad (1.6) \\ &\approx 0.791453793811934. \end{aligned}$$

We also consider a singular MVN problem given through the covariance matrix

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

with $\mathbf{a} = (-\infty, -\infty, -\infty)^t$ and $\mathbf{b} = (1, 1, 1)^t$. In this case the trivariate problem essentially reduces to a univariate problem and

$$\Phi_{3sex} = \Phi(1) \approx 0.841344746068543. \qquad (1.7)$$

1.3.2 Five-Dimensional Examples

For a second set of examples, consider a problem over the hyper-rectangle $[\mathbf{a}, \mathbf{b}]$ with $\mathbf{a} = (-1, -2, -3, -4, -5)^t$, $\mathbf{b} = (2, 3, 4, 5, 6)^t$ and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}.$$

The associated pentavariate normal probability is

$$\begin{aligned} \Phi_{5ex} &= \frac{1}{\sqrt{|\boldsymbol{\Sigma}|(2\pi)^5}} \int_{-1}^{2} \int_{-2}^{3} \int_{-3}^{4} \int_{-4}^{5} \int_{-5}^{6} e^{-\frac{1}{2}\mathbf{x}^t \boldsymbol{\Sigma}^{-1}\mathbf{x}} dx_5 dx_4 dx_3 dx_2 dx_1 \\ &\approx 0.4741284. \qquad (1.8) \end{aligned}$$

A related integration problem over a region within the positive orthant has $\mathbf{a} = (0, 0, 0, 0, 0)^t$ with

$$\Phi_{5oex} = \Phi_5(\mathbf{0}, \mathbf{b}; \mathbf{\Sigma}) \approx 0.11353418,$$

and the related cdf example has $\mathbf{a} = -\infty$ with

$$\Phi_{5cex} = \Phi_5(-\infty, \mathbf{b}; \mathbf{\Sigma}) \approx 0.81031466.$$

A related MVT example, with $\nu = 8$, is

$$T_{5ex} = \frac{\Gamma(\frac{13}{2})}{\Gamma(4)\sqrt{|\mathbf{\Sigma}|(8\pi)^5}} \int_{-5}^{6}\int_{-4}^{5}\int_{-3}^{4}\int_{-2}^{3}\int_{-1}^{2} \left(1 + \frac{\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}}{8}\right)^{-\frac{13}{2}} dx_5 dx_4 dx_3 dx_2 dx_1,$$
$$\approx 0.447862.$$

1.3.3 Eight-Dimensional Examples

Finally, we consider the eight-dimensional problem

$$\Phi_{8ex} = \frac{1}{\sqrt{|\mathbf{\Sigma}|(2\pi)^5}} \int_{-1}^{2}\int_{-2}^{3}\int_{-3}^{4}\int_{-4}^{5}\int_{-5}^{6}\int_{-6}^{7}\int_{-7}^{8}\int_{-8}^{9} e^{-\frac{1}{2}\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}} d\mathbf{x}$$
$$\approx 0.32395, \tag{1.9}$$

with

$$\mathbf{\Sigma} = \begin{bmatrix} 1&1&1&1&1&1&1&1 \\ 1&2&2&2&2&2&2&2 \\ 1&2&3&3&3&3&3&3 \\ 1&2&3&4&4&4&4&4 \\ 1&2&3&4&5&5&5&5 \\ 1&2&3&4&5&6&6&6 \\ 1&2&3&4&5&6&7&7 \\ 1&2&3&4&5&6&7&8 \end{bmatrix}.$$

The related example over the positive orthant has $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = \infty$ with

$$\Phi_{8oex} = \Phi_8(\mathbf{0}, \infty; \mathbf{\Sigma}) \approx 0.076586,$$

and the related cdf example has $\mathbf{a} = -\infty$ with

$$\Phi_{8cex} = \Phi_8(-\infty, \mathbf{b}; \mathbf{\Sigma}) \approx 0.69675.$$

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, © Springer-Verlag Berlin Heidelberg 2009

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),

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

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

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

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

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

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

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

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,

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

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 $||\mathbf{Wz}||^{-k}$, where $\mathbf{z}$ is distributed on a $(k-1)-$dimensional manifold and $\mathbf{W}$ is the decomposition $\mathbf{W} = \mathbf{D}\,\mathrm{diag}(||\mathbf{d}_1||^{-1},\ldots,||\mathbf{d}_k||^{-1})$ 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

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

which can be written in terms of Φ_k as

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

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

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

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

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

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

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

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,

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int_{a_1/l_{11}}^{b_1/l_{11}} \phi(y_1) \int_{(a_2-l_{21}y_1)/l_{22}}^{(b_2-l_{21}y_1)/l_{22}} \phi(y_2) \int_{(a_3-l_{31}y_1-l_{32}y_2)/l_{33}}^{(b_3-l_{31}y_1-l_{32}y_2)/l_{33}} \phi(y_3) \cdots \int_{(a_k-l_{k,k-2}y_{k-2}-l_{k,k-1}y_{k-1})/l_{k,k}}^{(b_k-l_{k,k-2}y_{k-2}-l_{k,k-1}y_{k-1})/l_{k,k}} \phi(y_n)d\mathbf{y}.$$

If we define

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

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

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

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

then

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

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).

3

Methods That Approximate the Problem

In this chapter we approximate the original MVN or MVT problem by using one or more MVN or MVT problems which are relatively easier to solve. In Section 3.1 we consider various approximations to Boole's formula, providing MVN and MVT approximations that consist of combinations of lower dimensional problems. In Section 3.2 we briefly describe methods that replace the original problem with one or more problems using covariance matrix structures that simplify the computation. In Section 3.3 we review further approximations. Finally, in Section 3.4 we review some asymptotic expansions of historical interest.

3.1 Boole's Formula Approximations and Bounds

As in Chapter 1, we define $A_i = \{X_i : a_i \le X_i \le b_i\}$. Let A_i^c denote the complement of A_i, and let further $P(\mathbf{A})$ denote $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{R})$ or $T_k(\mathbf{a}, \mathbf{b}; \mathbf{R}, \nu)$. If we use

$$P(\mathbf{A}) = P\left(\bigcap_{j=1}^{k} A_j\right) = 1 - P\left(\bigcup_{j=1}^{k} A_j^c\right),$$

then Boole's formula for $P(\mathbf{A})$ can be written in the form

$$P(\mathbf{A}) = 1 - \sum_{j=1}^{k} P(A_j^c) + \sum_{j<i} P(A_j^c \cap A_i^c) + \ldots + (-1)^k P\left(\bigcap_{i=1}^{k} A_i^c\right).$$

The simplest approximation to $P(\mathbf{A})$ in this form uses only the constant term and the univariate probabilities. If we let

$$S_1 = \sum_{j=1}^{k} P(A_j^c),$$

A. Genz, *Computation of Multivariate Normal and t Probabilities*
DOI: 10.1007/978-3-642-01689-9_3, © Springer-Verlag Berlin Heidelberg 2009

we obtain the well-known Bonferroni lower bound

$$1 - S_1 \leq P(\mathbf{A}).$$

A simple upper bound based only on univariate probabilities is

$$P(\mathbf{A}) \leq 1 - \min_j P(A_j^c).$$

Better approximations and sharper bounds result if using bivariate probabilities. The Hunter-Worsley bound (Hunter, 1976; Worsley, 1982) guarantees that

$$L_2 = 1 - S_1 + \sum_{(i,j)\in T} P(A_i^c \cap A_j^c) \leq P(\mathbf{A}),$$

where T is a maximal spanning tree for the complete graph of order k with edge weights $P(A_i^c \cap A_j^c)$; but see Maurer (1983) for improved results by using bivalent trees. An upper bound, derived by Dawson and Sankoff (1967), is given by

$$P(\mathbf{A}) \leq U_2 = 1 - 2(S_1 - S_2/m)/(m+1),$$

where $m = 1 + \lfloor 2S_2/S_1 \rfloor$ and

$$S_2 = \sum_{i<j} P(A_i^c \cap A_j^c).$$

We can determine even better approximations using higher order bounds based on hyper-trees (Tomescu, 1986). Bukszár and Prékopa (2001) showed a cherry tree (hyper-tree of order three) bound always exists which is better than the Hunter-Worsley bound. These *hybrid bounds*, which we denote by $L_{(2,3)}$, are lower bounds for $P(\mathbf{A})$ that require only $k-2$ trivariate values. Determining the optimal cherry tree bound requires evaluating all of the $k(k-1)(k-2)/6$ trivariate $P(A_m^c \cap A_j^c \cap A_i^c)$ values, but a bound that is often optimal or nearly optimal is defined by

$$L_{(2,3)} = L_2 + \sum_{(i,j)\in T^*} (P(A_j^c \cap A_i^c) - P(A_{m_{(i,j)}}^c \cap A_j^c \cap A_i^c))$$

The edge set T^* contains $k-2$ of the $k-1$ edges in the L_2 bound edge set T. For each edge in T^*, the vertex $m_{(i,j)}$ is selected using a method described by Bukszár and Prékopa (2001). It is always true that $L_{(2,3)} \geq L_2$, because

$$P(A_j^c \cap A_i^c) \geq P(A_{m_{(i,j)}}^c \cap A_j^c \cap A_i^c).$$

If we define S_3 by

$$S_3 = \sum_{m<j<i} P(A_m^c \cap A_j^c \cap A_i^c),$$

then sharp bounds that use S_1, S_2 and S_3 (Boros and Prékopa, 1989) are given by

$$\begin{aligned} L_3 &= 1 - S_1 + 2\frac{(2j-1)S_2 - 3S_3}{j(j+1)} \\ &\leq P(\mathbf{A}) \leq U_3 = 1 - \frac{(i+2k-1)S_1 - 2((2i+k-2)S_2 - 3S_3)/i}{k(i+1)}, \end{aligned}$$

where

$$i = 1 + \lfloor 2((k-2)S_2 - 3S_3)/((k-1)S_1 - 2S_2) \rfloor \text{ and } j = 2 + \lfloor 3S_3/S_2 \rfloor.$$

Unfortunately, it is not always true that $L_2 \leq L_3$, but Boros and Prékopa (1989) showed that $U_3 \leq U_2$. The bounds L_3 and U_3 require evaluating all of the $k(k-1)(k-2)/6$ trivariate probabilities, which might be expensive to compute for higher dimensional problems. The L_3 bound is more expensive to compute than $L_{(2,3)}$ (which does not require all of the trivariate probabilities) and often $L_{(2,3)}$ is sharper than L_3. All of the bounds become sharper if the integration limits defining $\mathbf{A}$ approach $\pm\infty$.

Tomescu (1986) described a class of hybrid upper bounds. Let

$$U_{(2,3)} = 1 - S_1 + S_2 - \sum_{E(T_k^3)} P(A_m^c \cap A_j^c \cap A_i^c),$$

where $E(T_k^3)$ is the set of $(k-1)(k-2)/2$ hyperedges (i,j,k) for a cherry tree. Tomescu showed that $P(\mathbf{A}) \leq U_{(2,3)}$. The optimal hypertree bound could also be expensive to compute because all possible trivariate distribution values are needed, but a bound that is often nearly optimal can be determined using the maximal spanning tree T. The $k-2$ terminal vertices from T are successively deleted. Each time a terminal vertex (along with its edge) is deleted that vertex is adjoined to each remaining T edge to form a set of hyperedges. The union of the sets of $k-2, k-3, \ldots, 1$ hyperedges found in this way form the hypertree T_k^3 that is used for $U_{(2,3)}$. This bound is always less than or equal to the second order Bonferroni $1 - S_1 + S_2$ bound but is not necessarily less than or equal to U_2 or U_3.

In practice, if a lower bound L and an upper bound U have been computed, the quantity $(L+U)/2$ is an approximation to $P(\mathbf{A})$ with maximum error of $E = (U-L)/2$. If this error is small enough for the chosen application, then no further computations are necessary. A useful strategy involves first computing L_1 and U_1, checking the error, and then successively computing L_2 and U_2, $L_{(2,3)}$, and L_3 and U_3 (if the dimension is not too large). At each stage the error bound can be computed, and the calculation terminated if the error is small enough. If none of the errors is small enough, then a more accurate method will be necessary (see Chapter 4).

As an example, consider the five-dimensional example problem (1.8) where

$$(L_2, L_3, L_{(2,3)}, \hat{P}, U_{(2,3)}, U_3, U_2) = (0.4732, 0.4719, 0.474113, 0.474129, 0.474146, 0.4937, 0.526),$$

with $\hat{P} = [\max(L_2, L_3, L_{(2,3)}) + \min(U_{(2,3)}, U_3, U_2)]/2$ and associated error bound $\hat{E} = [\min(U_{(2,3)}, U_3, U_2) - \max(L_2, L_3, L_{(2,3)})]/2 = 0.000016$. These results are consistent with results from methods described in Chapter 4, where a more accurately computed $\bar{P} = 0.4741284$. The related five-dimensional cdf example problem, with all lower limits $a_i = -\infty$, has

$$(L_2, L_3, L_{(2,3)}, \hat{P}, U_{(2,3)}, U_3, U_2) = (0.81012, 0.81014, 0.810313, 0.810315, 0.810316, 0.814, 0.820),$$

with error bound $\hat{E} = 0.0000013$, and $\bar{P} = 0.81031466$. The eight-dimensional example problem (1.9) has

$$(L_2, L_3, L_{(2,3)}, \hat{P}, U_{(2,3)}, U_3, U_2) = (0.28, 0.16, 0.296, 0.343, 0.404, 0.398, 0.448),$$

with $\hat{E} = 0.046$, and $\bar{P} = 0.32395$. The related eight-dimensional cdf example problem, with all lower limits $a_i = -\infty$, has

$$(L_2, L_3, L_{(2,3)}, \hat{P}, U_{(2,3)}, U_3, U_2) = (0.689, 0.632, 0.693, 0.700, 0.708, 0.726, 0.748),$$

with $\hat{E} = 0.008$, and $\bar{P} = 0.69675$.

Bukszár (2001) and Szántai (2001) compared several lower and upper bounds for MVN probabilities as an application of hyper-trees. Naiman and Wynn (1992, 1997) gave sufficient conditions on improving the classical inclusion-exclusion identities and inequalities by decreasing the maximum depth and the number of terms at each depth. The resulting inequalities are guaranteed to be at least as sharp as the classical approaches. Further developments and applications of these improved inequalities are given for example by McCann and Edwards (1996, 2000) and Dohmen (2000). An application of these bounds for the approximate calculation of MVT probabilities for multiple comparison problems is discussed in Genz and Bretz (2000) and Genz et al (2004), see also Section 6.1.2. For further references on probability inequalities we refer to Tong (1980, 1990), Galambos and Simonelli (1996), and Dohmen (2003).

3.2 Correlation Matrix Approximations

Several authors have considered replacing a given correlation matrix $\mathbf{R}$ by a more convenient $\mathbf{R}'$, so that equation (2.16) may be used. Based on theoretical and numerical justifications, Iyengar (1988) and Iyengar and Tong (1989) proposed replacing the individual correlations by their common average. Similar

application-specific approximations based on different averaging techniques for the correlation coefficients can be found in Hochberg and Tamhane (1987, p. 141) and Hochberg (1994). Royen (1987) and Hsu (1992) independently provided different techniques to find the "closest" $\mathbf{R}'$ with a product correlation structure $\rho_{ij} = \lambda_i \lambda_j$. Factor-analytic and linear programming methods that find the "closest" $\mathbf{R}'$ with a product correlation structure were investigated by Hsu (1992) and Hsu and Nelson (1998). These methods typically do not provide information about the approximation error. Some numerical tests, however, have often shown a reasonable overall performance.

3.3 Other Approximations

In this section we review other approximation methods, which do not quite fit into the framework of the previous sections. Several authors have investigated the decomposition of $P(\mathbf{A})$ into a product of conditional probabilities,

$$P(\mathbf{A}) = P\left(\bigcap_{i=1}^{m} A_i\right) \prod_{i=m+1}^{k} P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right), \tag{3.1}$$

for a fixed $m \in \{1, \ldots k\}$. Different approximations to the conditional probabilities were proposed. Pearson (1903) suggested approximating the bivariate normal distribution by

$$P(X_1 > a_1, X_2 > a_2) = \Phi(-a_1)\Phi\left(\frac{\mu_{2|1} - a_2}{\sigma_{2|1}}\right),$$

where $\mu_{2|1}$ and $\sigma_{2|1}$ are the conditional mean and variance of X_2 given $X_1 > a_1$. Mendell and Elston (1974) generalized this early result to the multivariate case by the recursion

$$P(\mathbf{A}) = \prod_{i=1}^{k} \Phi(-Z_{i|i-1}),$$

with the standardized thresholds

$$Z_{i|j} = \frac{Z_{i|j-1} - \mu_{j|j-1} r_{ij|j-1}}{\sigma_{j|j-1}},$$

starting with $Z_{1|0} = a_1$, and using

$$\mu_{j|j-1} = \frac{\phi(Z_{j|j-1})}{\Phi(-Z_{j|j-1})},$$

$$\sigma^2_{i|j} = 1 - r^2_{ji|j-1} \mu_{j|j-1}(\mu_{j|j-1} - Z_{j|j-1})$$

and

$$r_{mn|j} = \frac{r_{mn|j-1} - r_{jm|j-1} r_{jn|j-1} \mu_{j|j-1} (\mu_{j|j-1} - Z_{j|j-1})}{\sigma_{m|j} \sigma_{n|j}},$$

with $r_{mn|0} = \rho_{mn}$. Rice et al (1979) introduced minor modifications of this basic procedure. Kamakura (1989) compared these approaches with the approximation of Clark (1961) and the separated split method of Langdon (1984a,b). (This latter approach is based on the observation that the decomposition (3.1) is not unique and selects the "best" permutation with minimum variance.)

Solow (1990) investigated (3.1) for $m = 2$ and approximated each conditional probability using a linear regression model. Let I_A denote the indicator function of the event A. Then,

$$P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j \right) = E(I_{A_i} | I_{A_j} = 1, j = 1, ..., i-1)$$

is approximated by

$$E(I_{A_i}) + \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} (1 - E(I_{A_1}), \ldots, 1 - E(I_{A_{i-1}}))^t$$

where $\mathbf{\Sigma}_{21}$ is a row vector with the entries

$$\mathrm{cov}(I_{A_i}, I_{A_j}) = E(I_{A_i} I_{A_j}) - E(I_{A_i}) E(I_{A_j})$$

and $\mathbf{\Sigma}_{11}$ is a $(i-1) \times (i-1)$ matrix with the entries

$$\mathrm{cov}(I_{A_{j_1}}, I_{A_{j_2}}) = E(I_{A_{j_1}} I_{A_{j_2}}) - E(I_{A_{j_1}}) E(I_{A_{j_2}}),$$

for $1 \le j, j_1, j_2 \le i-1$. Note that

$$E(I_{A_i}) = \Phi(b_i) - \Phi(a_i)$$

and

$$E(I_{A_i} I_{A_j}) = E(I_{A_i \cap A_j}) = \Phi_2(\mathbf{a}, \mathbf{b}; \rho).$$

Joe (1995) extended this approach using tri- and quadrivariate marginal probabilities, that is, $m = 4$ in (3.1). To overcome the non-uniqueness of the decomposition (3.1), Joe proposed averaging over all approximation results of the $k!/m!$ equivalent permutation to (3.1). Tests by Joe showed that the resulting approximations are often very accurate for selected test problems, although these methods do not provide direct information about the approximation error.

3.4 Asymptotic Expansions

As indicated in Chapter 1, asymptotic expansions played a major historical role in evaluating MVN integrals. In particular, the tetrachoric series has often

been referenced as a series method for MVN integrals. Let $H_n(x)$ denote the Chebyshev-Hermite polynomial of degree n satisfying

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)\phi(x)dx = \begin{cases} n!, \, n = m, \\ 0, \, n \neq m, \end{cases}$$

(Stuart and Ord, 1987) and let $H_{-1}(x) = (1 - \Phi(x))/\phi(x)$. Kendall (1941) proposed the asymptotic expansion

$$\Phi_k(\mathbf{a}, \infty; \mathbf{R}) = \sum_{n_{12}=0}^{\infty} \cdots \sum_{n_{k-1,k}=0}^{\infty} \prod_{1 \leq i < j \leq k} \frac{\rho_{ij}^{n_{ij}}}{n_{ij}!} \prod_{i=1}^{k} H_{n_i - 1}(a_i)\, \phi(a_i) \qquad (3.2)$$

where $n_i = \sum_{j:i<j} + \sum_{j:j<i}$. The right hand side of equation (3.2) is called tetrachoric series (Pearson, 1901). Kibble (1945) obtained the corresponding series for the MVN density function. In the literature prior to 1980, the tetrachoric series has often been introduced with the comment that convergence may be slow unless the $|\rho_{ij}|$ are small for all i and j. Kendall (1941) noted that equation (3.2) "... *will be found to be convergent.*" Similarly, Kibble (1945) stated in view of his series "... *is absolutely convergent for all real values of the variables if* $\mathbf{R}$ *is positive definite.*" Harris and Soms (1980) reexamined above series and showed that in general both the tetrachoric and the Kibble series can diverge. They proved that both series converge absolutely if $|\rho_{ij}| < 1/(k-1)$. The authors introduced several counterexamples, showing among other things that the tetrachoric series (3.2) will diverge for $a_i = 0, i = 1, \ldots, k$, k even, whenever $|\rho_{ij}| > 1/(k-1)$.

Only a few methods have been published after 1980 using asymptotic expansions. Moran (1983) derived a new expansion using the elements of $\mathbf{R}^{-1}$. The author partially identified regions of convergence and divergence, which are somewhat different from those for the tetrachoric series. Because of the restricted convergence properties of the classical approaches, Royen (1987) proposed shifting the center of the series to a correlation matrix $\mathbf{R}'$ close to the given $\mathbf{R}$. To simplify the computations, $\mathbf{R}'$ is assumed to be of $l-$factorial structure (2.13). Royen (1987) proved that under these assumptions the series is absolutely convergent for $|h_{ij}| \leq \boldsymbol{\tau}_i \boldsymbol{\tau}_j/(k-1)$, with

$$\mathbf{R} - \mathbf{R}' = (h_{ij})_{ij} \quad \text{and} \quad \boldsymbol{\tau}_i = \sqrt{(1 - |\mathbf{v}_i|)^2},$$

where the $\mathbf{v}_i$ are the rows of $\mathbf{V}$ in (2.13). Other Taylor expansions were proposed by Olson and Weissfeld (1991), Wang and Kennedy (1992, 1997), and Dash (2007). Breitung and Hohenbichler (1989) and Breitung and Richter (1996) discuss asymptotic approximations to large deviation probabilities involving MVN variables.

4

Methods That Approximate the Integral

In this chapter we describe selected numerical integration methods that can be applied to general MVN and MVT problems. Such problems are often given as a multiple integral over an unbounded integration domain. The application of most integration methods requires a transformation or reparameterization of the original problem to one where the integration domain is bounded. Consequently, we first discuss various reparameterizations in Section 4.1. In Section 4.2 we then describe several multidimensional integration methods.

4.1 Reparameterizations

4.1.1 Spherical-Radial Transformation Methods

A key issue for many numerical integration methods is a suitable reparameterization of the original integral. The aim is to transform mode and scale appropriately, thus making the integrand more suitable for integration. In addition, efficient integration methods may require the variables to range over a particular set. The first class of methods that we discuss uses the combination of an initial standardization with a *spherical-radial* (SR) transformation. These reparameterizations have been developed independently by several authors for MVN integrals (Deák, 1980, 1990; Richter, 1994; Monahan and Genz, 1997; Somerville, 1997, 1998; Ni and Kedem, 1999). Similar derivations for MVT problems starting from either of equations (1.2) to (1.4) are given by Somerville (1997, 1998) and Genz and Bretz (2002).

All of the methods begin with a Cholesky decomposition. This yields a lower triangular $k \times k$ matrix $\mathbf{L} = (l_{ij})$ such that $l_{ii} > 0$ and $\mathbf{L}^{-1}\mathbf{\Sigma}\mathbf{L}^{-t} = \mathbf{I}_k$. Let $|\mathbf{A}|$ denote the determinant of a matrix $\mathbf{A}$. Then the linear transformation $\mathbf{x} = \mathbf{L}\mathbf{y}$, with its Jacobian given by

$$d\mathbf{x} = |\mathbf{L}|d\mathbf{y} = \sqrt{|\mathbf{\Sigma}|}d\mathbf{y},$$

reduces equation (1.1) to

A. Genz, *Computation of Multivariate Normal and t Probabilities*

DOI: 10.1007/978-3-642-01689-9_4, © Springer-Verlag Berlin Heidelberg 2009

$$\Phi_k(\mathbf{a} \le \mathbf{x} \le \mathbf{b}; \mathbf{\Sigma}) = \Phi_k(\mathbf{a} \le \mathbf{L}\mathbf{y} \le \mathbf{b}; \mathbf{I}_k) \tag{4.1}$$
$$= \int_{a_1 \le l_{11} y_1 \le b_1} \phi(y_1) \quad \ldots \quad \int_{a_k \le \mathbf{l}_k^t \mathbf{y} \le b_k} \phi(y_k) d\mathbf{y},$$

where $\mathbf{l}_j^t$ is the j-th row of $\mathbf{L}$. Note that other methods exist (for example, the principal component decomposition) which lead to similar standardizations. A second change of variables transforms the standardized vector $\mathbf{y}$ to a radius r and direction vector $\mathbf{z}$, $\mathbf{y} = r\mathbf{z}$ with $||\mathbf{z}|| = 1$, so that $\mathbf{y}^t\mathbf{y} = r^2$ for $r \ge 0$. This changes the integral (4.1) accordingly to

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \int_{||\mathbf{z}||=1} \frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \int_{r_l(\mathbf{z})}^{r_u(\mathbf{z})} r^{k-1} e^{-\frac{r^2}{2}} dr dU(\mathbf{z}), \tag{4.2}$$

where $U(.)$ is the joint cdf of the uniform distribution on the unit hypersphere $\{\mathbf{z} : ||\mathbf{z}|| = 1\}$. Letting

$$R(\mathbf{z}) = \{r : r \ge 0, \mathbf{a} \le r\mathbf{L}\mathbf{z} \le \mathbf{b}\},$$

the integration limits of the inner integral are defined as $r_l(\mathbf{z}) = \min\{r : r \in R(\mathbf{z})\}$ and $r_u(\mathbf{z}) = \max\{r : r \in R(\mathbf{z})\}$. If we let $\mathbf{v} = \mathbf{L}\mathbf{z}$, the limits for the r-variable integration can be given more explicitly by

$$r_l(\mathbf{z}) = \max\left\{0, \max_{v_i>0}\{a_i/v_i\}, \max_{v_i<0}\{b_i/v_i\}\right\},$$

$$r_u(\mathbf{z}) = \max\left\{0, \min\{\min_{v_i>0}\{b_i/v_i\}, \min_{v_i<0}\{a_i/v_i\}\}\right\}.$$

These limits are the distances from the origin to the two points where the vector with direction $\mathbf{z}$ intersects the boundary of the integration region.

For the MVT case, the expressions remain similar, with the exception of integrating a density for the $F_{k,\nu}$ distribution along the direction $\mathbf{z}$, instead of the χ_k distribution. We have

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu) = \int_{||\mathbf{z}||=1} F_{k,\nu}(r_U(\mathbf{z})) - F_{k,\nu}(r_l(\mathbf{z})) dU(\mathbf{z}), \tag{4.3}$$

with

$$F_{k,\nu}(r) = \frac{2\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{\nu}{2})\nu^{\frac{k}{2}}} \int_0^r \frac{t^{k-1}}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+k}{2}}} dt.$$

This can also be written in terms of a $Beta_{a,b}$ distribution using the relation $F_{k,\nu}(r) = Beta_{\frac{k}{2},\frac{\nu}{2}}(r^2/(\nu + r^2))$.

Extensions of these transformations were provided by Somerville (2001) for elliptical regions and by Lohr (1993) for general star shaped regions. Different

integration techniques have been applied to equations (4.2), (4.3), and related expressions: completely randomized methods, randomization of fixed sets of integration nodes, mixed methods using fixed quadrature on the radius and randomized spherical integration, and other methods. Some of these methods will be discussed in Section 4.2.

We illustrate the SR transformation described in this section with an example. Consider the three-dimensional example problem (1.6) where $\mathbf{a} = -\infty$, $\mathbf{b} = (1, 4, 2)^t$, $\nu = 5$ and

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \frac{3}{5} & \frac{1}{3} \\ \frac{3}{5} & 1 & \frac{11}{15} \\ \frac{1}{3} & \frac{11}{15} & 1 \end{bmatrix} = \mathbf{L}\mathbf{L}^t, \quad \text{with } \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}.$$

The integration region determined by $\mathbf{a} \le r\mathbf{L}\mathbf{z} \le \mathbf{b}$ can be explicitly written

$$-\infty \le rz_1 \le 1, \quad -\infty \le \frac{3rz_1}{5} + \frac{4rz_2}{5} \le 4, \quad -\infty \le \frac{rz_1}{3} + \frac{2rz_2}{3} + \frac{r2z_3}{3} \le 2,$$

and the SR form for T_{3ex} is

$$T_{3ex} = \int\limits_{||\mathbf{z}||=1} \frac{2\Gamma(4)}{\Gamma(\frac{3}{2})\Gamma(\frac{5}{2})\nu^{\frac{3}{2}}} \int\limits_0^{r_u(\mathbf{z})} \frac{r^2}{\left(1 + \frac{r^2}{5}\right)^4} dr dU(\mathbf{z}).$$

Note that $r_l(\mathbf{z}) = 0$ because the integration region contains the origin. If we let $v_1 = 1$, $v_2 = 3z_1/5 + 4z_2/5$, and $v_3 = z_1/3 + 2z_2/3 + 2z_3/3$, then $r_u(\mathbf{z}) = \max\{0, \min\limits_{v_i>0}\{b_i/v_i\}\}$.

4.1.2 Separation-of-Variables Methods

A second set of reparameterizations effects a *separation-of-variables* (SOV). The methods considered here involve the evaluation of each element x_i conditional on the realized values of the preceding terms $x_1, \ldots, x_{i-1}$. The innermost integral − after transformation − depends on all integration variables $x_1, \ldots, x_k$, while the outermost integral depends solely on x_1. In addition, these transformations standardize the original integral to one over the unit hypercube $[0, 1]^k$, irrespective of the initial parameter values $\mathbf{a}, \mathbf{b}$ and $\mathbf{\Sigma}$. Geweke (1991a) and Genz (1992) independently derived such transformations for the MVN problem. Geweke (1991a), Vijverberg (1996), Genz and Bretz (1999), and Genz and Bretz (2002) derived similar methods for MVT probabilities, starting from either of the equations (1.2) through (1.4). It is worth pointing out that these transformations, in principle, have already been described by Rosenblatt (1952). The MVT decomposition is also stated by Zellner (1971, pp. 383), though in a different context.

All methods start with a Cholesky decomposition $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$ (as given in equation (4.1) for the MVN integral) followed by substitutions which effect

the separation of variables. A $k-$fold application of the inverse normal or t function and subsequent standardization of the integration region to $[0,1]$ along each axis finishes the transformation.

We first provide some details for the MVN problem. After the Cholesky decomposition, equation (4.1) becomes

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int_{\tilde{a}_1}^{\tilde{b}_1} \phi(y_1) \int_{\tilde{a}_2(y_1)}^{\tilde{b}_2(y_1)} \phi(y_2) \cdots \int_{\tilde{a}_k(y_1,\ldots,y_{k-1})}^{\tilde{b}_k(y_1,\ldots,y_{k-1})} \phi(y_k) d\mathbf{y}, \tag{4.4}$$

with

$$\tilde{a}_i = \frac{a_i - \sum_{j=1}^{i-1} l_{i,j} y_j}{l_{i,i}} \quad \text{and} \quad \tilde{b}_i = \frac{b_i - \sum_{j=1}^{i-1} l_{i,j} y_j}{l_{i,i}}.$$

If we let $y_i = \Phi^{-1}(z_i)$, then

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int_{\Phi(\tilde{a}_1)}^{\Phi(\tilde{b}_1)} \int_{\Phi(\tilde{a}_2(\Phi^{-1}(z_1)))}^{\Phi(\tilde{b}_2(\Phi^{-1}(z_1)))} \cdots \int_{\Phi(\tilde{a}_k(\Phi^{-1}(z_1),\ldots,\Phi^{-1}(z_{k-1})))}^{\Phi(\tilde{b}_k(\Phi^{-1}(z_1),\ldots,\Phi^{-1}(z_{k-1})))} d\mathbf{z}.$$

Finally, we let $z_i = d_i + (e_i - d_i)w_i$, so $dz_i = (e_i - d_i)dw_i$, with

$$\begin{aligned} d_i(w_1,\ldots,w_{i-1}) &= \Phi\big(\tilde{a}_i(\Phi^{-1}(z_1(w_1)),\ldots,\Phi^{-1}(z_{i-1}(w_{i-1})))\big), \\ e_i(w_1,\ldots,w_{i-1}) &= \Phi\big(\tilde{b}_i(\Phi^{-1}(z_1(w_1)),\ldots,\Phi^{-1}(z_{i-1}(w_{i-1})))\big). \end{aligned}$$

Then

$$\begin{aligned} \Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = (e_1 - d_1) &\int_0^1 (e_2(w_1) - d_2(w_1)) \\ &\cdots \int_0^1 (e_k(w_1,\ldots,w_{k-1}) - d_k(w_1,\ldots,w_{k-1})) \int_0^1 d\mathbf{w}. \end{aligned} \tag{4.5}$$

The central MVT and non-central MVT distributions can be transformed in a similar manner. In the central MVT case, there is a choice between either equation (1.2) or (1.3). If equation (1.3) is used, then

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

If we define

$$\chi_\nu(s) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^t r^{\nu-1} e^{-\frac{r^2}{2}} dr,$$

and use the transformation $s = \chi_\nu^{-1}(t)$, then

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu) = \int_0^1 \Phi_k \left(\frac{\chi_\nu^{-1}(t)\mathbf{a}}{\sqrt{\nu}}, \frac{\chi_\nu^{-1}(t)\mathbf{b}}{\sqrt{\nu}}; \mathbf{\Sigma} \right) dt. \tag{4.6}$$

The transformations described previously for the MVN problem can then be applied, to produce a $(k+1)$-dimensional integral with $[0,1]$ for all of the integration limits. The non-central MVT problem can be transformed in a similar way by applying the MVN transformations to

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu, \boldsymbol{\delta}) = \int_0^1 \Phi_k \left(\frac{\chi_\nu^{-1}(t)\mathbf{a}}{\sqrt{\nu}} - \boldsymbol{\delta}, \frac{\chi_\nu^{-1}(t)\mathbf{b}}{\sqrt{\nu}} - \boldsymbol{\delta}; \mathbf{\Sigma} \right) dt.$$

If equation (1.2) is used for the central MVT definition, then after the Cholesky decomposition, we have

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^k}} \int_{\tilde{a}_1}^{\tilde{b}_1} \int_{\tilde{a}_2(y_1)}^{\tilde{b}_2(y_1)} \cdots \int_{\tilde{a}_k(\mathbf{y})}^{\tilde{b}_k(\mathbf{y})} \left(1 + \frac{\mathbf{y}^t\mathbf{y}}{\nu} \right)^{-\frac{\nu+k}{2}} d\mathbf{y}.$$

Now, we let

$$K_\nu^{(k)} = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{k}{2}}}, \tag{4.7}$$

and perform a partial separation of variables using the identity

$$\left(1 + \frac{\sum_{j=1}^k y_j^2}{\nu} \right) = \left(1 + \frac{y_1^2}{\nu} \right) \left(1 + \frac{y_2^2}{\nu + y_1^2} \right) \cdots \left(1 + \frac{y_m^2}{\nu + \sum_{j=1}^{k-1} y_j^2} \right).$$

Then we use the transformations

$$y_i = u_i \sqrt{\frac{\nu + \sum_{j=1}^{i-1} y_j^2}{\nu + i - 1}},$$

which can also be written as

$$y_i = u_i \sqrt{\frac{\prod_{j=1}^{i-1} (\nu + j - 1 + u_j^2)}{\nu + j}}, \tag{4.8}$$

combined with some more algebra (see Genz and Bretz (1999) for details), with the result

$$T_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma},\nu) = \int_{\tilde{a}_1}^{\tilde{b}_1} \frac{K_\nu^{(1)}}{(1+\frac{u_1^2}{\nu})^{\frac{1+\nu}{2}}} \int_{\tilde{a}_2(u_1)}^{\tilde{b}_2(u_1)} \frac{K_{\nu+1}^{(1)}}{(1+\frac{u_2^2}{\nu+1})^{\frac{2+\nu}{2}}} \cdots \tag{4.9}$$

$$\int_{\tilde{a}_k(\mathbf{y}(\mathbf{u}))}^{\tilde{b}_k(\mathbf{y}(\mathbf{u}))} \frac{K_{\nu+k-1}^{(1)}}{(1+\frac{u_k^2}{\nu+k-1})^{\frac{k+\nu}{2}}} d\mathbf{u}$$

$$= \int_{\tilde{a}_1}^{\tilde{b}_1} t(u_1;\nu) \int_{\tilde{a}_2(u_1)}^{\tilde{b}_2(u_1)} t(u_2;\nu+1) \cdots \int_{\tilde{a}_k(\mathbf{y}(\mathbf{u}))}^{\tilde{b}_k(\mathbf{y}(\mathbf{u}))} t(u_k;\nu+k-1) d\mathbf{u}.$$

If we let $u_i = T^{-1}(z_i;\nu+i-1)$, then

$$T_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma},\nu) = \int_{T(\tilde{a}_1;\nu)}^{T(\tilde{b}_1;\nu)} \int_{T(\tilde{a}_2(u_1(z_1));\nu+1)}^{T(\tilde{b}_2(u_1(z_1));\nu+1)} \cdots \int_{T(\tilde{a}_k(\mathbf{y}(\mathbf{u}(\mathbf{z})));\nu+k-1)}^{T(\tilde{b}_k(\mathbf{y}(\mathbf{u}(\mathbf{z})));\nu+k-1)} d\mathbf{z}.$$

A final transformation to $\mathbf{w}$ variables using

$$z_i = T(\tilde{a}_i(\mathbf{y}(\mathbf{u}(\mathbf{z})));\nu+k-1) + (T(\tilde{b}_i(\mathbf{y}(\mathbf{u}(\mathbf{z}))));\nu+k-1) - T(\tilde{a}_i(\mathbf{y}(\mathbf{u}(\mathbf{z}))));\nu+k-1)w_i$$

for $i = 1, 2, \ldots, k$, provides a k-dimensional integral with $[0,1]$ for all of the integration limits.

We consider the three-dimensional example problem (1.6) investigated in the previous section to illustrate the method described in this section. The SOV integration region is determined by $\mathbf{a} \le \mathbf{L}\mathbf{y} \le \mathbf{b}$, with

$$-\infty \le y_1 \le 1, \quad -\infty \le y_2 \le 5 - \frac{3y_1}{4}, \quad -\infty \le y_3 \le 3 - \frac{y_1}{2} - y_2.$$

We apply the transformation to the $\mathbf{u}$ variables using

$$y_1 = u_1, \quad y_2 = u_2\sqrt{\frac{5+u_1^2}{6}}, \quad y_3 = u_3\sqrt{\frac{5+u_1^2}{6}\frac{6+u_2^2}{7}},$$

so that

$$T_{3ex} = \frac{\Gamma(\frac{8}{2})}{\Gamma(\frac{5}{2})\sqrt{(5\pi)^3}} \int_{\mathbf{a}\le C\mathbf{y}\le\mathbf{b}} \left(1+\frac{\mathbf{y}^t\mathbf{y}}{5}\right)^{-\frac{8}{2}} d\mathbf{y}$$

$$= \int_{-\infty}^{1} \frac{K_5^{(1)}}{\left(1+\frac{u_1^2}{5}\right)^{\frac{6}{2}}} \int_{-\infty}^{\tilde{b}_2(u_1)} \frac{K_6^{(1)}}{\left(1+\frac{u_2^2}{6}\right)^{\frac{7}{2}}} \int_{-\infty}^{\tilde{b}_3(u_1,u_2)} \frac{K_7^{(1)}}{\left(1+\frac{u_1^2}{7}\right)^{\frac{8}{2}}} d\mathbf{u},$$

with

$$\tilde{b}_2(u_1) = \sqrt{\frac{6}{5+u_1^2}}\left(5 - \frac{3}{4}u_1\right),$$

and

$$\tilde{b}_3(u_1, u_2) = \sqrt{\frac{6}{5+u_1^2}\frac{7}{6+u_2^2}}\left(3 - \frac{1}{2}u_1 - u_2\sqrt{\frac{6}{5+u_1^2}}\right).$$

Then we use $e_1 = T(1;5), z_1 = e_1 w_1, u_1 = T^{-1}(z_1;5)$, $e_2 = T(\tilde{b}_2(u_1);6)$, $z_2 = e_2 w_2$, $u_2 = T^{-1}(z_2;6)$, and $e_3 = T(\tilde{b}_3(u_1,u_2);7)$, so that

$$T_{3ex} = e_1 \int_0^1 e_2(w_1) \int_0^1 e_3(w_1, w_2) dw_2 dw_1. \tag{4.10}$$

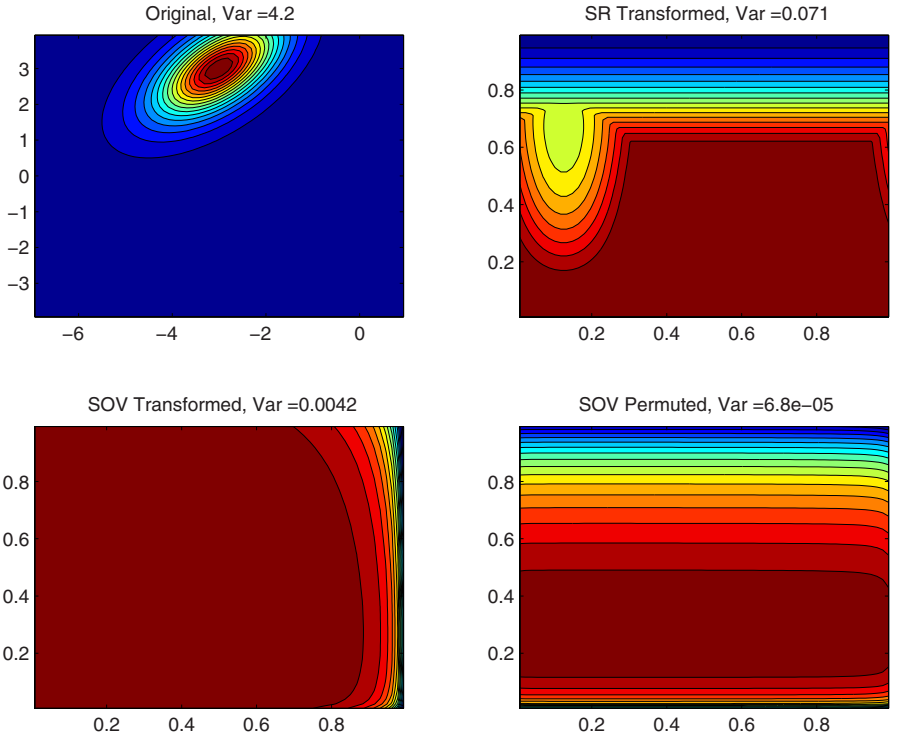

Fig. 4.1. Example graphs for T_{3ex}

In Figure 4.1 we provide four graphs to illustrate the different methods. The top left graph is a contour plot for the original density function in this example, with the last variable integrated out. The top right graph is a contour

graph for the SR integrand from equation (4.3), using the spherical surface parameterization described in Section 4.1.1; see equation (4.15) for the explicit expression. The bottom left graph is a contour plot for the SOV integrand $e_1 e_2(w_1) e_3(w_1, w_2)$ from equation (4.10). The bottom right graph also shows the SOV integrand for the example in this section but variables 2 and 3 are interchanged (see next section). These graphs show the integrands that would be sampled by Monte Carlo methods applied to the transformed problems. The `Var` quantities given for each graph are variances for the data used to plot each graph. These graphical representations suggest that such transformed integrands are indeed better suited for multidimensional integration than the original problems (Genz and Bretz, 2002). It is also possible to truncate the unbounded domain to simplify the numerical integration problem, but it is difficult to provide good *a priori* truncated integration limits. For a discussion of this issue with examples involving trivariate t orthant probabilities we refer to Fang and Wang (1994, p. 78).

4.1.3 Variable Reordering

Many of the resulting numerical evaluation problems can be transformed into easier computational problems if the variables are reordered. There are $k!$ possible reorderings of the variables for a MVN or MVT problem. These reorderings do not change the MVN or MVT value as long as the integration limits and corresponding rows and columns of the covariance matrix $\mathbf{\Sigma}$ are also permuted. Schervish (1984) originally proposed sorting the variables so that the variables with shortest integration interval widths are the outer integration variables. This approach is expected to reduce the overall variation of the integrand and consequently to result in an easier numerical integration problem. Gibson et al (1994) suggested an improved prioritization of the variables. They proposed sorting the variables so that the innermost integrals have the largest expected values. The first (outermost) integration variable is chosen by selecting a variable

$$i = \arg \min_{1 \le i \le k} \left\{ \Phi\left(\frac{b_i}{\sqrt{\sigma_{i,i}}} \right) - \Phi\left(\frac{a_i}{\sqrt{\sigma_{i,i}}} \right) \right\}.$$

The integration limits as well as the rows and columns of $\mathbf{\Sigma}$ for variables 1 and i are interchanged. Then the first column of the Cholesky decomposition $\mathbf{L}$ of $\mathbf{\Sigma}$ is computed using $l_{1,1} = \sqrt{\sigma_{1,1}}$ and $l_{i,1} = \sigma_{i,1}/l_{1,1}$, for $i = 2, \ldots, m$, and we set

$$y_1 = \frac{1}{\Phi(b_1) - \Phi(a_1)} \int_{a_1}^{b_1} s\phi(s) ds.$$

Given this (expected) value for y_1, the second integration variable is chosen by selecting a variable

$$i = \arg\min_{2\le i\le k} \left\{ \Phi\left(\frac{b_i - l_{i,1}y_1}{\sqrt{\sigma_{i,i} - l_{i,1}^2}} \right) - \Phi\left(\frac{a_i - l_{i,1}y_1}{\sqrt{\sigma_{i,i} - l_{i,1}^2}} \right) \right\}.$$

The integration limits, rows and columns of $\mathbf{\Sigma}$, and rows of $\mathbf{L}$ for variables 2 and i are interchanged. Then the second column of $\mathbf{L}$ is computed using

$$l_{2,2} = \sqrt{\sigma_{2,2} - l_{2,1}^2} \quad \text{and} \quad l_{i,2} = \frac{\sigma_{i,2} - l_{2,1}l_{i,1}}{l_{2,2}},$$

for $i = 3, \ldots, k$, and we compute the expected value for y_2 using

$$y_2 = \frac{1}{\Phi(\tilde{b}_2) - \Phi(\tilde{a}_2)} \int_{\tilde{a}_2}^{\tilde{b}_2} s\phi(s)ds.$$

At stage j, given the expected values for $y_1, y_2, \ldots, y_{j-1}$, the j-th integration variable is chosen by selecting a variable

$$i = \arg\min_{j\le i\le m} \left\{ \Phi\left(\frac{b_i - \sum_{m=1}^{j-1} l_{i,m}y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}} \right) - \Phi\left(\frac{a_i - \sum_{m=1}^{j-1} l_{i,m}y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}} \right) \right\}.$$

The integration limits, rows and columns of $\mathbf{\Sigma}$, and rows of $\mathbf{L}$ for variables j and i are interchanged. Then the j-th column of $\mathbf{L}$ is computed using

$$l_{j,j} = \sqrt{\sigma_{j,j} - \sum_{m=1}^{j-1} l_{j,m}^2} \quad \text{and} \quad l_{i,j} = \frac{1}{l_{j,j}} \left(\sigma_{i,j} - \sum_{m=1}^{j-1} l_{j,m}l_{i,m} \right),$$

for $i = j+1, \ldots, k$, and we let

$$y_j = \frac{1}{\Phi(\tilde{b}_j) - \Phi(\tilde{a}_j)} \int_{\tilde{a}_j}^{\tilde{b}_j} s\phi(s)ds.$$

The complete $k-1$ stage process has overall cost $O(k^3)$, which is usually insignificant compared to the rest of the computation cost for the methods discussed here, and is therefore a relatively cheap preconditioning step that can be used with the algorithms. The y_i's that are produced by the algorithm can also be used to approximate $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})$ following Mendell and Elston (1974),

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) \approx \int_{\tilde{a}_1}^{\tilde{b}_1} \phi(z_1) \int_{\tilde{a}_2(y_1)}^{\tilde{b}_2(y_1)} \phi(z_2) \cdots \int_{\tilde{a}_k(y_1,\ldots,y_{k-1})}^{\tilde{b}_k(y_1,\ldots,y_{k-1})} \phi(z_k)d\mathbf{z}, \tag{4.11}$$

which is just a product of one-dimensional normal probabilities (see also Section 3.3).

The variance for $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})$ in the SOV form given by equation (4.5) is

$$\int_{\mathbf{0}}^{\mathbf{1}} \prod_{i=1}^{k} (e_i(w_1, \ldots, w_{i-1}) - d_k(w_1, \ldots, w_{i-1}))^2 d\mathbf{w} - \Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})^2.$$

The variable reordering algorithm above is a "greedy" algorithm for minimizing the Φ_k variance by successively selecting the variable orderings to minimize the $(e_i(w_1, \ldots, w_{i-1}) - d_k(w_1, \ldots, w_{i-1}))^2$ terms in the variance product for $i = 1, 2, \ldots, k$.

A related consequence of the algorithm is that the w_i variables with smaller indices have smaller domains than the variables with larger indices. But the variations in a variable w_i cause variations in all w_j with $j > i$, so a reduction in the domain for w_i should reduce the variation in the variables w_j with $j > i$, and the overall variance for $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})$. These considerations lead to a slightly more sophisticated algorithm which minimizes the variance for the variables selected at each stage. We define the variance for a normal variable x over the truncated interval $[a, b]$ by

$$v(a, b) = \frac{\int_a^b x^2 e^{-x^2/2} dx}{\Phi(b) - \Phi(a)} - \left(\frac{\int_a^b x e^{-x^2/2} dx}{\Phi(b) - \Phi(a)} \right)^2,$$

which simplifies to

$$v(a, b) = 1 + \frac{a\phi(a) - b\phi(b)}{\Phi(b) - \Phi(a)} - \left(\frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)} \right)^2.$$

At stage j, with this revised strategy, the j-th integration variable is chosen to minimize v, by selecting variable

$$i = \arg \min_{j \le i \le m} \left\{ v \left(\frac{a_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}}, \frac{b_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}}, \right) \right\}.$$

Examples show that the variance of the estimated integral can vary by an order of magnitude or more for different orderings. If we consider the trivariate normal problem (1.5), then the original variable ordering produces an SOV integrand (as given by equation (4.5)) with approximate variance 0.002, compared to an approximate variance of 0.00006 for the SOV integrand which is produced by either of the reordering strategies discussed in this section (and which resulted in a $(1, 3, 2)$ variable reordering). If we consider the five-dimensional problem (1.8), then the original variable ordering leads to a SOV integrand with approximate variance 0.07, compared to 0.0001 for the SOV integrand produced by either of the reordering strategies discussed in this section and which resulted in a $(5, 4, 3, 2, 1)$ variable reordering.

Corresponding variable reordering strategies can also be used for MVT SOV integrands. If definition (1.2) is used, then strategies similar to the ones discussed is this section can be developed if the Φ's are replaced by appropriate t_ν's; see Genz and Bretz (2002) for further details. For the trivariate t problem (1.6), we achieve a variance reduction from 0.042 to 0.000068, see also Figure 4.1. If definition (1.3), is used, then a simple strategy, suggested by Genz and Bretz (2002), uses $\sqrt{\nu}$ for an approximation to the expected value of s for the outer χ_k integral, and then (after cancelation of the $\sqrt{\nu}$ terms) applies the Gibson et al (1994) strategy to $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})$.

Tests by Genz and Bretz (2002) show that a significant improvement in efficiency is possible when the Gibson et al (1994) strategy is used. These variable reordering strategies will be discussed further when test results are reviewed in Chapter 5. These ideas are closely related to the problem of determining the *effective dimension* of multidimensional integrals, where recent research has focused on finding sufficient conditions for the *tractability* of certain classes of integrals. Results relating to MVN integrals are given by Woźniakowski (2000), Curbera (2000) and Sloan (2002).

4.1.4 Tridiagonal Decomposition Methods

In Section 2.3.2 we described how the cost for MVN and MVT computations can be reduced when the covariance matrix or its inverse is tridiagonal. In this section we describe methods for decomposing a general MVN cdf problem into a set of tridiagonal MVN cdf problems. Similar methods can also be used for MVT cdf problems. Related methods were originally described by Schläfli (1858), but more recently by Miwa et al (2003) and Craig (2008).

We consider the method by Miwa et al (2003) in more detail, where the decomposition is based on the covariance matrix. Here, we provide an alternate description of this in terms of a decomposition of the inequalities for the integration region. After the Cholesky decomposition $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$, the general MVN cdf is

$$\Phi_k(-\infty \le \mathbf{x} \le \mathbf{b}; \mathbf{\Sigma}) = \Phi_k(-\infty \le \mathbf{L}\mathbf{y} \le \mathbf{b}; \mathbf{I}_k).$$

The basic idea is to successively transform $\Phi_k(-\infty \le \mathbf{L}\mathbf{y} \le \mathbf{b}; \mathbf{I}_k)$ into a set of problems with $\mathbf{L}$, each of which has a lower bidiagonal matrix for $\mathbf{L}$. Focusing on the set of integration region inequalities $-\infty \le \mathbf{L}\mathbf{y} \le \mathbf{b}$, we begin by replacing $\mathbf{L}$ with $\mathbf{L}\mathbf{Q}$, where $\mathbf{Q}$ is an orthogonal (reflector) matrix (Golub and Van Loan, 1996, p. 209) chosen such that $\mathbf{L}\mathbf{Q}$ has zeros in the first $k-2$ positions in the last row, that is

$$\mathbf{L}\mathbf{Q} = \begin{bmatrix} \star & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ ? & ? & \cdots & ? & \star & 0 \\ ? & ? & \cdots & \cdots & ? & \star \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} ? & ? & \cdots & \cdots & ? & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & ? & \cdots & \cdots & ? & 0 \\ 0 & 0 & \cdots & 0 & ? & \star \end{bmatrix},$$

where $\star$'s are nonzero entries and ?'s are either zero or nonzero entries. Now, $-\infty < \mathbf{Ly} \le \mathbf{b}$ becomes $-\infty < \mathbf{LQQ}^t\mathbf{y} = \mathbf{LQz} \le \mathbf{b}$, with $\mathbf{z} = \mathbf{Q}^t\mathbf{y}$, and

$$\Phi_k(-\infty \le \mathbf{x} \le \mathbf{b}; \mathbf{\Sigma}) = \Phi_k(-\infty \le \mathbf{LQz} \le \mathbf{b}; \mathbf{I}_k).$$

because $e^{-\mathbf{z}^t\mathbf{z}/2} = e^{-\mathbf{y}^t\mathbf{QQ}^t\mathbf{y}/2} = e^{-\mathbf{y}^t\mathbf{y}/2}$. Then, Fourier-Motzkin elimination (Schechter, 1998) can be used to replace $-\infty < \mathbf{LQz} \le \mathbf{b}$ by $k-1$ inequalities of the form $-\infty < \mathbf{M}_j^{(1)}\mathbf{z} \le \mathbf{b}_j$, for $j = 1, 2, \ldots, k-1$, where

$$\mathbf{M}_j^{(1)} = \begin{bmatrix} ? & ? & ? & ? & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & ? & \cdots & ? & 0 & 0 \\ ? & ? & \cdots & ? & \star & 0 \\ 0 & 0 & \cdots & 0 & ? & \star \end{bmatrix}.$$

Now, with signs $s_j = \pm 1$, depending on the $\mathbf{M}_j^{(1)}$ diagonal entries,

$$\Phi_k(-\infty \le \mathbf{x} \le \mathbf{b}; \mathbf{\Sigma}) = \sum_{j=1}^{k-1} s_j \int\limits_{-\infty < \mathbf{M}_j^{(1)}\mathbf{z} \le \mathbf{b}_j} \frac{e^{-\frac{\mathbf{z}^t\mathbf{z}}{2}}}{\sqrt{(2\pi)^m}} d\mathbf{z}.$$

The next stage in this process chooses orthogonal $\mathbf{Q}$'s so that the $\mathbf{M}_j^{(1)}\mathbf{Q}$ matrices have zeros in the first $k-3$ positions in row $k-1$, followed by Fourier-Motzkin elimination to create zeros in the first $k-3$ positions in the columns numbered $k-3$. Now, there are $(k-1)(k-2)$ inequalities of the form $-\infty < \mathbf{M}_j^{(2)}\mathbf{z} \le \mathbf{b}_j$, with

$$\mathbf{M}_j^{(2)} = \begin{bmatrix} ? & ? & ? & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & \cdots & ? & 0 & 0 & 0 \\ ? & \cdots & ? & \star & 0 & 0 \\ 0 & \cdots & 0 & ? & \star & 0 \\ 0 & 0 & \cdots & 0 & ? & \star \end{bmatrix}.$$

After $k-1$ stages, there are $(k-1)!$ inequality sets with the form $-\infty < \mathbf{M}_j^{(k-1)}\mathbf{z} \le \mathbf{b}_j$, where (after some diagonal entry scaling)

$$\mathbf{M}_j^{(k-1)} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ c_1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{k-2} & 1 & 0 \\ 0 & 0 & \cdots & 0 & c_{k-1} & 1 \end{bmatrix},$$

so that

$$\Phi_k(-\infty \le \mathbf{x} \le \mathbf{b}; \mathbf{\Sigma}) = \sum_{i=j}^{(k-1)!} s_j \int\limits_{-\infty < \mathbf{M}_j^{(k-1)}\mathbf{z} \le \mathbf{b}_j} \frac{e^{-\frac{\mathbf{z}^t\mathbf{z}}{2}}}{\sqrt{(2\pi)^m}} d\mathbf{z}$$

$$= \sum_{j=1}^{(k-1)!} s_j \Phi_k(-\infty \le \mathbf{x} \le \mathbf{b}_j; \mathbf{T}_j), \tag{4.12}$$

with $\mathbf{T}_j = \mathbf{M}_j^{(j)}(\mathbf{M}_j^{(k-1)})^t$, and all of the $\mathbf{T}_j$ matrices are tridiagonal. All of the $\Phi_k(-\infty \le \mathbf{T}_j\mathbf{z} \le \mathbf{b}_j)$ integrals can be efficiently computed using the recursive algorithm described in Section 2.3.2.

The result of this is an $O((k-1)!)$ algorithm for MVN cdf probabilities, which limits the use of this algorithm for large values of k. However, for some practical applications, there can be fewer than $(k-1)!$ tridiagonal subproblems, because of extra zero entries in the reduction matrices (Miwa et al, 2003; Craig, 2008). An alternate algorithm with a reduction to at most $(k-1)!$ subproblems, which all have tridiagonal inverse covariance matrices, is also described by Craig (2008). These algorithms have the potential to efficiently compute low to moderate dimensional MVN (and MVT) cdf probabilities to much higher accuracy than the simulation based SOV algorithms for the full covariance matrix problems, described in the next section. Unfortunately, the only general reduction to tridiagonal problems for MVN and MVT probabilities when the lower integration limits $\mathbf{a}$ are bounded first requires a decomposition into 2^k cdf problems, so the overall time complexity increases to $O(2^k(k-1)!)$.

4.2 Integration Methods

4.2.1 Monte Carlo Methods

The (simple) *Monte Carlo* (MC) method is a standard device for evaluating multidimensional integrals. Deák (1990) and Hajivassiliou et al (1996) provide good reviews of MC integration methods while emphasizing the MVN distribution. For general discussions on MC methods we refer to Hammersley and Handscomb (1965) and Ripley (1987).

Discrete simulators (acceptance-rejection methods, crude frequency simulators, etc.) were some of the first MC methods used. These simulators use expressions such as

$$\frac{1}{N}\sum_{i=1}^{N} I_{\mathbf{a} \le \mathbf{x}_i \le \mathbf{b}},$$

where I is an indicator function, N denotes the number of samples, and the $\mathbf{x}_i$ are sampled from some given distribution. These methods are known to

behave poorly in comparison to other methods and we refer to them only for the sake of completeness. Crude MC methods are based on

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)$$

for a given integrand f and where $\mathbf{x}_i$ are sampled from the domain. In the following we discuss several major variance reduction techniques, which have been shown to perform well for the present integration problems. To make the notation concrete, we start by discussing the MVN case.

A standard method of variance reduction is the use of *antithetic variates*. Their implementation has been recommended by Deák (1980), Vijverberg (1997) and others. Genz and Bretz (2002), for example, suggested replacing $f(\mathbf{x}_i)$ by $(f(\mathbf{x}_i)+f(\mathbf{1}-\mathbf{x}_i))/2$ for the separation-of-variables reparameterizations leading to integrals over the unit hypercube.

For *importance sampling* techniques, let $h_k(\mathbf{x}) = \prod_{j=1}^{k} h(x_j)$ denote the continuous density of k independent and identically distributed random variables, which is decomposable into a product of univariate densities h. Since according to equation (4.1)

$$\Phi_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma}) = \int\limits_{\mathbf{a}\le\mathbf{L}\mathbf{y}\le\mathbf{b}} \phi_k(\mathbf{y};\mathbf{I}_k)d\mathbf{y} = \int\limits_{\mathbf{a}\le\mathbf{L}\mathbf{y}\le\mathbf{b}} \frac{\phi_k(\mathbf{y};\mathbf{I}_k)}{h_k(\mathbf{y})} h_k(\mathbf{y})d\mathbf{y},$$

Φ_k can be estimated by

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\phi_k(\mathbf{y}_i;\mathbf{I}_k)}{h_k(\mathbf{y}_i)} = \sum_{i=1}^{N}\prod_{j=1}^{k} \frac{\phi(y_{ij})}{h(y_{ij})}, \tag{4.13}$$

where $\mathbf{y}_1,\ldots,\mathbf{y}_N$ is a sample from the importance density h_k. The basic idea of importance sampling is to choose h_k so that the generated random numbers are concentrated in the region where $\phi_{k,\mathbf{I}_k}$ is large. If h_k is chosen to be the truncated normal density $\phi(y_{ij})/\Phi(a_j \le \mathbf{l}_j^t\mathbf{y}_i \le b_j)$, expression (4.13) simplifies to

$$\frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{k} \Phi(a_j \le \mathbf{l}_j^t\mathbf{y}_i \le b_j).$$

This importance sampler is known as *GHK simulator* (Geweke, 1991a; Hajivassiliou, 1993a; Keane, 1993) and was independently derived by Genz (1992) using SOV techniques (see Section 4.1.2). If simple Monte Carlo sampling is applied to the integral (4.5), then the resulting computations are equivalent to GHK importance sampling. Vijverberg (1997) investigated the exponential density, the truncated logit, t and transformed $Beta_{2,2}$ densities as alternative choices for h_k in equation (4.13). Vijverberg (2000) extended above importance samplers to integrals over wedge-shaped regions.

A third variance reduction technique is to use *control variates*. This requires a function h_k to be closely related to $\phi_{k,\mathbf{R}}$, where the integral is known or easy to compute. Then, the first integral of the decomposition

$$\int\limits_{-\infty \le \mathbf{x} \le \mathbf{b}} \phi_{k,\mathbf{R}}(\mathbf{x}) - h_k(\mathbf{x}) d\mathbf{x} \quad + \int\limits_{-\infty \le \mathbf{x} \le \mathbf{b}} h_k(\mathbf{x}) d\mathbf{x}$$

can be estimated using standard MC techniques. Stern (1992) proposed specifying h_k as $\phi_{k,\mathbf{I}}$ and estimating the remaining integral by frequency simulation. A different approach, which combines control variate techniques with importance sampling, has been proposed by Breslaw (1994). For convenience in terminology, let $s = k(k-1)/2$ and consider any correlation matrix $\mathbf{R}$ as defining a $s \times 1$ vector $\boldsymbol{\rho}(\mathbf{R})$ with coordinates ρ_{ij}. For a given starting point $\mathbf{R}'$, the differential element $d\Phi_k$ is then integrated along the line $\mathbf{R}'\mathbf{R}$ to obtain

$$\Phi_k(-\infty, \mathbf{b}; \mathbf{R}) = \Phi_k(-\infty, \mathbf{b}; \mathbf{R}') + \sum_{i<j} \int\limits_{\rho'_{ij}}^{\rho_{ij}} \frac{\partial \Phi_k}{\partial \kappa_{ij}}(\mathbf{K}) d\kappa_{ij},$$

see also Plackett (1954). Breslaw (1994) proposed choosing $\mathbf{R}'$ so that it has the product correlation structure (2.16). The elements are chosen to lie "close" to $\mathbf{R}$ by minimizing $\boldsymbol{\rho}(\mathbf{R} - \mathbf{R}')^t \boldsymbol{\rho}(\mathbf{R} - \mathbf{R}')$. The author then showed how the gradient $\frac{\partial \Phi_k}{\partial \kappa_{ij}}(\mathbf{K})$ can be evaluated efficiently using the GHK simulator with the help of the dimension reduction from Plackett (1954); recall equation (2.6) for the trivariate expression. Numerical comparisons of Breslaw (1994) and Bommert (1999) suggest that this *line integral simulator* is more accurate than the GHK simulator for a variety of parameter combinations. Szántai (2001) has studied the use of control variate simulators based on some of the probability bounds described in Section 3.1.

A different set of simulators is based on the SR transformation (4.2). The generation of points that are uniformly random on the unit hypersphere is extensively discussed by Fang and Wang (1994). One useful transformation described by Fang and Wang from the unit $k-1$ dimensional hypercube to the surface of the unit $k-$ sphere U_k is defined as follows. If $\mathbf{w} \in [0,1]^{k-1}$ and $\mathbf{z} \in U_k$, then the transformation is given by

$$z_{k-2i+2}(\mathbf{w}) = \sin(2\pi w_{k-2i+1}) \sqrt{1 - w_{k-2i}^{\frac{2}{k-2i}}} \prod_{j=1}^{i-1} w_{k-2j}^{\frac{1}{k-2j}}$$

$$z_{k-2i+1}(\mathbf{w}) = \cos(2\pi w_{k-2i+1}) \sqrt{1 - w_{k-2i}^{\frac{2}{k-2i}}} \prod_{j=1}^{i-1} w_{k-2j}^{\frac{1}{k-2j}}$$

for $i = 1, 2, \ldots l$, where $l = \lfloor \frac{k}{2} \rfloor - 1$, and ending with

$$z_2(\mathbf{w}) = \sin(2\pi w_1) \prod_{j=1}^{l} w_{k-2j}^{\frac{1}{k-2j}}$$

and

$$z_1(\mathbf{w}) = \cos(2\pi w_1) \prod_{k=1}^{l} w_{k-2j}^{\frac{1}{k-2j}},$$

when k is even, or ending with

$$z_3(\mathbf{w}) = (2w_1 - 1) \prod_{j=1}^{l} w_{k-2j}^{\frac{1}{k-2j}},$$

$$z_2(\mathbf{w}) = 2\sin(2\pi w_2)\sqrt{w_1(1-w_1)} \prod_{j=1}^{l} w_{k-2j}^{\frac{1}{k-2j}}$$

and

$$z_1(\mathbf{w}) = 2\cos(2\pi w_2)\sqrt{w_1(1-w_1)} \prod_{j=1}^{l} w_{k-2j}^{\frac{1}{k-2j}},$$

when k is odd. This transformation has a constant Jacobian, so integral (4.2) becomes

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \int_0^1 \int_0^1 \cdots \int_0^1 \chi_k\left(r_u(\mathbf{z}(\mathbf{w}))\right) - \chi_k\left(r_l(\mathbf{z}(\mathbf{w}))\right) d\mathbf{w}. \tag{4.14}$$

An MC algorithm for the MVN problem can use uniform points from $[0,1]^{k-1}$.

There are a variety of further simulators or variations on the methods described above. Many of them were investigated by Hajivassiliou et al (1996) and were found to perform poorly in comparison to the methods above. Further references include Deák (1986, 2001), Gassmann (1988), McFadden (1989), Börsch-Supan and Hajivassiliou (1993), Ambartzumian et al (1998), and Genz and Bretz (2002).

Most MC methods are directly generalized to the MVT case. Geweke (1991a) and Vijverberg (1996) independently examined using importance sampling techniques. Both authors reformulated the MVT pdf as a product of univariate t pdf's and applied the same set of candidate importance densities as described above for the MVN case. Westfall (1997) proposed a hybrid method of combing a control variate method with crude MC estimates by means of generalized least squares. Applications of MC techniques for SR transformations tailored to the MVT case are given by Genz and Bretz (2002). In this case the integral for (4.3) becomes

$$T_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}, \nu) = \int_0^1 \int_0^1 \cdots \int_0^1 F_{k,\nu}\left(r_u(\mathbf{z}(\mathbf{w}))\right) - F_{k,\nu}\left(r_l(\mathbf{z}(\mathbf{w}))\right) d\mathbf{w}. \tag{4.15}$$

Formulas similar to this based on other sphere surface parameterizations and other formulas for the MVT case are given by Somerville (1997, 1998).

4.2.2 Quasi-Monte Carlo Methods

As a consequence of the central limit theorem, crude MC integration yields a probabilistic error bound in $O(\sigma N^{-1/2})$, where $\sigma = \sigma(f)$ is the square root of the variance of the integrand f. Thus, halving the integration error requires quadrupling the number of sample points. One reason for this relatively low accuracy is that the sample points are independent. Some parts of the integration region typically contain clusters of points, while other parts remain almost empty. Error reduction of MC methods can be performed in two ways. Variance reduction methods, such as those discussed in the previous section, transform the integrand so that the constant σ is reduced. A second approach − to be discussed in this section − is to replace the independent random variables by alternative sequences, which improve the $-1/2$ exponent.

Quasi-Monte Carlo (QMC) methods (also sometimes known as *number theoretic methods*) use carefully selected, deterministic sequences of sample points. QMC methods are often justified by the Koksma-Hlawka inequality bound (Hlawka, 1961; Niederreiter, 1992) for the integration error

$$E_N(f) = \left| \int\limits_{[0,1]^k} f(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \right| \leq D(S_N) V(f).$$

Here, $V(f) < \infty$ denotes the $k-$dimensional variation of f in the sense of Hardy and Krause (Zaremba, 1968; Niederreiter, 1992) and

$$D(S_N) = \sup_{\mathbf{v} \in [0,1]^k} \left| \frac{\#\{i : \mathbf{x}_i \leq \mathbf{v}, \;\; i = 1, \ldots, N\}}{N} - v_1 v_2 \ldots v_k \right|$$

measures the *discrepancy* of any sequence $S_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ from $U[0,1]^k$. The Koksma-Hlawka inequality ensures that the integration error is practically bounded by $D(S_N)$. If the absolute difference between the ratio of the number of points lying in the rectangle $[\mathbf{0}, \mathbf{v}]$ to the total number N of points and the volume $v_1 v_2 \ldots v_k$ of the rectangle $[\mathbf{0}, \mathbf{v}]$ is small, then the set of points S_N is considered uniformly scattered. *Low-discrepancy sets* are typically of order $O(N^{-1+\epsilon})$, where N^ϵ stands for $\log N$ to some power. Thus, QMC is asymptotically better than MC. However, because the Koksma-Hlawka inequality is a gross worst-case bound, practical implementations of QMC integration methods typically behave much better than the inequality suggests (Caflisch et al, 1997; Sloan and Woźniakowski, 1998; Tezuka, 2002).

To see the effect, compare the two graphs in Figure 4.2. They depict the sets of 313 integration points within the unit hypercube $[0,1] \times [0,1]$ of a two-dimensional problem. The figure suggests that the MC sampling points, though obtained pseudo randomly, are not well equidistributed throughout the integration region. Note the gap $[0.2, 0.3] \times [0.5, 0.8]$, which contains almost no integration nodes. Other parts of $[0,1]^2$ have clusters of nodes. By contrast,

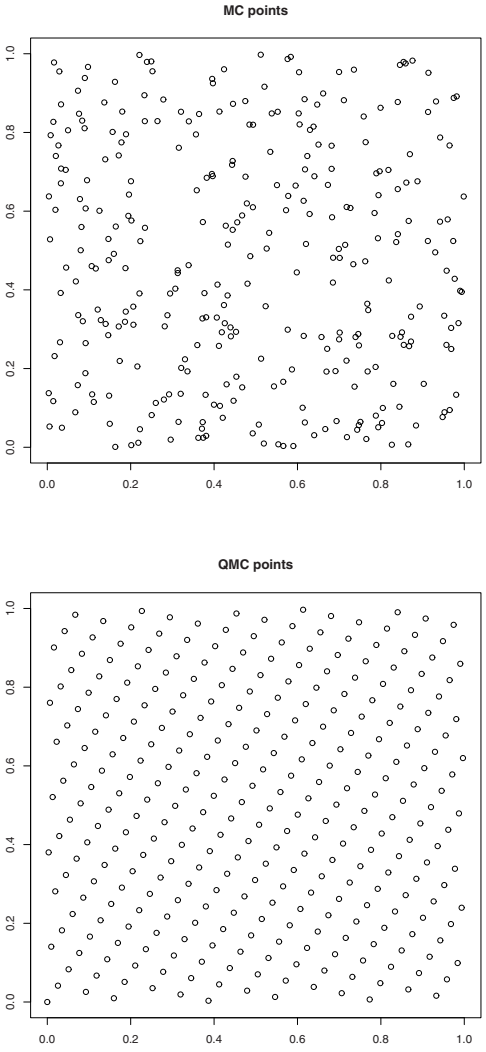

Fig. 4.2. Comparison of MC and QMC integration point sets.

the set of QMC points is much more regular and seems to fill the integration region more evenly.

There is much literature available on QMC integration methods. We refer the reader to Niederreiter (1992), Caflisch (1998) and Owen (2000) for general reading and further references. The latest developments in QMC methods are published in the biennial conference proceedings Niederreiter et al (1998),

Niederreiter and Spanier (2000), Fang et al (2002), Niederreiter and Talay (2006), and Keller et al (2008). In the following we focus on QMC methods for MVN and MVT integration problems.

Three important classes of low-discrepancy sets are *Kronecker sequences* (also called Weyl sequences), *lattices* and *digital nets*. Kronecker sequences (Drmota and Tichy, 1997, p. 67) are point sets

$$K_N = \{i\mathbf{v} \mod 1,\ i = 1, \ldots, N\},$$

where the components of $\mathbf{v}$ are irrational and linearly independent over the rational numbers. Simple choices for $\mathbf{v}$ are the Richtmyer (1951) vector, which has $v_i = \sqrt{p_i}$, with $p_i = i^{th}$ prime, the Niederreiter (1972) vector, which has $v_i = 2^{i/(k+1)}$, and the Zinterhof (1969) vector, which has $v_i = e^{i/k}$. These low-discrepancy sequences are easy to use in implementations, but often have larger discrepancies than good lattices and digital nets.

Lattice rules are particularly accurate for periodic functions over the unit hypercube (Niederreiter, 1992; Sloan and Joe, 1994; Wang and Hickernell, 2002). A rank-1 lattice is a point set in the form

$$L_N = \{i\mathbf{v}/N \mod 1,\ i = 0, \ldots, N-1\},$$

where $\mathbf{v}$ is a $k-$dimensional integer generating vector that depends on N. A common choice for $\mathbf{v}$ is the *Korobov form*

$$\mathbf{v}(h) = (1, h, h^2 \mod N, \ldots, h^{k-1} \mod N)$$

for a given h with $1 \leq h \leq \lfloor \frac{N}{2} \rfloor$ and prime N (Korobov, 1960). In order to achieve $O(N^{-1+\epsilon})$ integration errors, a "good" lattice parameter h must be determined. Simple methods for choosing a good h require an $O(N^2)$ cost minimization, which is usually too expensive for run-time computations, compared to the lattice rule application cost. Consequently, Korobov lattice rule implementations often store good h values for selected prime N sequences for a range of k values.

More recent work on the construction of good lattice rules has led to considering component-by-component (CBC) algorithms for determining good lattice parameters in the more general $\mathbf{v}$ form, and there are now $O(N \log(N))$ algorithms available for this purpose (Dick and Kuo, 2004; Nuyens and Cools, 2006a,b). These algorithms allow for the feasible run-time construction of good lattice parameters for QMC MVN and MVT computations.

An inherent problem with any standard deterministic QMC integration is that no simple error bounds are available. Randomized lattice rules have therefore been suggested, which give unbiased integral estimates along with standard MC error estimates. These lattice rules shift the whole lattice L_N by a $k-$dimensional shift $\boldsymbol{\Delta} \sim U[0,1]^k$, so that

$$L_N^{\text{shift}} = \{\mathbf{z} + \boldsymbol{\Delta} \mod 1 : \mathbf{z} \in L_N\}$$

(Cranley and Patterson, 1976; Joe, 1990). Averages of these randomly shifted lattice rule integral estimates are "randomized" QMC integral estimates. Beckers and Haegemans (1992) first studied randomized lattice rules for MVN computations. Genz and Bretz (2002) considered the use of

$$I_M = \frac{1}{M}\sum_{i=1}^{M} I_{N,i},$$

with

$$I_{N,i} = \frac{1}{2N}\sum_{j=1}^{N}\big(f(|2\{\mathbf{z}_j + \boldsymbol{\Delta}_i\} - \mathbf{1}|) + f(\mathbf{1} - |2\{\mathbf{z}_j + \boldsymbol{\Delta}_i\} - \mathbf{1}|)\big) \qquad (4.16)$$

to approximate both MVN and MVT probabilities. In this expression, $\{\cdot\}$ denotes the remainder mod 1. The number of shifts, M, is usually small, but large enough so that the standard error

$$\hat{\sigma}_M = \left(\frac{1}{M(M-1)}\sum_{i=1}^{M}(I_{N,i} - I_M)^2\right)^{\frac{1}{2}}$$

provides an error estimate for I_M. In practice, $\lambda\hat{\sigma}_M$ provides a robust error estimate when $\lambda = 3$ and $M = 12$, based on approximate 99% confidence intervals. The "baker" transformation $|2\mathbf{x} - \mathbf{1}|$ in (4.16) periodizes the integrand. Hickernell (2002) showed this type of periodization combined with a randomization step provides $O(N^{-2+\epsilon})$ asymptotic errors for sufficiently smooth functions. Other periodizing transformations (Sloan and Joe, 1994) introduce non-constant Jacobian factors which can significantly increase the difficulty of the integration problem.

Previous versions of (4.16) have been used by Genz (1993) for MVN and Genz and Bretz (1999) for MVT computations, where test results showed that these methods were very efficient, compared to other methods. Hickernell and Hong (1997) and Hickernell et al (2001) extended above lattices by introducing extensible lattice sequences. The authors constructed infinite sets of points, the first b^m of which form a lattice for a given base b and any $m \in \mathbb{N}$. Thus, if quadrature error using an initial lattice is too large, the lattice can be extended without discarding the original points. In related work, Cools et al (2006) have described how to construct embedded lattice rule sequences using CBC algorithms.

Digital nets form the second major class of low-discrepancy sets (Niederreiter, 1992; Larcher, 1998). For any $i \in \mathbb{N}$, let

$$i = \cdots i_3 i_2 i_1 \text{ (base } b) = \sum_{j=1}^{\infty} i_j b^{j-1}$$

be the b-nary representation, with $i_1, i_2, \ldots \in \{0, \ldots, b-1\}$, only a finite amount of which are non-zero. Define the function

$$\psi_b(i) = 0.i_1 i_2 i_3 \cdots = \sum_{j=1}^{\infty} i_j b^{-j} \in [0,1),$$

which reverses the digits for i about the decimal point. Given a set of generator matrices $\mathbf{C}_1, \ldots, \mathbf{C}_k$, a digital sequence is defined as

$$\big\{\big(\psi_b(z_{i1}), \psi_b(z_{i2}), \ldots, \psi_b(z_{ik})\big) : i = 0, 1, \ldots\big\},$$

where $z_{ij} = \cdots z_{ij3} z_{ij2} z_{ij1}$ (base b) and

$$\begin{pmatrix} z_{ij1} \\ z_{ij2} \\ z_{ij3} \\ \vdots \end{pmatrix} = \mathbf{C}_j \begin{pmatrix} i_1 \\ i_2 \\ i_3 \\ \vdots \end{pmatrix}.$$

A digital net is then the set containing the first b^m points of a digital sequence. Different choices of $\mathbf{C}_j$ are given in Halton (1960), Sobol (1967), Faure (1982), Niederreiter (1992) and Niederreiter and Xing (1998). Scrambled digital sequences were originally considered by Warnock (1972) and Braaten and Weller (1979). Braaten and Weller (1979) pointed out that scrambled Halton sequences have a significantly improved rate of convergence over non-scrambled sequences. Scrambling methods were popularized in a series of articles by Owen (1995, 1997a,b). Starting with the first coordinate, the cube is divided into b equal pieces. The results are then randomly permuted. This procedure is repeated for each of these b pieces by cutting them individually into b smaller pieces, which are then randomly permuted within the larger slices. This process is carried out several times. Finally, these permutations are repeated independently for the remaining coordinates; see Hickernell and Hong (2002) for an illustrative example. The resulting point set is again a digital set with probability one. Note that in general, scrambled lattices lose their lattice structure. Hong and Hickernell (2003) proposed efficient implementations of scrambling and compared the performances of these sequences on MVN integrals. The authors compared subregion adaptive methods (see Section 4.2.4) with the randomly shifted extensible lattices introduced above, the non-scrambled Niederreiter and Xing (1998) sequence and the Owen-scrambled Sobol sequence. Their results suggest that randomized QMC methods indeed perform better than competing methods. Sándor and András (2002) came to a similar conclusion; they also compared randomized QMC methods to antithetic MC and Latin hypercube sampling.

A Randomized QMC Algorithm for MVN and MVT Problems

We conclude this section by providing an algorithm to compute MVN and MVT probabilities for hyper-rectangular regions using the SOV reparameterization. The algorithm below uses a periodized, randomized QMC rule (Richtmyer, 1951) in the form

$$T_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma},\nu) \approx \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{j=1}^{N}\big(f(|2\{j\sqrt{\mathbf{p}}+\mathbf{\Delta}_i\}-\mathbf{1}|)\big),$$

with $\mathbf{p} = (2, 3, 5, \ldots, p_k)$, where p_j is the j-th prime, $\mathbf{\Delta}_i$ is a random $U[0,1]^k$ shift vector, and the integrand function f is the SOV integrand from equation (4.6).

1. **Input** $\mathbf{\Sigma}$, $\mathbf{a}$, $\mathbf{b}$, ν, N, M, and α.
2. Compute lower triangular Cholesky factor $\mathbf{L}$ for $\mathbf{\Sigma}$, permuting $\mathbf{a}$ and $\mathbf{b}$, and rows and columns of $\mathbf{\Sigma}$ for variable prioritization.
3. Initialize $T = 0, N = 0, V = 0$, and $\mathbf{q} = \sqrt{\mathbf{p}}$.
4. **For** $i = 1, 2, \ldots, M$
   a) Set $I_i = 0$, and generate uniform random $\mathbf{\Delta} \in [0,1]^k$
   b) **For** $j = 1, 2, \ldots, N$
      i. Set $\mathbf{w} = |2\{j\mathbf{q}+\mathbf{\Delta}\}-\mathbf{1}|$, $s = \chi_\nu^{-1}(w_k)$,
      $\mathbf{a}' = \frac{s\mathbf{a}}{\sqrt{\nu}}$, $\mathbf{b}' = \frac{s\mathbf{b}}{\sqrt{\nu}}$,
      $d_1 = \Phi\left(\frac{a_1'}{l_{1,1}}\right)$, $e_1 = \Phi\left(\frac{b_1'}{l_{1,1}}\right)$, and
      $f_1 = e_1 - d_1$.
      ii. **For** $m = 2, 3, \ldots, k$ set
      $y_{m-1} = \Phi^{-1}\left(d_{m-1} + w_{m-1}(e_{m-1} - d_{m-1})\right)$,
      $d_m = \Phi\left(\frac{a_m' - \sum_{n=1}^{m-1} l_{m,n}y_j}{l_{m,m}}\right)$,
      $e_m = \Phi\left(\frac{b_m' - \sum_{n=1}^{m-1} l_{m,n}y_j}{l_{m,m}}\right)$, and
      $f_m = (e_m - d_m)f_{m-1}$.
      End m loop
      iii. Set $I_i = I_i + (f_m - I_i)/j$;
      End j loop
   c) Set $\delta = (I_i - T)/i$, $T = T + \delta$, $V = (i-2)V/i + \delta^2$ and $E = \alpha\sqrt{V}$.
   End i loop
5. **Output** $T \approx T_k(\mathbf{a},\mathbf{b};\mathbf{\Sigma},\nu)$ with error estimate E.

The output error estimate E is the usual Monte Carlo standard error based on N samples of the randomly shifted QMC rule, and scaled by the confidence factor α. A sample size of $M = 12$, with $\alpha = 3$, produces E with at least 99% confidence. The algorithm can easily be modified to compute MVN probabilities by using $\mathbf{a}' = \mathbf{a}$ and $\mathbf{b}' = \mathbf{b}$ and removing the s computation in step 4(b)i.

4.2.3 Polynomial Integration Methods

For this section and the following section we assume that we have reparameterized a selected MVN or MVT problem so that we need to approximate a multivariate integral in the form

$$I(f) = \int_{\mathbf{0}}^{\mathbf{1}} f(\mathbf{u}) d\mathbf{u}$$

Traditional methods for this type of problem have used integration formulas (or rules), in the form

$$I(f) \approx \ B_N(f) = \sum_{i=1}^{N} w_i f(\mathbf{u}_i),$$

with points $\mathbf{u}_i$ and weights w_i (and N) chosen so that the rule $B_N(f)$ will integrate exactly all polynomials with degree $\leq d$ for some d. One motivation for this approach is that smooth integrands $f(\mathbf{u})$ can be closely approximated by polynomials, so a good rule $B_N(f)$ for polynomials should also be good for smooth functions. Another motivation is that once a good polynomial-integrating rule is found, it can be copied (transformed linearly) to any other bounded hyper-rectangular integration region without changing its degree d, and it can be applied to subdivisions of the initial integration region $[0,1]^k$ to increase accuracy.

An extensive amount of research has been devoted to finding multidimensional polynomial integrating rules with minimal point numbers N for specified d and dimension k; see the books by Stroud (1971) and Davis and Rabinowitz (1984) as well as the review article by Cools (1999). One focus for this work was the construction of good local integration rules for subregion adaptive methods (see Section 4.2.4). The most efficient rules were often constructed using many different methods, but more recently, large families of efficient rules based on "sparse grids" have been constructed from good one-dimensional rules (Genz, 1986; Gerstner and Griebel, 1998; Bungartz and Dirnstorfer, 2003), and some of these rules have been used for efficient MVN computation implementations.

4.2.4 Subregion Adaptive Methods

Subregion adaptive methods begin with the original integration region $[0,1]^k$, and dynamically construct a finer and finer subdivision of this integration region, with smaller subregions concentrated where the integrand is most irregular. Within each subregion a moderate degree local (*basic*) integration rule (or formula) is used to provide an estimate for the integral. These local results are combined to produce the global estimate. Although various basic rule types could be used, polynomial integrating basic rules are usually used with the subregion adaptive methods because they can provide rapid convergence once the subdivision has been refined enough that a low degree polynomial approximation can provide an accurate approximation to the integrand. Along with the chosen basic integration rule B, an associated error estimation method is also needed. Error estimates for polynomial rules typically use the difference between the basic rule and a lower degree local rule.

Typical input for this type of algorithm consists of (i) the integrand, (ii) an error tolerance ϵ and (iii) a work limit W on the total number of subregions allowed (this can either be used to limit the maximum time allowed or to limit the maximum workspace allowed). There are many possible adaptive strategies that may be used to dynamically refine a list of subregions. Below, we give a generic *globally adaptive integration* algorithm.

1. **Input** $f(\mathbf{w})$, an absolute error tolerance ϵ and a work limit W
2. Main loop:
   a) Begin with global estimates (from $B(f)$) for $I(f)$ and the error.
   b) **While** (total error $> \epsilon$) and (total work $<$ W) **Do**
      i. Determine new subdivision by subdividing largest error subregion.
      ii. Apply B to $f(\mathbf{w})$ in the new subregions.
      iii. Update the subregion list, and estimates for $I(f)$ and error.
      End While
3. **Output** estimates for $I(f)$ and the absolute error.

In order to avoid a rapid growth in the number of subregions, the subdivision step 2(a) in the generic globally adaptive algorithm must be done in a controlled way. A natural subdivision strategy subdivides the chosen region into 2^k pieces by halving along each coordinate axis. But with this strategy the number of subregions grows too rapidly and the differences in the behavior of the integrand as $\mathbf{w}$ varies in different directions is not accounted for. A better strategy was developed by van Dooren and de Ridder (1976), where the subdivision step 2(a) divides the selected subregion in half along the coordinate axis where the integrand is (locally) changing most rapidly. This allows the computational time to increase slowly and adaptation to occur only in those variables that cause most of the variation in the integrand. A detailed development of this type of algorithm with a robust error estimation method, and a carefully tested implementation was provided by Berntsen et al (1991a,b). This type of algorithm was used in tests by Genz (1993) and Genz and Bretz (2002) for SOV reparameterized MVN and MVT probabilities, and was very efficient for low-dimensional problems

4.2.5 Sparse-Grid Methods

Sparse-grid methods are based on a construction originally developed by Smolyak (1963). These have been carefully analyzed more recently because of efficiency and implementation considerations. The Smolyak construction starts with a family of one-dimensional integration rules

$$B_j(f) = \sum_{i=1}^{N_j} w_{i,j} f(u_{i,j}) \approx \int_0^1 f(u)du,$$

for $j = 0, 1, \ldots$, with $N_0 \geq 1$ and $N_{j+1} > N_j$. A (tensor) product rule, which uses one dimensional rules from this sequence, can be written in the form

$$B_{\mathbf{j}}(f) = \sum_{i_1=1}^{N_{j_1}} \sum_{i_2=1}^{N_{j_2}} \cdots \sum_{i_k=1}^{N_{j_k}} \left(\prod_{l=1}^{k} w_{i_l,j_l} f(u_{i_1,j_1}, u_{i_2,j_2}, \ldots, u_{i_k,j_k}) \right) \approx \int_{\mathbf{0}}^{\mathbf{1}} f(\mathbf{u}) d\mathbf{u}.$$

A family of $k-$dimensional sparse-grid rules using product rules constructed from the one-dimensional family $B_j(f)$ is then given by

$$S_d(f) = \sum_{\max\{0,d-k+1\} \le |\mathbf{j}| \le d} (-1)^{d-|\mathbf{j}|} \binom{k-1}{d-|\mathbf{j}|} B_{\mathbf{j}}(f),$$

with $|\mathbf{j}| = j_1 + j_2 + \ldots + j_k$. For example, $S_0(f) = B_{(0,0,\ldots,0)}(f)$,

$$S_1(f) = B_{(1,0,\ldots,0)}(f) + B_{(0,1,\ldots,0)}(f) + \ldots + B_{(0,0,\ldots,1)}(f) - (k-1) B_{(0,0,\ldots,0)}(f),$$

for $k > 1$, with $S_1(f) = B_1(f)$ when $k = 1$,

$$\begin{aligned} S_2(f) = {} & B_{(2,0,\ldots,0)}(f) + B_{(1,1,\ldots,0)}(f) + \ldots + B_{(0,0,\ldots,2)}(f) \\ & - (k-1)(B_{(1,0,\ldots,0)}(f) + B_{(0,1,\ldots,0)}(f) + \ldots + B_{(0,0,\ldots,1)}(f)) \\ & + \frac{(k-1)(k-2)}{2} B_{(0,0,\ldots,0)}(f), \end{aligned}$$

for $k > 2$, with $S_2(f) = B_2(f)$ when $k = 1$, and

$$S_2(f) = B_{(2,0)}(f) + B_{(1,1)}(f) + B_{(0,2)}(f) - B_{(1,0)}(f) - B_{(0,1)}(f)$$

when $k = 2$.

For efficiency, the one-dimensional rules usually have $N_0 = 1$ and $B_0(f) = f(1/2)$ (the mid-point rule). Popular choices for these one-dimensional rule families are Gauss, Patterson, or Chebyshev point rules (Genz, 1986; Gerstner and Griebel, 1998; Bungartz and Dirnstorfer, 2003). These $k-$dimensional rules inherit polynomial integrating properties for the one-dimensional rules, and if the $\{N_j\}$ sequence is not increasing too rapidly, $S_d(f)$ uses significantly fewer function values than a product rule $B_{(d,d,\ldots,d)}(f)$ with the same polynomial degree. If the function evaluation points for the one-dimensional rules form a nested sequence, then the points for $S_d(f)$ are a (sparse) subset of the (grid) points for $B_{(d,d,\ldots,d)}(f)$, and this motivates the "sparse-grid" nomenclature for these rules. Error estimates for these rules are usually computed using differences from successive rules in the $\{S_d(f)\}$ sequence. For lower dimensional problems, sparse-grid methods, when applied to the SOV reparameterized MVN and MVT problems, can be more efficient than randomized QMC methods (see Section 5.4 for further discussion).

4.2.6 Other Numerical Integration Methods

A number of other numerical integration methods have been used for MVN and MVT calculations. There is a good overview of the methods described

so far and other methods emphasizing on statistical applications in the book by Evans and Swartz (2000). The use of specialized Gauss rules (Steen et al, 1969) for MVN problems has been studied by Drezner (1992) and Yang (1998). Schervish (1984) has described the implementation of an algorithm that uses Newton-Cotes rules with error bounds, see also Bohrer and Schervish (1981) and Milton (1972). The specialized algorithms described in Sections 2.1.1 and 2.1.2 are examples of this for $k = 2$ and $k = 3$. Wang and Kennedy (1990) and Wang and Kennedy (1992) have developed self-validated MVN methods that use Taylor series with interval arithmetic.

5

Further Topics

In this chapter we consider subjects which were not discussed in the previous chapters. Topics include MVN and MVT problems with an integration region defined by a set of linear inequalities; evaluation of singular distribution functions; numerical tests of different integration methods; and integration problems having an application specific expectation function with MVN or MVT weight. A description of current software implementations in MATLAB and R is also given.

5.1 Linear Inequality Constraints

Some MVN and MVT problems have an integration region defined by a set of linear inequalities in the form

$$\mathbf{a} \leq \mathbf{M}\mathbf{x} \leq \mathbf{b},$$

where $\mathbf{M}$ is a $k \times k$ nonsingular matrix, in contrast to the simpler initial integration regions defined by $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$ (with a $k \times k$ identity matrix for M), which we have been considering so far. We will also consider a more general case, where M is not a square matrix, in the next section. If we make the Cholesky decomposition ($\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t$) change of variables $\mathbf{x} = \mathbf{L}\mathbf{y}$, then the integration region is determined by

$$\mathbf{a} \leq \mathbf{M}\mathbf{L}\mathbf{y} \leq \mathbf{b}.$$

Integration methods for the spherical-radial transformed problem (4.2) are easily adapted to this more general region. We simply define

$$R(\mathbf{z}) = \{r : r \geq 0, \mathbf{a} \leq r\mathbf{M}\mathbf{L}\mathbf{z} \leq \mathbf{b}\},$$

and there is no explicit change in the definition of the integration limits (which depend on $R(\mathbf{z})$) for the inner integral in equation (4.2). As was mentioned

A. Genz, *Computation of Multivariate Normal and t Probabilities*
DOI: 10.1007/978-3-642-01689-9_5, © Springer-Verlag Berlin Heidelberg 2009

previously, Lohr (1993) investigated using some spherical-radial methods for this type of problem.

Methods which use the separation-of-variables reparameterizations described in Section 4.1.2 can also be used for this type of problem if we make an additional change of variables. First, we need to determine an orthogonal matrix $\mathbf{Q}$ so that $\mathbf{MLQ}$ is lower triangular. This is a standard problem in numerical linear algebra (see Golub and Van Loan (1996)), where all we need to do is determine the QR decomposition for $(\mathbf{ML})^t$. The result is $(\mathbf{ML})^t = \mathbf{QR}$, with $\mathbf{Q}$ $k \times k$ orthogonal and $\mathbf{R}$ $k \times k$ upper triangular. This equation can be rewritten in the form $\mathbf{MLQ} = \mathbf{R}^t$ (a lower triangular matrix), and the set of inequalities can be rewritten in the form $\mathbf{a} \leq \mathbf{R}^t\mathbf{Q}^t\mathbf{y} \leq \mathbf{b}$. If the change of variables $\mathbf{y} = \mathbf{Qv}$ is used, then (because $\mathbf{y}^t\mathbf{y} = \mathbf{v}^t\mathbf{Q}^t\mathbf{Qv} = \mathbf{v}^t\mathbf{v}$) the resulting MVN and MVT integration problems have integration regions given by $\mathbf{a} \leq \mathbf{R}^t\mathbf{v} \leq \mathbf{b}$. This is the same type of integration region that was needed for use with equations (4.4) and (4.9), except that the $\mathbf{y}$ variables have been replaced by $\mathbf{v}$ variables. This analysis can be generalized for problems with an $\mathbf{M}$ matrix which is a singular or non-square matrix. The matrix $(\mathbf{LM})^t$ still has a QR factorization (with singular or non-square $\mathbf{R}$) and the integration region is still determined by a system of linear inequalities in the form $\mathbf{a} \leq \mathbf{R}^t\mathbf{v} \leq \mathbf{b}$. These types of problems are discussed in the next section.

5.2 Singular Distributions

If $|\mathbf{\Sigma}| = 0$ for a given covariance matrix, $\mathbf{\Sigma}^{-1}$ does not exist and (1.1) does not appear to be well-defined. For such *singular* MVN distributions the probability mass is concentrated on a linear subspace of $\mathbb{R}^k$. Assume that $(X_1, \ldots, X_k)^t$ follows a k-variate multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, that is, $(X_1, \ldots, X_k)^t \sim N_k(\boldsymbol{\mu}, \mathbf{\Sigma})$, such that $\text{rank}(\mathbf{\Sigma}) = r < k$. We can then find a $r \times k$ transformation matrix $\mathbf{C}$, so that $\mathbf{y} = \mathbf{C}(X_1, \ldots, X_k)^t \sim N_r(\mathbf{C}\boldsymbol{\mu}, \mathbf{C\Sigma C}^t)$. Motivated by this, Genz and Kwong (2000) and Bansal et al (2000) derived different methods for computing general singular MVN probabilities. The Genz and Kwong method uses a generalized Cholesky decomposition $\mathbf{\Sigma} = \mathbf{LL}^t$ with $l_{ij} = 0$ for all $j > r$, when $\text{rank}(\mathbf{\Sigma}) = r < k$ (Healy, 1968), which was also used by Lohr (1993). Then, Φ_k can be written with variables separated in a form similar to equation (4.1) as

$$\Phi_k(\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}; \mathbf{\Sigma}) = \int\limits_{a_1 \leq l_{11} y_1 \leq b_1} \phi(y_1) \quad \ldots \int\limits_{a_r \leq \mathbf{l}_r^t \mathbf{y} \leq b_r} \phi(y_r)\, dy_r \ldots dy_1.$$

The final set of $k - r$ constraints $a_i \leq \mathbf{l}_i^t\mathbf{y} \leq b_i,\ i > r$, are still valid for the r integration variables and must be included in the definition of the transformed domain. The remaining $k - r$ variables $y_{r+1}, y_{r+2}, \ldots, y_k$ are not constrained, so the associated terms in the form $\int_{-\infty}^{\infty} \phi(y_i) dy_i$, for $i = r+1, r+2, \ldots, k$,

which have value equal to 1 are not included in the equation for Φ_k. The SR methods described in Section 4.1.1 can also be used for the problem in this form.

The efficient use of the SOV methods described in Section 4.1.2 requires a more detailed analysis of the inequalities for the integration region. We first assume that (if necessary) a permutation of inequalities and a scaling has been completed so that the $\mathbf{L}$ matrix has the structure

$$\mathbf{L} = [\ \mathbf{L}'\ \mathbf{0}\ ] = \begin{bmatrix} 1 & 0 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & \dots & \dots & 0 \\ \star & 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \star & 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \star & \star & \dots & \star & 1 & 0 & \dots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \star & \star & \dots & \star & 1 & 0 & \dots 0 \end{bmatrix} \begin{matrix} 1 \\ \vdots \\ k_1 \\ 1 \\ \vdots \\ k_2 \\ \vdots \\ 1 \\ \vdots \\ k_r \end{matrix}, \tag{5.1}$$

where $\sum_{j=1}^{r} k_j = k$, $\star$ denotes an entry which could be zero or nonzero, and $\mathbf{L}'$ contains $\mathbf{L}$'s scaled and permuted columns $1, \dots, r$. For those cases where the scaling of an inequality requires division by a negative number, the inequality directions must be reversed, so that after scaling by a negative number, a scaled upper limit becomes a lower limit and a scaled lower limit becomes an upper limit. We use $\mathbf{a}'$ and $\mathbf{b}'$ to denote the new limit vectors after permutations, scalings, and interchanges of the original limit vectors $\mathbf{a}$ and $\mathbf{b}$, so that the new set of inequalities for the integration region takes the form $\mathbf{a}' \le \mathbf{L}'\mathbf{y} \le \mathbf{b}'$.

We can now produce explicit expressions for the limits of the successive integration variables. If we let $m_i = \sum_{j=1}^{i-1} k_j$, then the revised limits for y_i can be written as

$$\tilde{a}_i(y_1, \dots, y_{i-1}) = \max_{m_i < u \le m_{i+1}} \left(a'_u - \sum_{j=1}^{i-1} l'_{u,j} y_j \right),$$

and

$$\tilde{b}_i(y_1, \dots, y_{i-1}) = \min_{m_i < u \le m_{i+1}} \left(b'_u - \sum_{j=1}^{i-1} l'_{u,j} y_j \right).$$

The resulting expression for Φ_k becomes

$$\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma}) = \int_{\tilde{a}_1}^{\tilde{b}_1} \phi(y_1) \int_{\tilde{a}_2(y_1)}^{\tilde{b}_2(y_1)} \phi(y_2) \dots \int_{\tilde{a}_r(y_1,\dots,y_{r-1})}^{\tilde{b}_r(y_1,\dots,y_{r-1})} \phi(y_r) d\mathbf{y}. \tag{5.2}$$

The integration methods discussed in Section 4.2 can be applied directly to the problem in this form. A further reduction to a set of trivariate problems is also possible for MVN and MVT cdf problems using the methods described in Section 4.1.4.

If (as described in the preceding section) the original problem is given with an integration region defined by a set of linear inequalities in the form

$$\mathbf{a} \leq \mathbf{M}\mathbf{x} \leq \mathbf{b},$$

but $\mathbf{M}$ is an $m \times k$ matrix, with k not necessarily equal to m, then the transformation $\mathbf{x} = \mathbf{L}\mathbf{y}$ still produces the system of inequalities

$$\mathbf{a} \leq \mathbf{M}\mathbf{L}\mathbf{y} \leq \mathbf{b}.$$

In the singular covariance case, where $\mathbf{L}$ has zeros in the last $k - r$ columns, the last $k - r$ variables may be removed, and the system with r variables now has the form

$$\mathbf{a} \leq \mathbf{M}\mathbf{L}''\mathbf{y} \leq \mathbf{b},$$

where $\mathbf{L}''$ has the same structure as $\mathbf{L}'$ above, but it has not yet been scaled to have ones for the last nonzero element in each row. However, $\mathbf{M}\mathbf{L}''$ is not generally in (the SOV required) lower triangular form, so a QR decomposition of $(\mathbf{M}\mathbf{L}'')^t = \mathbf{Q}\mathbf{R}$) can be followed by the transformation $\mathbf{y} = \mathbf{Q}\mathbf{v}$ to produce an inequality system in the form $\mathbf{a} \leq \mathbf{L}'''\mathbf{v} \leq \mathbf{b}$, with $\mathbf{L}''' = \mathbf{R}^t$. A final scaling and relabeling of the variables results in the system

$$\mathbf{a}' \leq \mathbf{L}'\mathbf{y} \leq \mathbf{b}',$$

where $\mathbf{L}'$ has the same structure as in (5.1), except that the number of rows, $\sum_{j=1}^{r} k_j$, is now m. If $m \geq r$, then the method described in the preceding paragraph can be used. If $m < r$, there are fewer inequalities than there are variables, but because of the structure of the QR factorization, the $\mathbf{L}'$ matrix will have zeros in the last $r - m$ columns, so the associated variables can be ignored, leaving a problem in a form that can be also handled by the method described in the preceding paragraph.

The reparameterization that we have just described generalizes the methods discussed by Genz and Kwong (2000). Brodtkorb (2006) considered further refinements of these methods for singular MVN problems associated with wave distribution applications. Bansal et al (2000) used a principal component decomposition to reduce equation (1.1) to an $r-$dimensional integral. The authors proposed the use of linear programming techniques combined with Gauss-Legendre quadrature to handle the resulting set of inequalities on the integration variables. This method is usually much more computationally intensive, because of the extra linear programming work required for each integration point. An explicit representation of these transformations for the bivariate and trivariate normal cases based on the principal component decomposition is given by Rudolfer and Watson (1993). de Los Reyes (1990) provided upper bounds for singular MVN probabilities by integrating the standard MVN distribution over a hypersphere.

5.3 Related Integrals

We briefly consider the more general integration problem

$$E(g) = \int_{\mathbf{A}} f(\mathbf{x})g(\mathbf{x})d\mathbf{x},$$

where $g(\mathbf{x})$ is an application specific expectation function with MVN or MVT weight $f(\mathbf{x})$, so that

$$E(g) = \int_{\mathbf{A}} \phi_k(\mathbf{x}; \mathbf{\Sigma})g(\mathbf{x})d\mathbf{x}, \tag{5.3}$$

or

$$E(g) = \int_{\mathbf{A}} t_k(\mathbf{x}; \mathbf{\Sigma}, \nu)g(\mathbf{x})d\mathbf{x}, \tag{5.4}$$

5.3.1 $\mathbf{A} = \mathbb{R}^k$

A significant set of applications from Bayesian analysis and computational finance have $\mathbf{A} = \mathbb{R}^k$; see, for example, Shaw (1988), Ninomiya and Tezuka (1996), Caflisch et al (1997), Curbera (1998, 2000), Jäckel (2002), and Glasserman (2004). These integrals can be transformed to integrals over the unit hypercube if the reparameterizations introduced in Section 4.1 are used. With the MVN case, the simplest methods use the transformation $\mathbf{x} = \mathbf{S}\Phi^{-1}(\mathbf{w})$, where $\mathbf{\Sigma} = \mathbf{S}\mathbf{S}^t$ ($\mathbf{S}$ is a "square root" of $\mathbf{\Sigma}$) and Φ^{-1} is applied componentwise to the vector $\mathbf{w}$. Then,

$$E(g) = \int_{\mathbf{0}}^{\mathbf{1}} g(\Phi^{-1}(\mathbf{w}))d\mathbf{w},$$

and standard numerical integration methods for the unit hypercube can be applied to estimate $E(g)$.

Other methods start with the transformation $\mathbf{x} = \mathbf{S}\mathbf{y}$, followed by the SR transformation $\mathbf{y} = r\mathbf{z}$ so that

$$E(g) = \frac{2^{1-k/2}}{\Gamma(k/2)} \int_{||\mathbf{z}||=1} \int_0^\infty r^{k-1} e^{-r^2/2} g(r\mathbf{S}\mathbf{z})drdU(\mathbf{z}).$$

Special numerical integration methods for problems in this form have been studied by Monahan and Genz (1997) and Genz and Monahan (1998, 1999). A detailed comparison of several MC and QMC techniques involving a three dimensional integral with Gaussian weight is given by Caflisch (1998). We discuss some methods for the choice of a good $\mathbf{S}$ and more general transformations for these and related problems in Section 6.2.

5.3.2 $\mathbf{A} = [\mathbf{a}, \mathbf{b}]$

If $\mathbf{A}$ is the truncated region $\mathbf{A} = [\mathbf{a}, \mathbf{b}]$, with some finite components in $\mathbf{a}$ or $\mathbf{b}$, then a normalized $E(g)$ in the form

$$E(g) = \frac{1}{\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})} \int_{\mathbf{A}} \phi_k(\mathbf{x}, \mathbf{\Sigma}) g(\mathbf{x}) d\mathbf{x},$$

is usually needed. The SOV method can be used (Genz, 1992) to reparameterize the problem so that equation (4.5) becomes

$$E(g) = \frac{e_1 - d_1}{\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})} \int_0^1 (e_2(w_1) - d_2(w_1) \\ \ldots \int_0^1 (e_k(w_1, \ldots, w_{k-1}) - d_k(w_1, \ldots, w_{k-1})) \int_0^1 g(\mathbf{L}\Phi^{-1}(\mathbf{z}(\mathbf{w})))) d\mathbf{w}.$$

The $\Phi_k(\mathbf{a}, \mathbf{b}; \mathbf{\Sigma})$ normalization constant can be computed in conjunction with the expectation integral, saving some computation time associated with the transformations.

5.4 Numerical Tests

Several numerical studies have been carried out to compare the various integration techniques with each other. In the following we first survey those major studies which have compared methods designed to compute general MVN probabilities. Then we summarize the findings and make some recommendations.

Genz (1993) compared the methods of Schervish (1984) and Deák (1980, 1990), the subregion adaptive method of Berntsen et al (1991a,b), as well as both MC and randomized lattice rule implementations of the SOV reparameterization for random correlation matrices and $k < 15$. Hajivassiliou et al (1996) compared 13 simulators for $k \leq 16$ and the correlation matrices of either one-factorial or AR(1) structure. The comparisons included Deák's method, the control variate method of Stern (1992), a Gibbs sampler and the GHK simulator. No QMC method was included. (Recall that the GHK simulator is the same as MC sampling after the SOV reparameterization.) Gassmann et al (2002) compared 10 integration methods for $k \leq 20$ and randomly chosen correlation matrices. The study included the methods considered by Genz (1993), as well as several variations of the bounding procedure of Szántai (1986) and of a hybrid method originally proposed by Gassmann (1988). Sándor and András (2002) compared Latin hypercube sampling (McKay et al, 1979; Tang, 1993) and orthogonal arrays (Owen, 1992)

with randomized Niederreiter (1988) and Halton sequences (Wang and Hickernell, 2000) as well as randomized $(0, 2, s)$-nets in base 32 (Owen, 1995). The comparisons were performed for correlation matrices of either one-factorial or AR(1) structure with $k \leq 50$. Hong and Hickernell (2003) compared the subregion adaptive method with randomly shifted extensible lattices, non-scrambled Niederreiter and Xing (1998) sequences and Owen-scrambled Sobol sequences ($k \leq 15$, random correlation matrices). The only major comparison study including MVT computations was performed by Genz and Bretz (2002), who compared methods based on SR and SOV reparameterizations for several MC methods and different randomized lattice rule implementations ($k \leq 20$, random correlation matrices).

Summarizing the numerical tests, we can draw the following conclusions. Acceptance-rejection methods and their derivatives should generally be avoided (Genz, 1993; Hajivassiliou et al, 1996). This is also true for the method of Schervish (1984), because of its rapidly increasing computation times for $k > 6$ (Genz, 1993; Gassmann et al, 2002). The subregion adaptive method performs well for $k \leq 10$ (Hong and Hickernell, 2003; Genz and Bretz, 2002), but problems with unreliable error control have been reported occasionally (Gassmann et al, 2002). Among the reparameterizations, SOV and their variations seem to be more efficient than SR based methods (Genz, 1993; Hajivassiliou et al, 1996; Genz and Bretz, 2002), although Gassmann et al (2002) found SR based methods competed well for some problems involving very high correlations. Randomized QMC methods seem to perform better than deterministic QMC methods (Hong and Hickernell, 2003). Both methods, however, outperform MC methods by at least an order of magnitude (Genz and Bretz, 2002; Sándor and András, 2002). The question of which QMC method to use remains open and is left for future research. MVT related computations usually perform better, if reparameterizations start from equation (1.3) instead of equation (1.2) (Genz and Bretz, 2002). With currently available hardware, computations of four-digit MVN and MVT probabilities using randomized QMC methods require on average a few seconds of workstation time.

5.5 Software Implementations

In this section we give the links to some relevant web sites, which contain source code for some methods described in this monograph. Implementations in MATLAB[1] and R will be discussed separately in the subsequent sections. The websites of Genz (www.math.wsu.edu/faculty/genz/homepage) and Hajivassiliou (econ.lse.ac.uk/∼vassilis/) contain most of the algorithms discussed in Chapter 4. The Genz website contains FORTAN 77, FORTRAN 90 and MATLAB software. The FORTRAN 90 MVDIST package includes special functions for univariate, bivariate and trivariate normal and t distributions (including the non-central case), along with general purpose QMC software for

[1] MATLAB is a registered trademark of MathWorks, Inc.

MVN and MVT problems that can handle general linear inequality constraints and singular correlation matrices; it also includes functions for the computation of second and third order bounds and equicoordinate quantiles. The website of Bretz (www.biostat.uni-hannover.de/staff/bretz/) contains SAS/IML implementations for general MVN and MVT problems based on randomized QMC methods along with code for orthant probabilities. Source code published in *Applied Statistics* is available from lib.stat.cmu.edu/apstat/. This site includes the algorithms AS 195 (Schervish, 1984), AS 251 (Dunnett, 1989) and AS 285 Lohr (1993). The algorithms of Joe (1995) are available using anonymous ftp from ftp.stat.ubc.ca/pub/hjoe/mvnapp. FORTRAN code for the calculation of Owen's $T-$function using the hybrid approach of Patefield and Tandy (2000) is available at www.jstatsoft.org/v05/i05/. Some parallel MVN and MVT software is available from the ParInt website www.cs.wmich.edu/∼parint/. Recent information on MC and QMC methods and links to other related sites are obtained from www.mcqmc.org. Some newer (after 2007) versions of MATLAB contain implementations for MVN and MVT cdf functions in the Statistics Toolbox based on some of the methods developed by the authors of this book (try "help mvncdf" or "help mvtcdf" within MATLAB for details). There are also some Mathematica[2] functions (also based on methods developed by the current authors) in the Mathematica Statistic Package for MVN computations (searching for "Multinormal Distribution" within the Help section will provide further information).

5.5.1 R

R is a language and environment for statistical computing and graphics (Ihaka and Gentleman, 1996). It provides a wide variety of statistical and graphical techniques, and is highly extensible. The latest version of R is available at the Comprehensive R Archive Network (CRAN), which can be accessed from www.r-project.org. In this section we illustrate the use of the R package `mvtnorm`, which computes MVN and MVT probabilities (Hothorn et al, 2001). The `mvtnorm` package includes additional functionality (calculation of density values, random variable generation, etc.), some of which is described in Appendix A.

The `mvtnorm` package uses a careful implementation of the methods described in Chapter 4. SOV methods are used, which provide a transformation of the original integration problem to the unit hypercube $[0, 1]^k$ (Section 4.1.2). Variable reordering, as described in Section 4.1.3, is additionally implemented to improve integration efficiency. Several suitable standard integration routines can be applied to this transformed integration problem. The `mvtnorm` package uses an extension of the lattice rules described in Section 4.2.2 combined with an antithetic variable sampling (Section 4.2.1). Robust integration error bounds are obtained by introducing additional shifts of the entire set of

[2] Mathematica is a registered trademark of Wolfram Research, Inc.

integration nodes in random directions. Since this additional randomization step is only performed to introduce a robust Monte Carlo error bound, 10 simulation runs are usually sufficient.

In order to illustrate the `mvtnorm` package, we start considering the three-dimensional problem (1.5). We can use the following statements to calculate the required probabilities,

```
> library(mvtnorm)
> k <- 3
> R <- diag(3)
> R[2,1] <- 3/5
> R[3,1] <- 1/3
> R[3,2] <- 11/15
> pmvnorm(mean=rep(0,k), R, lower=rep(-Inf,k),
+         upper=c(1,4,2))
[1] 0.8279847
attr(,"error")
[1] 2.870095e-07
attr(,"msg")
[1] "Normal Completion"
```

In the beginning, we call the `mvtnorm` library to load the required functions. We then enter the lower triangular of the correlation matrix $\mathbf{R}$. The non-centrality vector is passed to `pmvnorm` by specifying the argument `mean`. The integration region is given by the vectors `lower` and `upper`, both of which can have elements `-Inf` or `+Inf`. The value of `pmvnorm` is then the estimated integral value with the two attributes `error` (the estimated absolute error) and `msg` (a status message, indicating whether or not the algorithm terminated correctly). From the results above it follows that

$$\Phi_{3ex} = P(-\infty < X_1 \leq 1, -\infty < X_2 \leq 4, -\infty < X_3 \leq 2) \approx 0.82798$$

with an absolute error estimate of $2.9e-07$.

We next consider the singular problem (1.7), in which case the `mvtnorm` package uses the methods described in Section 5.2. As mentioned in Section 1.3, the trivariate problem essentially reduces to a univariate problem. If we call

```
> one <- rep(1,k)
> R <- one%*%t(one)
> p <- pmvnorm(mean=rep(0,k), R, lower=rep(-Inf,k),
+              upper=rep(1,k))
> p
[1] 0.8413447
attr(,"error")
[1] 0
attr(,"msg")
[1] "Normal Completion"
```

and compare the result with the univariate normal distribution call

```
> pnorm(1)
[1] 0.8413447
```

we see that the results coincide.

The `mvtnorm` package allows the user to pass either the covariance or the correlation matrix. Consider the following non-central MVT problem. We now use the `pmvt` function to calculate the MVT probabilities,

```
> a <- pmvt(lower=-Inf, upper=2, delta=rep(0,5), df=3,
+           sigma = diag(5)*2)
> b <- pmvt(lower=-Inf, upper=2/sqrt(2), delta=rep(0,5),
+           df=3, corr=diag(5))
> attributes(a) <- NULL
> attributes(b) <- NULL
> all.equal(round(a,3), round(b,3))
[1] TRUE
```

The second of the previous statements passes the correlation matrix to `pmvt`. Note that the integration bounds have to be standardized accordingly, but not the non-centrality parameter.

Finally, we compare `pmvt` and `pmvnorm` for large degrees of freedom. If we compute the MVT probabilities for increasing values of ν using

```
> foo <- function(x) pmvt(lower=-Inf, upper=1,
+        delta=rep(0,5), x, corr=diag(5))
> x <- as.matrix(c(1:19,10*(2:20)))
> p <- apply(x, 1, foo)
```

and plot the resulting probabilities, we obtain the graph shown in Figure 5.1, where the horizontal dotted line indicates the limiting MVN probability. As seen from Figure 5.1, the MVT and MVN probabilities are very similar for large values of ν.

5.5.2 MATLAB

MATLAB is a language and environment for numerical computations. It is widely used for scientific and engineering computations. In this section we illustrate the use of some MATLAB functions for MVN and MVT computations which are available from the Genz website by following the "Software" link (www.math.wsu.edu/faculty/genz/homepage). All functions have introductory comment sections which describe the function's input and output parameters; some of this information is given in Appendix B. In the following we describe the current list of available functions.

bvnl : a function for the computation of bivariate normal cdf probabilities. This function uses methods described in Section 2.1.1 and Genz (2004).

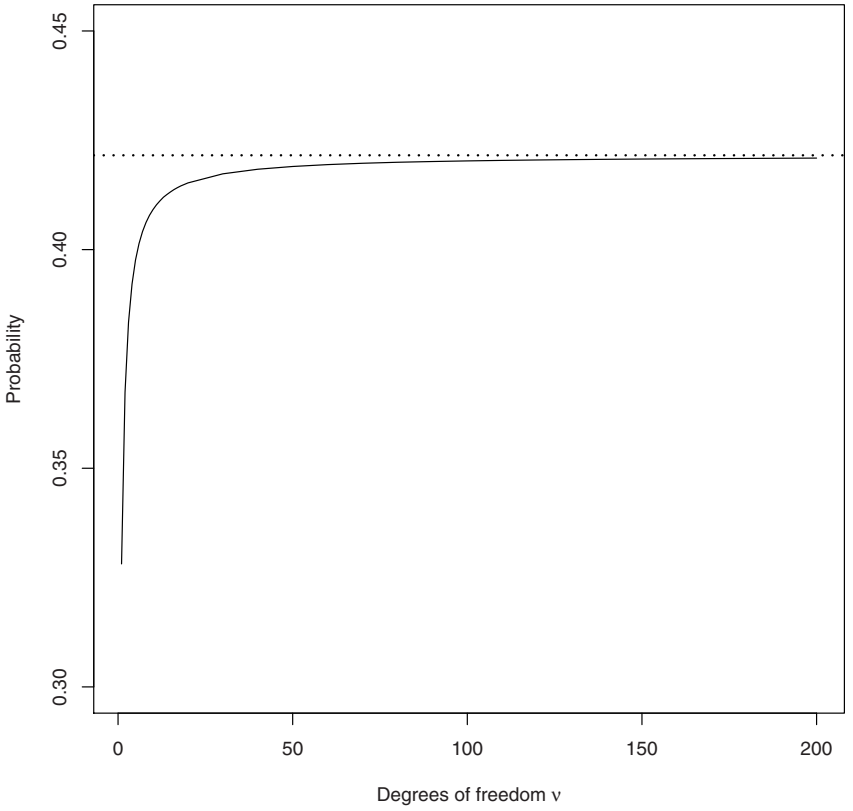

Fig. 5.1. MVT probabilities for increasing degrees of freedom

tvnl : a function for the computation of univariate, bivariate and trivariate cdf normal probabilities. These functions use methods described in Section 2.1.2 and Genz (2004).

tvnls : a simpler but less accurate version of **tvnl**.

qsimvn : a function for the numerical computation of multivariate normal distribution values. This function uses the SOV reparameterization described in Section 4.1.2 with variable reordering (see Section 4.1.3) and a quasi-random integration method with a randomized point set based on a Kronecker sequence (see Section 4.2.2).

qsimvnv : a vectorized version of **qsimvn**, with similar parameters.

qsilatmvnv: similar to **qsimvnv**, except that randomized lattice rules (see Section 4.2.2) are used for the numerical integration.

qscmvn : a function for the numerical computation of multivariate normal distribution values. The integration region may be specified by a set of linear inequalities in the form $\mathbf{a} < \mathbf{Cx} < \mathbf{b}$, where $\mathbf{C}$ is an $m \times k$ matrix. This function uses a method similar to **qsimvn** combined with methods for handling singular and inequality constrained problems described in Section 5.1.

qscmvnv : a vectorized version of **qscmvn**, with similar parameters.

qsclatmvnv : similar to **qscmvnv**, except randomized lattice rules (see Section 4.2.2) are used for the numerical integration.

mvnlps : a function for the numerical computation of multivariate normal distribution values for ellipsoidal integration regions. This function uses methods described in Section 2.2.2 and Sheil and O'Muircheartaigh (1977).

qsimvnef : a function for the numerical computation of multivariate normal distribution expected values. The method used is similar to the method used for **qsimvn**, but also computes the expected value of a user specified function; see Section 5.3.2.

qsimvnefv : a vectorized version of **qsimvnef**, with similar parameters.

tvtl : a set of functions for the computation of univariate, bivariate and trivariate cdf normal and t probabilities. These functions use methods described in Section 2.1.2 and Genz (2004).

qsimvt : a function for the numerical computation of multivariate t distribution values. This function uses the MVT distribution in the form (1.3) with the SOV reparameterization described in Section 4.1.2, variable reordering (see Section 4.1.3) and a Niederreiter sequence (see Section 4.2.2) quasi-random integration method.

qsilatmvtv : similar to **qsimvt**, except that randomized lattice rules (see Section 4.2.2) are used for the numerical integration and and the function references are vectorized.

qscmvt : a function for the numerical computation of MVT probabilities. The integration region may be specified by a set of linear inequalities in the form $\mathbf{a} < \mathbf{Mx} < \mathbf{b}$, where $\mathbf{M}$ is an $m \times k$ matrix. This function uses a method similar to **qsimvt** combined with methods for handling singular and inequality constrained problems described in Section 5.1.

qscmvtv : a vectorized version of **qscmvt**, with similar parameters.

We now provide some examples to illustrate the MATLAB functions. Consider again the Φ_{3ex} MVN problem (1.5), which was also used in Section 5.5.1. We can use the following MATLAB statements to obtain some example results.

```
>> b = [1 4 2]'; r3 = [3/5 1/3 11/15];
>> p = tvnl( b, r3, 1e-12 ); disp(p)
         0.827984897456834
>> a = -inf*[1 1 1]'; r = [1 3/5 1/3;3/5 1 11/15;1/3 11/15 1];
>> [ p e ] = qsilatmvnv( 10000, r, a, b ); disp([p e])
```

```
        0.827984872246485       8.05123962846801e-08
```

The first result for Φ_{3ex} is the highly accurate 0.827984897456834 value from **tvnl** and the second result comes from a 1000-point randomized lattice rule simulation. The error estimate for this result is consistent with the **tvnl** result. A related problem is the T_{3ex} MVT problem (1.6), which has the same parameters, except for the degrees of freedom ($\nu = 5$). We can use a similar set of MATLAB statements to obtain example results.

```
>> b = [1 4 2]'; r3 = [3/5 1/3 11/15];
>> p = tvlt( 5, b, r3, 1e-12 ); disp(p)
         0.791453793811934
>> a = -inf*[1 1 1]'; r = [1 3/5 1/3;3/5 1 11/15;1/3 11/15 1];
>> [ p e ] = qsimvt( 10000, 5, r, a, b ); disp([p e])
         0.791561147814799       0.00109331493304781
>> [ p e ] = qsilatmvtv( 100000, 5, r, a, b ); disp([p e])
         0.791454692453579      1.04944158384174e-06
```

The results are similar to the Φ_{3ex} results, but the simulated results require significantly more computation to achieve high accuracy. Some results from a related singular problem where all $\mathbf{\Sigma}$ entries are 1:

```
>> b = [1 4 2]'; r3 = [1 1 1];
>> p = tvnl( b, r3, 1e-14 ); disp(p)
         0.841344746068543
>> a = -inf*[1 1 1]'; r = ones(3);
>> [ p e ] = qsilatmvnv( 10000, r, a, b ); disp([p e])
         0.841344746068551                           0
```

In this case the simulated result is (within rounding) identical to the **tvnl** result because both functions reduce the problem to a one-dimensional problem.

Some MATLAB results for the 5-dimensional Φ_{5ex} problem (1.8) are

```
>> a = [-5:-1]'; b = -a + 1;
>> c = tril(ones(5)); r = c*c';
>> [ p e ] = qsilatmvnv( 10000, r, a, b ); disp([p e])
         0.474128241397226       2.6890463317249e-06
>> [ p e ] = qsilatmvnv( 100000, r, a, b ); disp([p e])
         0.474128319994297      2.67591852853959e-07
```

The two different simulation results illustrate approximately linear decrease in estimated errors as the number of samples is increased. Similar results were obtained with a positive orthant probability for the same covariance matrix:

```
>> ao = zeros(5,1); bo = inf*ones(5,1);
>> c = tril(ones(5)); r = c*c';
>> [ p e ] = qsilatmvnv(  10000, r, ao, bo ); disp([p e])
         0.246105114254968      2.09977198987622e-05
```

```
>> [ p e ] = qsilatmvnv( 100000, r, ao, bo ); disp([p e])
          0.24609474118263      3.95720168886173e-06
```

The use of the MATLAB function **qsclatmvnv** which allows the integration region to be defined in terms of a system of linear inequalities (see Section 5.1) can be illustrated with this problem. Suppose the integration region is the simplex defined by the inequalities $0 \leq \sum_{i=1}^{5} x_i \leq 1$ with $0 \leq x_i \leq 1,\ i = 1, \ldots, 5$, which can be written as $\mathbf{0} \leq \mathbf{Mx} \leq \mathbf{1}$ with

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Some MATLAB results are

```
>> as = zeros(6,1); bs = ones(6,1); M = [eye(5); ones(1,5)];
>> c = tril(ones(5)); r = c*c';
>> [ p e ] = qsclatmvnv(  10000, r, as, M, bs ); disp([ p e ])
   7.4946e-05    7.665e-07
>> [ p e ] = qsclatmvnv( 100000, r, as, M, bs ); disp([ p e ])
   7.5042e-05   9.6451e-08
```

5.6 Miscellaneous and Further References

In this subsection we give additional references which have not been discussed before. Kotz et al (2000) is a good starting point for further problems.

DiDonato et al (1980) and DiDonato and Hageman (1982) presented some methods for the computation of bivariate normal probabilities over polygons. Drezner (1990) and Hutchinson (1993) proposed simple approximations to tri- and quadrivariate normal probabilities. Wang (1987) partitioned the $(x_1, x_2)-$plane into fine rectangular regions to calculate bivariate normal probabilities and modeled the resulting contingency table by an association model from Goodman (1981). Extensions to higher dimensions were given by Rom and Sarkar (1990) and Wang (1997). A series of structural safety articles focused on approximating the MVN integral after the Cholesky decomposition (4.1); see Hohenbichler and Rackwitz (1983, 1985), Gollwitzer and Rackwitz (1987), Tang and Melchers (1987), Ambartzumian et al (1998) and Pandey (1998), to name a few. Other approaches for computing general MVN probabilities include James (1991), Drezner (1992), and Berens et al (1995). Henery (1981) approximated truncated MVN probabilities based on a Gram-Charlier expansion. Siddiqui (1967), Krishnan (1972) and Bulgren et al (1974) considered computational aspects of bivariate t distributions with different standardizations than those considered in this book.

Most of the algorithms for MVN and MVT computations involve computationally intensive but highly repetitive evaluations of integrands for reparameterized probability distribution integrals. These algorithms can therefore be easily parallelized, with a resulting significant reduction in total computation time. Various authors have described parallelization and implementation of MVN and MVT algorithms, see, for example, Hajivassiliou (1993b), Wang and Kennedy (1997), and de Doncker et al (1999, 2001).

6

Applications

In this chapter we consider a variety of applications, which require efficient numerical MVN and/or MVT integration methods. In Section 6.1 we discuss the numerical computation of high-dimensional integrals for multiple comparison problems. In Section 6.2 we discuss the application of MVN and MVT integration methods for problems in computational finance and Bayesian statistical analysis.

6.1 Multiple Comparison Procedures

We consider classes of high dimensional integrals that are needed for the application of multiple comparison procedures. The numerical integration problems involve computation of MVT or MVN probabilities with integration over regions determined by sets of linear inequalities (Hochberg and Tamhane, 1987; Hsu, 1996; Bretz et al, 2010). In Section 6.1.1 we introduce some notation and discuss how the methods from the previous chapters can be applied to common multiple test problems. In Section 6.1.2 we consider the calculation of critical values, which leads to the numerical problem of combining an appropriate optimization method with an efficient numerical integration method. We refer to Bretz et al (2001) for a more detailed review of numerical integration problems arising in the context of multiple comparison procedures.

6.1.1 Multiple Testing in Linear Models

We consider a general linear model with fixed effects,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

We assume that we are given an $n \times 1$ data vector $\mathbf{Y}$, a fixed and known $n \times p$ design matrix $\mathbf{X}$, an unknown $p \times 1$ parameter vector $\boldsymbol{\beta}$ and a $n \times 1$ error vector $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ with unknown variance σ^2. Let

A. Genz, *Computation of Multivariate Normal and t Probabilities*
DOI: 10.1007/978-3-642-01689-9_6, © Springer-Verlag Berlin Heidelberg 2009

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^-\mathbf{X}^t\mathbf{Y}$$

and

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\nu}$$

denote the least square unbiased estimates of $\boldsymbol{\beta}$ and σ, respectively, where $\nu = n - \text{rank}(\mathbf{X})$ and $(\mathbf{X}^t\mathbf{X})^-$ denotes some generalized inverse of $\mathbf{X}^t\mathbf{X}$.

The setup for multiple comparison problems provides a constant $p \times k$ matrix $\mathbf{C}$. The covariance matrix for the multiple comparison problem is $\boldsymbol{\Sigma} = \mathbf{C}^t(\mathbf{X}^t\mathbf{X})^-\mathbf{C}$, an $k \times k$ positive semi-definite matrix. The matrix $\mathbf{C}$ specifies the k two-sided (lower-tailed, upper-tailed) null hypotheses $H_i^= : \mathbf{c}_i^t\boldsymbol{\beta} = 0$ ($H_i^\leq : \mathbf{c}_i^t\boldsymbol{\beta} \leq 0, H_i^\geq : \mathbf{c}_i^t\boldsymbol{\beta} \geq 0$), where $\mathbf{c}_i^t$ denotes the $i-$th column of $\mathbf{C}$, $i = 1, \ldots, k$. Each null hypothesis is tested using the test statistics

$$T_i = \frac{\mathbf{c}_i^t\hat{\boldsymbol{\beta}}}{\hat{\sigma}\sqrt{\mathbf{c}_i^t(\mathbf{X}^t\mathbf{X})^-\mathbf{c}_i}}, \quad i = 1, \ldots, k.$$

The joint distribution of $T_1, \ldots, T_k$ is as follows. Let $Z_i = \hat{\sigma}T_i/\sigma$. Then it follows that $\mathbf{Z} = (Z_1, \ldots, Z_k)^t \sim N_k(\boldsymbol{\delta}, \mathbf{R})$, where $\mathbf{R}$ is obtained by standardizing $\boldsymbol{\Sigma}$ and the non-centrality vector $\boldsymbol{\delta}$ is given by

$$\boldsymbol{\delta} = \left(\frac{\mathbf{c}_i^t\boldsymbol{\beta}}{\sigma\sqrt{\mathbf{c}_i^t(\mathbf{X}^t\mathbf{X})^-\mathbf{c}_i}}\right)_{i=1,\ldots,k}.$$

Because under the conditions above $\nu\hat{\sigma}^2$ is χ^2_ν distributed and independent of $\mathbf{Z}$, the vector $(T_1, \ldots, T_k)^t = \mathbf{Z}/\sqrt{\chi^2_\nu}$ follows a $k-$variate t distribution with parameters $\mathbf{R}$, ν, and $\boldsymbol{\delta}$ (Cornish, 1954; Dunnett and Sobel, 1954). Under the null hypothesis, $\boldsymbol{\delta} = \mathbf{0}$ and we need to calculate central MVT probabilities based on either (1.2) or (1.3). Thus, if t_i denotes the observed value for T_i, we reject

$$\begin{cases} H_i^=, & \text{if } 1 - T_k(-\mathbf{t}, \mathbf{t}; \boldsymbol{\Sigma}, \nu) \leq \alpha, \\ H_i^\leq, & \text{if } 1 - T_k(-\infty, \mathbf{t}; \boldsymbol{\Sigma}, \nu) \leq \alpha, \\ H_i^\geq, & \text{if } 1 - T_k(-\mathbf{t}, \infty; \boldsymbol{\Sigma}, \nu) \leq \alpha, \end{cases} \tag{6.1}$$

where $\mathbf{t} = (t_i, \ldots, t_i)^t$ denotes an $k \times 1$ vector with equal entries t_i and $\alpha \in (0, 1)$ denotes the significance level. Otherwise, $\boldsymbol{\delta} \neq \mathbf{0}$ and we need to calculate non-central MVT probabilities based on (1.4), as required, for example, for power calculations and sample size determination (Bretz et al, 2001). In the asymptotic case $\nu \to \infty$ or if σ is known, the calculations reduce to the evaluation of the MVN probabilities (1.1).

To illustrate the framework described above, we discuss two common multiple comparison procedures in more detail. We first consider the Dunnett test for comparing k treatments with a control in a standard one-factorial analysis-of-variance (ANOVA) model (Dunnett, 1955). Here, $p = k + 1$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{k+1})^t$ denotes the parameter vector with the treatment means

β_i, where the index $i = 1$ denotes the control, to which the remaining k treatments are compared with. The associated $\mathbf{C}$ matrix is

$$\mathbf{C}_{k+1\times k} = \begin{bmatrix} -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Let n_i denote the sample size of treatment group i, such that $n = \sum_{i=1}^{k+1} n_i$. Straightforward algebra shows that after standardizing $\mathbf{\Sigma} = \mathbf{C}^t(\mathbf{X}^t\mathbf{X})^-\mathbf{C}$ the $k \times k$ correlation matrix $\mathbf{R}$ can be written as $\mathbf{R} = \mathbf{D} + \mathbf{v}\mathbf{v}^t$, with $\mathbf{D} = \text{diag}(1 - \lambda_i^2)$, $v_i = \lambda_i$, and $\lambda_i = \sqrt{\frac{n_i}{n_1+n_i}}$. This leads to a product correlation structure and equation (2.16) can be used. Thus, the computation of MVN probabilities for the Dunnett test reduces to the computation of a one-dimensional integral over $\mathbb{R}$ with a Gaussian weight function, see Dunnett (1989). In the MVT case, we have an additional integral for the χ_ν variable. Note that the product correlation structure does not hold in general ANOVA models, see Hsu (1996, Chapter 7). In such cases the general purpose methods from Chapter 4 have to be used.

As a second example for a multiple comparison procedure we consider all pairwise comparisons between p treatments, resulting in a total of $k = p(p-1)/2$ two-sided hypotheses. The studentized range test of Tukey (1953) is a standard approach. For $p = 3$ we have $k = 3$ and

$$\mathbf{C}_{3\times 3} = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Figure 6.1 illustrates the associated integration region, which consists of an infinite hexagonal cylinder, see also Hsu (1996, p. 120). One immediately recognizes this is a singular integration problem and the methods from Section 5.2 have to be applied. Consider an all pairwise comparison problem with $p = 10$ and sample sizes $12, 14, \dots, 30$. In this case $(\mathbf{X}^t\mathbf{X})^-$ is a diagonal matrix with entries $1/12, 1/14, \dots, 1/30$, and the matrix $\mathbf{C}$ has 45 columns in the form

$$\mathbf{C}_{10\times 45} = \begin{bmatrix} -1 & -1 & \dots & -1 & 0 & \dots & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & -1 & \dots & -1 & \dots & \vdots \\ 0 & 1 & \dots & \vdots & 1 & \dots & 0 & \dots & \vdots \\ \vdots & 0 & \dots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots & 0 \\ \vdots & \vdots & \dots & 0 & \vdots & \dots & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 & 0 & \dots & 1 & \dots & 1 \end{bmatrix}.$$

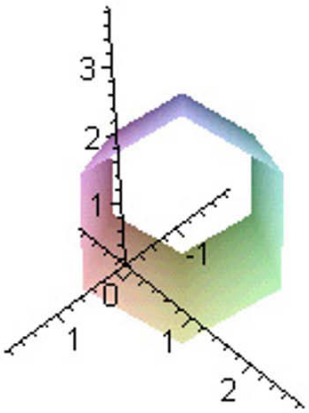

Fig. 6.1. Cylindric hexagonal integration region for Tukey test, $p = 3$

If the transformations from Section 5.2 are applied, then the final Cholesky factor $\mathbf{L}^t$ takes the form

$$\mathbf{L}^t = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ \star & 1 & 0 & \cdots & \cdots & 0 \\ \star & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \star & \star & \cdots & \cdots & \star & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \star & \star & \cdots & \cdots & \star & 1 \end{bmatrix};$$

$\mathbf{L}^t$ has nine columns.

The framework described above follows the canonical linear model theory and thus includes classical regression and ANOVA models allowing for covariates and/or factorial treatment structures with identically and independently distributed normal errors and constant variance, see Bretz et al (2008) and the references therein. This framework can be extended to more general parametric and semi-parametric models based on standard asymptotic results, thus requiring the efficient computation of multivariate normal probabilities (Hothorn et al, 2008). We refer the reader to Bretz et al (2010) for further

details and a detailed description of the `multcomp` package in R, which provides a general software implementation of multiple comparisons procedures in generalized linear models, mixed-effects models, survival models, etc.

6.1.2 Numerical Computation of Critical Values

A different problem from equation (6.1) is the determination of simultaneous confidence intervals for $\mathbf{c}_i^t\boldsymbol{\beta}$, which – in the two-sided case – take the form

$$\left[\mathbf{c}_i^t\hat{\boldsymbol{\beta}} - t_{1-\alpha}\hat{\sigma}\sqrt{\mathbf{c}_i^t(\mathbf{X}^t\mathbf{X})^-\mathbf{c}_i};\ \mathbf{c}_i^t\hat{\boldsymbol{\beta}} + t_{1-\alpha}\hat{\sigma}\sqrt{\mathbf{c}_i^t(\mathbf{X}^t\mathbf{X})^-\mathbf{c}_i}\right].$$

The actual numerical problem consists of finding the *critical value* $t_{1-\alpha}$ where $P(t_{1-\alpha}) = 1 - \alpha$, with

$$P(t) = \begin{cases} T_k(-\mathbf{t}, \mathbf{t}; \boldsymbol{\Sigma}, \nu) & \text{(two-sided)}, \\ T_k(-\infty, \mathbf{t}; \boldsymbol{\Sigma}, \nu) & \text{(lower-tailed)}, \\ T_k(-\mathbf{t}, \infty; \boldsymbol{\Sigma}, \nu) & \text{(upper-tailed)}. \end{cases}$$

Here, $\mathbf{t}$ denotes a $k \times 1$ vector with equal entries t. The numerical problem is therefore a combined problem of using an appropriate numerical optimization method to determine $t_{1-\alpha}$ with an efficient numerical integration method for evaluating $P(t)$.

The numerical optimization-integration methods considered below use the function

$$h(t) = P(t) - (1 - \alpha).$$

These methods involve finding $t_{1-\alpha}$, the point where $h(t_{1-\alpha}) = 0$, using a numerical optimization method. Since $h(t)$ is often expensive to compute using numerical integration, particularly for large values of k, a numerical optimization method that requires only a few iterations is needed. We can accomplish this if we combine a method for getting good starting points for the optimization step with an optimization method that converges rapidly. An additional complication with the combined numerical optimization-integration method is the presence of numerical integration errors, which must be controlled along with the numerical optimization errors. Below, we discuss these issues along the lines of Genz and Bretz (2000).

Starting Interval Selection

For the purpose of obtaining a good starting interval for $t_{1-\alpha}$, the methods from Section 3.1 can be used. Let

$$L_1(t) = 1 - S_1(t) \leq P(t)$$

and

$$P(t) \leq 1 - \min_j P(A_j^c(t)) = U_1(t)$$

denote the lower and upper bounds introduced in Section 3.1 based on one-dimensional distribution values. If $t_{1,a}$ and $t_{1,b}$ are determined by solving $U_1(t) = 1 - \alpha$ and $L_1(t) = 1 - \alpha$, respectively, then $t_{1-\alpha} \in [t_{1,a}, t_{1,b}]$. This bounding interval for $t_{1-\alpha}$ can be found directly using the appropriate one-dimensional inverse distribution function. For example, with the two-sided case, $[t_{1,a}, t_{1,b}] = [T^{-1}(1 - \frac{\alpha}{2}; \nu), T^{-1}(1 - \frac{\alpha}{2k}; \nu)]$. As shown in Section 3.1, shorter intervals can be found using bivariate distribution values if the modified Bonferroni bound $U_2(t)$ from Dawson and Sankoff (1967) is combined with the Hunter-Worsley bound $L_2(t)$ (Hunter, 1976; Worsley, 1982). If $U(t_{2,a}) = 1 - \alpha$ and $L(t_{2,b}) = 1 - \alpha$ then $t_{1,a} \le t_{2,a} \le t_{1-\alpha} \le t_{2,b} \le t_{1,b}$. Starting with $[t_{1,a}, t_{1,b}]$, we can use numerical optimization, applied to $L_2(t)$, to determine $t_{2,b}$, then use numerical optimization, applied to $U_2(t)$ starting with $[t_{1,a}, t_{2,b}]$, to determine $t_{2,a}$. We refer the reader to Genz et al (2004) for an investigation of such lower and upper bounds for multiple comparison problems.

Choice of Optimization Method

A primary goal for selecting an optimization method is finding a method that, given good starting points, requires only a few iterations for a large class of problems. Thus, it would be desirable to use a second order method like Newton's method to find $t_{1-\alpha}$. The Newton iteration method for improving an estimate for t_c approximating $t_{1-\alpha}$, successively replaces t_c by $t_c - h(t_c)/h'(t_c)$. This method requires values for both $h(t)$ and $h'(t)$. If we make a simple change of variable $\mathbf{y} = t\mathbf{x}$ in the detailed expression for $h(t)$ determined from the definition of the MVT distribution function, we have (for the two-sided case)

$$h(t) = \alpha - 1 + t^k \, \frac{2^{1-\frac{\nu}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2})(2\pi)^{\frac{k}{2}}} \int_0^{\infty} s^{\nu-1} e^{-\frac{s^2}{2}} \int_{-\frac{s}{\sqrt{\nu}}}^{\frac{s}{\sqrt{\nu}}} e^{-\frac{t^2 \mathbf{y}^t \mathbf{\Sigma}^{-1} \mathbf{y}}{2}} d\mathbf{y} ds.$$

Differentiating $h(t)$, and then changing the variables back to $\mathbf{x}$, we find

$$h'(t) = \frac{1}{t}(kP(t) - H(t)),$$

where $H(t)$ is given by

$$\frac{2^{1-\frac{\nu}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2})(2\pi)^{\frac{k}{2}}} \int_0^{\infty} s^{\nu-1} e^{-\frac{s^2}{2}} \int_{-\frac{s\mathbf{t}}{\sqrt{\nu}}}^{\frac{s\mathbf{t}}{\sqrt{\nu}}} \mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x} \; e^{-\frac{\mathbf{x}^t \mathbf{\Sigma}^{-1} \mathbf{x}}{2}} d\mathbf{x} ds.$$

For the lower-tailed (upper-tailed) case, the lower (upper) limits for the inner integral are all $-\infty$ (∞). In either case, $H(t)$ may be computed with only a little extra work during the computation of $P(t)$.

Given a starting interval $[t_a, t_b]$ with $t_{1-\alpha} \in [t_a, t_b]$ and a required error tolerance τ for the final estimate of $t_{1-\alpha}$, we let $t_c = \frac{t_a+t_b}{2}$ and use a Newton algorithm that repeats the following steps:

IF $t_b - t_a > 2\tau$ and $|h(t_c)| > \tau|h'(t_c)|$ **THEN**
(a) if $h(t_c) < 0$, set $(t_a, h(t_a)) = (t_c, h(t_c))$;
otherwise set $(t_b, h(t_b)) = (t_c, h(t_c))$;
(b) set $t_c = t_c - \frac{h(t_c)}{h'(t_c)}$;
ELSE stop and output $t_{1-\alpha} \approx \frac{t_a+t_b}{2}$ or $t_{1-\alpha} \approx t_c$.

Genz and Bretz (2000) also investigated the use of Secant-like methods for solving $h(t) = 0$ but believed that the simple Secant method is not suitable for many problems because $|h'(t)|$ is sometimes very small near $t_{1-\alpha}$ (particularly when α is small), and this can result in divergence unless a very good starting value is available. Various bisection-Secant hybrid methods were considered and after some experiments, the Pegasus method (Ralston and Rabinowitz, 1978) was selected. This method has asymptotic order of convergence similar to that of the Secant method and at each iteration provides a bracketing interval for $t_{1-\alpha}$. The Pegasus method that was finally implemented (starting with τ and $[t_a, t_b]$) initially sets $t_c = \frac{t_a+t_b}{2}$. If $h(t_c) < 0$, then $(t_a, h(t_a)) = (t_c, h(t_c))$; otherwise, $(t_b, h(t_b)) = (t_c, h(t_c))$. The basic iteration repeats the following steps:

IF $t_b - t_a > 2\tau$ and $|h(t_c)| > \tau \left|\frac{h(t_b)-h(t_a)}{t_b-t_a}\right|$ **THEN**
(a) compute $t_c = t_b - \frac{h(t_b)(t_b-t_a)}{h(t_b)-h(t_a)}$;
(b) if $h(t_c) < 0$, set $(t_a, h(t_a)) = (t_c, h(t_c))$;
otherwise set $h(t_a) = \frac{h(t_a)h(t_b)}{h(t_b)+h(t_c)}$ and $(t_b, h(t_b)) = (t_c, h(t_c))$
ELSE stop and output $t_{1-\alpha} \approx \frac{t_a+t_b}{2}$ or $t_{1-\alpha} \approx t_c$.

The Pegasus method is the same as the linearly convergent False-Position method except for the h_a modification at step (b), which improves the approximate order of convergence to 1.64.

Error Control

When numerical integration is used to evaluate $h(t)$, what is actually computed is $\hat{h}(t) = h(t) + \epsilon_I$, where ϵ_I is the numerical integration error. This error can, in principle, be made arbitrarily small, but at the expense of more work (computer time). Let $\hat{t} = t_{1-\alpha} + \epsilon_t$ be an approximation to $t_{1-\alpha}$ with error ϵ_t. What is actually computed at each step in a combined numerical optimization-integration algorithm is

$$\hat{h}(\hat{t}) = h(\hat{t}) + \epsilon_I \approx \epsilon_t h'(t_{1-\alpha}) + \epsilon_I, \tag{6.2}$$

for sufficiently small $|\epsilon_t|$. In order to balance the numerical integration and optimization errors, an estimate for $|h'(t_{1-\alpha})|$ is needed. This quantity does

not need to be very accurate, so a simple difference-quotient approximation, in the form

$$h'(t_{1-\alpha}) \approx \hat{h}' = \frac{\hat{h}(t_b) - \hat{h}(t_a)}{t_b - t_a},$$

can be determined using quantities from the Pegasus method iterations. If the Newton method is used, then an accurate value of $h'(t_{1-\alpha})$ is already available.

Given $\hat{h}'$ and τ (the desired error tolerance for $t_{1-\alpha}$), the numerical integration error tolerance must be set at a level at least as small as $\hat{h}'\tau$. Otherwise the numerical integration errors may dominate the total error in $\hat{h}(\hat{t})$ and it will be impossible to reliably determine when the optimization method has converged. The strategy used for the example test results reported below was to set the error tolerance for the numerical integration at $\tau|\hat{h}'|$.

Examples

The following examples illustrate the use of the numerical optimization-integration algorithms described above. Tables 6.1 and 6.2 summarize the numerical results. Total work is measured in terms of the number of density function evaluations needed for the integration to compute $P(t)$. The basic starting interval $[t_a, t_b]$ was computed using the Pegasus method; the cost is minimal because only univariate and bivariate distribution function values are required. The initial bracketing interval for $t_{1-\alpha}$ is given in the first row for each of the problems considered in Tables 6.1 and 6.2. The work reported in that row is the work to compute the initial $h(t_a)$, $h(t_c)$ and $h(t_b)$ values (with only $h(t_c)$ required to start the Newton iteration). Subsequent rows show results of the iterations necessary to determine $t_{1-\alpha}$ to the specified accuracy

The first example is based on data taken from Hsu and Nelson (1998). We calculate two-sided confidence intervals for six treatment-control comparisons with $P(t) = T_6(-\mathbf{t}, \mathbf{t}; \mathbf{\Sigma}, 86)$ and

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & - & - & - & - & - \\ 0.3958 & 1 & - & - & - & - \\ 0.5677 & 0.4936 & 1 & - & - & - \\ 0.5468 & 0.4621 & 0.7598 & 1 & - & - \\ 0.5140 & 0.4488 & 0.7675 & 0.6930 & 1 & - \\ 0.5505 & 0.4922 & 0.8651 & 0.7738 & 0.7915 & 1 \end{bmatrix}.$$

For $\alpha = 0.1$, we obtain $[t_{1,a}, t_{1,b}] = [1.663, 2.442]$ and $[t_{2,a}, t_{2,b}] = [2.116, 2.324]$. Similarly, for $\alpha = 0.05$, we compute $[t_{1,a}, t_{1,b}] = [1.988, 2.701]$ and $[t_{2,a}, t_{2,b}] = [2.429, 2.606]$. The numerical optimization-integration results are summarized in Table 6.1, where $\tau = 0.001$ throughout.

The second example involves all pairwise comparisons of four treatments based on an one-factorial ANOVA model with sample sizes $n_i = 20, 3, 3, 15$ (Westfall et al, 1999). We calculate two-sided confidence intervals with $P(t) = T_6(-\mathbf{t}, \mathbf{t}; \mathbf{\Sigma}, 37)$, given

Table 6.1. Numerical optimization-integration results for the first example

Iteration	t_a, t_c, t_b	$h(t_a), h(t_c), h(t_b)$	$\hat{h'}$	Work
	Pegasus method, $\alpha = 0.1$			
#1	2.220, 2.264, 2.324	–0.01, 0.00050, 0.0100	0.22	39952
#2	2.220, 2.262, 2.264	–0.01, 0.00001, 0.0005	0.22	51152
	Newton method, $\alpha = 0.1$			
#1	2.116, 2.220, 2.324	– , –0.01, –	0.26	11200
#2	2.220, 2.261, 2.324	–0.01, –0.0002, –	0.24	28752
	Pegasus method, $\alpha = 0.05$			
#1	2.518, 2.561, 2.606	–0.005, 0.00020, 0.0060	0.12	146432
#2	2.518, 2.559, 2.561	–0.005, 0.00004, 0.0002	0.12	187984
	Newton method, $\alpha = 0.05$			
#1	2.429, 2.518, 2.606	– , –0.005, –	0.14	41552
#2	2.518, 2.557, 2.606	–0.005, –0.0002, –	0.13	104880
#3	2.557, 2.559, 2.606	–0.0002, 0.00001, –	0.12	168208

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & - & - & - & - & - \\ 0.1304 & 1 & - & - & - & - \\ 0.2364 & 0.2364 & 1 & - & - & - \\ -0.6594 & 0.6594 & 0 & 1 & - & - \\ -0.8513 & 0 & 0.3086 & 0.6455 & 1 & - \\ 0 & -.8513 & 0.3086 & -0.6455 & 0.1667 & 1 \end{bmatrix}.$$

For $\alpha = 0.1$, we obtain $[t_{1,a}, t_{1,b}] = [1.687, 2.508]$ and $[t_{2,a}, t_{2,b}] = [2.265, 2.383]$. Similarly, for $\alpha = 0.05$, we compute $[t_{1,a}, t_{1,b}] = [2.050, 2.788]$ and $[t_{2,a}, t_{2,b}] = [2.026, 2.684]$. The numerical optimization-integration results are summarized in Table 6.2, again with $\tau = 0.001$.

The results from these example tests demonstrate that the algorithms described here provide feasible methods for computing $t_{1-\alpha}$ values for confidence intervals. Thus, efficient numerical optimization-integration methods are available which only require a few iteration steps to achieve an accurate estimate of $t_{1-\alpha}$ for a given error tolerance τ. Given good starting intervals determined from bivariate distribution values, the numerical optimization based on the use of the Newton method seems to be more efficient than optimization based on the Pegasus method.

6.2 Bayesian Statistics and Finance Applications

Many integrals that arise in Bayesian statistical analysis applications (Evans and Swartz, 1995; Shaw, 1988) have the general form

Table 6.2. Numerical optimization-integration results for the second example

Iteration	t_a, t_c, t_b	$h(t_a), h(t_c), h(t_b)$	$\hat{h}'$	Work
Pegasus method, $\alpha = 0.1$				
#1	2.324, 2.339, 2.383	–0.003, 0.0003, 0.0090	0.21	86144
#2	2.324, 2.337, 2.339	–0.003, –0.0002, 0.0003	0.21	127696
Newton method, $\alpha = 0.1$				
#1	2.265, 2.324, 2.383	– , –0.003, –	0.25	41552
#2	2.324, 2.338, 2.383	–0.003, –0.00006, –	0.21	68592
Pegasus method, $\alpha = 0.05$				
#1	2.648, 2.654, 2.684	–0.0006, –0.00001, 0.003	0.11	255680
Newton method, $\alpha = 0.05$				
#1	2.612, 2.648, 2.684	– , –0.0006, –	0.13	63328
#2	2.648, 2.654, 2.684	–0.0006, 0.00003, –	0.12	159504

$$I(g) = \int_{-\infty}^{\infty} p(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{6.3}$$

where $\boldsymbol{\theta}$ is a k-dimensional vector of integration variables, $p(\boldsymbol{\theta})$ is a posterior density (usually not normalized), and $g(\boldsymbol{\theta})$ is some elementary real-valued function (for example, $g(\boldsymbol{\theta}) = \theta_j$, a component of $\boldsymbol{\theta}$). If the posterior $p(\boldsymbol{\theta})$ is characterized by the presence of a dominant peak, then the integral can often be efficiently computed by combining a multivariate normal or multivariate t approximation to $p(\boldsymbol{\theta})$ together with a good numerical integration method.

Many computational finance problems require the computation of integrals in the form

$$I = \int_{-\infty}^{\infty} \phi_k(\mathbf{x}; \mathbf{I}_k) q(\mathbf{x}) d\mathbf{x},$$

see Hull (1993), Jäckel (2002), and Glasserman (2004). These problems are similar in form to the (6.3) problems, with $g \equiv 1$ and a peaked posterior density $p(\boldsymbol{\theta}) = \phi_k(\boldsymbol{\theta}; \mathbf{I}_k) q(\boldsymbol{\theta})$, dominated by a simple product MVN density (with $\boldsymbol{\Sigma} = \mathbf{I}_k$ and $\boldsymbol{\delta} = \mathbf{0}$), but the location of the peak for $p(\boldsymbol{\theta})$ is often shifted away from the $\boldsymbol{\theta} = \mathbf{0}$ position for the MVN density. Many of these problems have very large values for k, which might count a number of time periods (for example, days in a year, or months in the life of a mortgage). So, methods which use the location of the peak and information about the behavior of $p(\boldsymbol{\theta})$ near the peak can often lead to more efficient computation.

In this section we discuss some of the details for methods which construct a multivariate normal or multivariate t approximation to $p(\boldsymbol{\theta})$ and use this approximation with an appropriate numerical integration method, focusing on

the Bayesian statistical analysis problems. We finish with several illustrations of these methods, including several computational finance problems.

6.2.1 The Standardizing Transformation

Below, we follow the conventions used in computational finance and refer to the non-centrality parameter $\boldsymbol{\delta}$ as the "mean" and denote it by $\boldsymbol{\mu}$. We begin with a standardizing transformation of the posterior density in the form $\boldsymbol{\theta} = \boldsymbol{\mu} + \mathbf{C}\mathbf{y}$, where $\boldsymbol{\mu}$ is the location of the peak for $p(\boldsymbol{\theta})$ (the *mode*), and $\mathbf{C}$ is the lower triangular Cholesky factor for what is often called the *modal covariance matrix* $\boldsymbol{\Sigma}$ (with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^t$). The modal covariance matrix $\boldsymbol{\Sigma}$ is defined as the inverse of the negative of the Hessian matrix for $\log(p(\boldsymbol{\theta}))$ at $\boldsymbol{\mu}$. Note that if $p(\boldsymbol{\theta})$ is a multivariate normal density with mean $\boldsymbol{\mu}$, then the modal covariance matrix is the covariance matrix for the multivariate normal density. We also note that any $k \times k$ matrix $\mathbf{C}$ satisfying $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^t$ could be used for the standardizing transformation, but the Cholesky $\mathbf{C}$ is often chosen for computational convenience.

The mode $\boldsymbol{\mu}$ can be determined using a numerical method to determine the point $\boldsymbol{\mu}$ where the maximum of the function $F(\boldsymbol{\theta}) = \log(p(\boldsymbol{\theta}))$ is attained. Then the inverse of the modal covariance matrix $\boldsymbol{\Sigma}$ can be approximated using finite difference approximations to the partial derivatives of $F(\boldsymbol{\theta})$. For example, if $\mathbf{H}$ is used to denote the Hessian matrix at $\boldsymbol{\mu}$, the following standard finite difference approximations to the partial derivatives could be used:

$$\begin{aligned}\mathbf{H}_{ij} \approx \frac{1}{h^2}\big(&F(\boldsymbol{\theta} + h(\mathbf{e}_i + \mathbf{e}_j)/2) + F(\boldsymbol{\theta} - h(\mathbf{e}_i + \mathbf{e}_j)/2) \\ &-F(\boldsymbol{\theta} + h(\mathbf{e}_i - \mathbf{e}_j)/2) - F(\boldsymbol{\theta} - h(\mathbf{e}_i - \mathbf{e}_j)/2)\big),\end{aligned}$$

when $i \neq j$, otherwise

$$\mathbf{H}_{ii} \approx \frac{F(\boldsymbol{\theta} + h\mathbf{e}_i) + F(\boldsymbol{\theta} - h\mathbf{e}_i) - 2F(\boldsymbol{\theta})}{h^2}.$$

Here $\mathbf{e}_i$ is the i-th standard Euclidean basis vector with a 1 for the i-th component and 0 otherwise, and h is a step size chosen to provide moderately accurate derivative approximations but avoiding large subtractive cancellation errors that often result from these approximations (for example, $h = 0.001$). Then the standardizing Cholesky factor $\mathbf{C}$ is determined by $\mathbf{C}\mathbf{C}^t = -\mathbf{H}^{-1}$. After applying the standardizing transformation $\boldsymbol{\theta} = \boldsymbol{\mu} + \mathbf{C}\mathbf{y}$, we have

$$\begin{aligned}I(g) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= |\mathbf{C}| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}) p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}) d\mathbf{y}.\end{aligned}$$

6.2.2 The Multivariate Normal Transformation

If we assume that the posterior density is approximately multivariate normal, then after the standardizing transformation, we multiply and divide by the MVN density so that

$$I(g) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_k(\mathbf{y}; \mathbf{I}_k) g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}) h(\mathbf{y}) d\mathbf{y}, \tag{6.4}$$

with

$$h(\mathbf{y}) = \frac{|\mathbf{C}| p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y})}{\phi_k(\mathbf{y}; \mathbf{I}_k)}.$$

If the MVN approximation is good, then $h(\mathbf{y}) \approx 1$ and the $I(g)$ integrals should be easy to estimate, as long as g is not too complicated.

A final set of univariate normal transformations of the form $y_i = \Phi^{-1}(z_i)$ can be used to put $I(g)$ into a form suitable for standard multidimensional integration methods. Then, $I(g)$ becomes

$$I(g) = \int_0^1 \cdots \int_0^1 g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}(\mathbf{z})) h(\mathbf{y}(\mathbf{z})) d\mathbf{z},$$

where $\mathbf{y}(\mathbf{z}) = (\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_k))^t$.

6.2.3 The Multivariate t Transformation

A more general unimodal mode, that can be used for the posterior densities with thicker tails than those for the MVN density, is the MVT density from (1.2),

$$t_k(\mathbf{x}; \boldsymbol{\Sigma}, \nu) = K_\nu^{(k)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left(1 + \frac{\mathbf{x}^t \boldsymbol{\Sigma}^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+k}{2}},$$

where $K_\nu^{(k)}$ is defined in equation (4.7), see also Evans and Swartz (1995, 2000). With this model, the modal covariance matrix ($-\mathbf{H}^{-1}$ at $\mathbf{x} = \mathbf{0}$) and the actual covariance matrix $\boldsymbol{\Sigma}$ for the model differs by a factor of $(\nu + k)/\nu$, so that the correct standardizing transformation matrix is $\mathbf{C} = \sqrt{(\nu + k)/\nu}\hat{\mathbf{C}}$, where $\hat{\mathbf{C}}\hat{\mathbf{C}}^t = -\mathbf{H}^{-1}$.

An approximate ν can be determined by comparing the behavior of the log posterior ratio near $\boldsymbol{\mu}$, $\log\left(p(\boldsymbol{\mu} + \sqrt{\frac{\nu+k}{\nu}}\hat{\mathbf{C}}\mathbf{y})/p(\boldsymbol{\mu})\right)$, with the log of the MVT model, $-\frac{\nu+k}{2}\log\left(1 + \frac{\mathbf{y}^t\mathbf{y}}{\nu}\right)$, for selected $\mathbf{y}$ vectors. If we define an average log posterior ratio near $\boldsymbol{\mu}$ by

$$S(\delta, \nu) = \frac{1}{2k} \sum_{i=1}^{n} \left(\log\left(\frac{p(\boldsymbol{\mu} + \delta\sqrt{\frac{\nu+k}{\nu}}\hat{\mathbf{C}}\mathbf{e}_i)}{p(\boldsymbol{\mu})} \right) + \log\left(\frac{p(\boldsymbol{\mu} - \delta\sqrt{\frac{\nu+k}{\nu}}\hat{\mathbf{C}}\mathbf{e}_i)}{p(\boldsymbol{\mu})} \right) \right),$$

where $\mathbf{e}_i$ denotes the i-th column of the $k \times k$ identity matrix, then a simple heuristic strategy is to choose the ν value (ν_{min}) which minimizes the absolute difference between $S(\delta, \nu)$ and $-\frac{\nu+k}{2} \log\left(1+\frac{\delta^2}{\nu}\right)$ for selected values of δ. Genz and Kass (1997) minimized the function

$$\left| S(1,\nu) + \frac{\nu+k}{2} \log\left(1+\frac{1}{\nu}\right) \right| + \left| S(2,\nu) + \frac{\nu+k}{2} \log\left(1+\frac{4}{\nu}\right) \right|,$$

for $\nu \in [1, \nu_{max}]$, given some heuristic threshold ν_{max} (for example, $\nu_{max} = 15$). If $\nu_{min} = \nu_{max}$, then an MVN model is selected.

Once a good ν has been chosen, and if an MVT model is selected, the transformation $\boldsymbol{\theta} = \boldsymbol{\mu} + \mathbf{C}\mathbf{y}$ produces $I(g)$ in the form

$$I(g) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_k(\mathbf{y}; I, \nu) g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}) h(\mathbf{y}) d\mathbf{y}, \tag{6.5}$$

with

$$h(\mathbf{y}) = \frac{|\mathbf{C}||p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y})}{t_k(\mathbf{y}; I, \nu)}.$$

At this stage, the transformation $\mathbf{y} = \mathbf{w}\sqrt{\nu}/s$ (which introduces the additional s variable, see equations (1.2) and (1.3)) could be used to express $I(g)$ as a weighted MVN integral

$$I(g) = \frac{2^{1-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^{\infty} s^{\nu-1} e^{-\frac{s^2}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_k(\mathbf{w}; \mathbf{I}_k) h\left(\frac{\mathbf{w}\sqrt{\nu}}{s}\right) g\left(\boldsymbol{\mu} + \mathbf{C}\frac{\mathbf{w}\sqrt{\nu}}{s}\right) d\mathbf{w} ds,$$

In this case, a final transformation to $\mathbf{z} \in [0,1]^{k+1}$ combines $s = \chi_\nu^{-1}(z_{k+1})$, and $\mathbf{w}(\mathbf{z}) = (\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_k))^t$ for computation with standard numerical integration methods for integration over $[0,1]^{k+1}$, so that $I(g)$ is given by

$$I(g) = \int_0^1 \int_0^1 \cdots \int_0^1 h\left(\frac{\Phi^{-1}(\mathbf{z})\sqrt{\nu}}{\chi_\nu^{-1}(z_{k+1})}\right) g\left(\boldsymbol{\mu} + \mathbf{C}\frac{\Phi^{-1}(\mathbf{z})\sqrt{\nu}}{\chi_\nu^{-1}(z_{k+1})}\right) d\mathbf{z}.$$

Alternatively, a final transformation to $\mathbf{z} \in [0,1]^k$, using inverse univariate T transformations, starts with the transformations (4.8) so that $I(g)$ becomes

$$I(g) = \int_{-\infty}^{\infty} t(u_1; \nu) \int_{-\infty}^{\infty} t(u_k; \nu + k - 1) g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}(\mathbf{u})) h(\mathbf{y}(\mathbf{u})) d\mathbf{u}.$$

The final transformations to $[0,1]$ variables z_i uses the univariate inverse T transformations $u_i = T^{-1}(z_i; \nu + i - 1)$ for $i = 1, 2, \ldots, k$, with the final form for $I(g)$ given by

$$I(g) = \int_0^1 \int_0^1 \cdots \int_0^1 g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}(\mathbf{u}(\mathbf{z})))h(\mathbf{y}(\mathbf{u}(\mathbf{z})))d\mathbf{z},$$

where $\mathbf{u}(\mathbf{z}) = (T^{-1}(z_1; \nu), \ldots, T^{-1}(z_k; \nu + k - 1))^t$.

6.2.4 The Split-t Transformations

Some problems have skewed tail behavior, instead of the symmetric non-normal tail behavior which can be handled with the MVT transformation discussed in the previous section. For these skewed problems, a set of "split" transformations has been studied for use after the standardizing transformation $\boldsymbol{\theta} = \boldsymbol{\mu} + \mathbf{C}\mathbf{y}$. With the split-$t$ transformations, the univariate normal transformations $z_i = \Phi(y_i)$ described in Section 6.2.2 have to be modified. These univariate transformations are split at the origin and replaced by (possibly) different transformation functions, chosen from the MVT family, for each of the $\mathbf{y}$ variables. With this method, the function $p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y})$ is investigated along both the positive and negative directions for each of the $\mathbf{y}$ variables to determine the appropriate degrees of freedom $\nu_i^{\pm}$ and scale factors $\delta_i^{\pm}$, for $i = 1, 2, \ldots, k$. The split transformations take the form

$$y_i = S_i^{-1}(z_i) = \begin{cases} \delta_i^{\pm} T^{-1}(z_i; \nu_i^{\pm}), \text{ if } \pm(z - 0.5) > 0 \text{ and } \nu_i^{\pm} < \nu_{max} \\ \delta_i^{\pm}, \Phi^{-1}(z_i), \quad \text{if } \pm(z - 0.5) > 0 \text{ and } \nu_i^{\pm} \geq \nu_{max}, \end{cases}$$

with all $z_i \in [0,1]$ and ν_{max} set to some heuristic threshold value (for example, $\nu_{max} = 10$). A numerical method for the determination of good values for the δ and ν parameters based on evaluations of $p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y})$ is described in detail by Genz and Kass (1997).

After the transformations, the integral over the unit k-cube is then

$$I(g) = \int_0^1 \cdots \int_0^1 g(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}(\mathbf{z}))h(\mathbf{y}(\mathbf{z}))d\mathbf{z},$$

where $\mathbf{y}(\mathbf{z}) = (S_1^{-1}(z_1), \ldots, S_m^{-1}(z_k))^t$ and

$$h(\mathbf{y}(\mathbf{z})) = |\mathbf{C}| p(\boldsymbol{\mu} + \mathbf{C}\mathbf{y}(\mathbf{z})) \prod_{i=1}^{k} J_i(y_i),$$

with the transformation Jacobian factors J_i defined by

$$J_i(y) = \begin{cases} \delta_i^{\pm} \left(K^{(1)}_{\nu_i^{\pm}}\right)^{-1} \left(1 + \frac{y^2}{\nu_i^{\pm}(\delta_i^{\pm})^2}\right)^{\frac{\nu_i^{\pm}+1}{2}}, & \text{if } \pm y_i > 0 \text{ and } \nu_i^{\pm} < \nu_{max}, \\ \delta_i^{\pm}\sqrt{2\pi} e^{\frac{y^2}{2(\delta_i^{\pm})^2}}, & \text{if } \pm y_i > 0 \text{ and } \nu_i^{\pm} = \nu_{max}. \end{cases}$$

If the split-t transformations are used with Monte Carlo integration, then this is equivalent to using the split-t distributions as importance functions for Monte Carlo integration (Geweke, 1989, 1991b; Evans, 1991).

6.2.5 Some Examples

Stanford Heart Transplant Data Example

This example has been considered in a number of papers; see, for example, Naylor and Smith (1982) and Genz and Kass (1997). The integrals of interest have the form

$$I(g) = \int_0^\infty \int_0^\infty \int_0^\infty g(\lambda, \tau, \rho) p(\lambda, \tau, \rho) d\lambda d\tau d\rho,$$

with

$$p(\lambda, \tau, \rho) = \prod_{i=1}^{n} \frac{\rho \lambda^\rho}{(\lambda + x_i)^{p+1}} \prod_{i=n+1}^{N} \left(\frac{\lambda}{\lambda + x_i}\right)^\rho \prod_{j=1}^{m} \frac{\tau \rho \lambda^\rho}{(\lambda + y_j + \tau z_j)^{\rho+1}} \prod_{j=m+1}^{M} \left(\frac{\lambda}{\lambda + y_j + \tau z_j}\right)^\rho,$$

where we use the Naylor and Smith (1982) data values for $(x, y, z)_i$. Approximate expected values for (λ, τ, ρ) were needed, so a computation which uses the vector $(1, \lambda, \tau, \rho)$ provides approximations to $(\bar{\lambda}, \bar{\tau}, \bar{\rho}) = I(\lambda, \tau, \rho)/I(1)$.

Following Naylor and Smith (1982), we first make the transformation $(\lambda, \tau, \rho) = (e^{\theta_1}, e^{\theta_2}, e^{\theta_3})$, so that the integrals now have the form

$$I(g) = \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty g\left(e^{\theta_1}, e^{\theta_2}, e^{\theta_3}\right) p\left(e^{\theta_1}, e^{\theta_2}, e^{\theta_3}\right) e^{\theta_1+\theta_2+\theta_3} d\boldsymbol{\theta}.$$

Then, numerical optimization is applied to the function

$$q(\boldsymbol{\theta}) = \log\left(p\left(e^{\theta_1}, e^{\theta_2}, e^{\theta_3}\right)\right) + \theta_1 + \theta_2 + \theta_3$$

to obtain the mode $\boldsymbol{\mu}$ and modal covariance matrix $\boldsymbol{\Sigma}$, with Cholesky factor $\mathbf{C}$. The sum $\theta_1 + \theta_2 + \theta_3$, which arises from the Jacobian for the transformation to the $\boldsymbol{\theta}$ variables, is included in the objective function during

the optimization, because the transformation selection and integration algorithms also use this sum. Applying numerical optimization, we then obtain $\boldsymbol{\mu} \approx (3.3850, -0.092401, -0.72291)$ with (the finite difference approximated)

$$\boldsymbol{\Sigma} \approx \begin{bmatrix} 0.21468 & -0.0092513 & 0.093005 \\ -0.0092513 & 0.17276 & -0.49946 \\ 0.093005 & -0.049946 & 0.06893 \end{bmatrix}$$

and the resulting

$$\mathbf{C} = \begin{bmatrix} 0.46333 & 0 & 0 \\ -0.19967 & 0.41516 & 0 \\ 0.20073 & -0.11065 & 0.12804 \end{bmatrix}.$$

The maximum value of q is approximately -375, so the constant 375 is added to $\log(p)$ during the integral computations to avoid underflow problems when computing p.

Table 6.3. Heart Transplant Results

	Expected Values			
p's	$\bar{1}$	$\bar{\lambda}$	$\bar{\tau}$	$\bar{\rho}$
2025	0.2961	32.556	1.0406	0.4976
4365	0.2963	32.585	1.0421	0.4976
9135	0.2964	32.597	1.0441	0.4975
Err. Est.	0.0005	0.037	0.0033	0.0007

The results in Table 6.3 were computed using adaptive integration software (Berntsen et al, 1991b) applied to the transformed integral in the form (6.4). The errors given are estimated errors from the adaptive integration method. Similar results were also obtained using split transformations because the transformation selection algorithm selected a normal/normal split for all three variables with the scale factors $\delta^{\pm} \simeq 1$.

Photocarcinogenicity Data Example

This example was discussed Dellaportas and Smith (1993). The integrals of interest have the form

$$I(g) = \int_0^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\boldsymbol{\theta}, \rho) p(\boldsymbol{\theta}, \rho) d\boldsymbol{\theta} d\rho,$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^t$,

$$p(\boldsymbol{\theta}, \rho) = \left(\prod_{i=1}^{n} \rho t_i^{\rho - 1} e^{\mathbf{z}_i \boldsymbol{\theta}} \right) \left(\prod_{i=1}^{n+m} e^{-t_i^{\rho} \mathbf{z}_i \boldsymbol{\theta}} \right),$$

$n = 65$, $m = 15$, and data values given in the $\mathbf{t}$ vector and the 80×4 matrix $\mathbf{Z}$ with rows $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_{80}$. Approximate expected values for $(\boldsymbol{\theta}, \rho)$ were needed, so a computation which uses the vector $(1, \boldsymbol{\theta}, \rho)$ provides approximations to $(\bar{\boldsymbol{\theta}}, \bar{\rho}) = I(\boldsymbol{\theta}, \rho)/I(1)$. The transformation $\rho = e^{\theta_5}$ produces integrals in the form

$$I(g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\boldsymbol{\theta}, e^{\theta_5} p(\boldsymbol{\theta}, e^{\theta_5}) e^{\theta_5} d\boldsymbol{\theta} d\theta_5.$$

Numerical optimization is applied to the function

$$q(\boldsymbol{\theta}, \theta_5) = \log\left(p\left(\boldsymbol{\theta}, e^{\theta_5}\right)\right) + \theta_5$$

to obtain the mode $\boldsymbol{\mu}$ and modal covariance matrix $\boldsymbol{\Sigma}$, with Cholesky factor $\mathbf{C}$. For this example, we have $\boldsymbol{\mu} \approx (-10.847, -1.1791, -0.35168, 0.40198, 1.1903)^t$,

$$\boldsymbol{\Sigma} \approx \begin{bmatrix} 1.3442 & 0.015174 & -0.023852 & -0.12772 & -0.11510 \\ 0.015174 & 0.13312 & 0.054143 & 0.048690 & -0.0060429 \\ -0.023852 & 0.054143 & 0.11577 & 0.050959 & -0.0025648 \\ -0.12772 & 0.048690 & 0.050959 & 0.11582 & 0.0066915 \\ -0.11510 & -0.0060429 & -0.0025648 & 0.0066915 & 0.010258 \end{bmatrix},$$

and

$$\mathbf{C} = \begin{bmatrix} 1.1594 & 0 & 0 & 0 & 0 \\ 0.013088 & 0.36461 & 0 & 0 & 0 \\ -0.020573 & 0.14923 & 0.30509 & 0 & 0 \\ -0.11016 & 0.13749 & 0.092348 & 0.27614 & 0 \\ -0.099279 & -0.013010 & -0.008738 & -0.0059725 & 0.010972 \end{bmatrix}.$$

The minimum value of $-q$ is approximately 243.6, so the constant 243.6 is added to $\log(p)$ during the integral computations to avoid underflow problems when computing p.

Table 6.4. Photocarcinogenicity Results

	Expected Values					
p's	$\bar{1}$	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	$\bar{\rho}$
10000	0.038613	-10.862	-1.1940	-0.35565	0.39772	3.2852
20000	0.038927	-10.873	-1.1926	-0.35936	0.39676	3.2887
40000	0.039068	-10.891	-1.1936	-0.36597	0.39300	3.2940
Err. Est.	0.0008	0.3	0.03	0.02	0.008	0.08

The results in Table 6.4 were computed using randomized lattice rules (4.16) for integrals in the form (6.4). The errors given are estimated errors (using 3 times the standard errors) for the last expected value row, using 10 randomly shifted lattice rules. Similar results were also obtained using split transformations because the transformation selection algorithm selected a normal/normal split for all five variables with the scale factors $\delta^{\pm} \approx 1$.

Asian Option Example

A set of integrals that are needed for pricing an Asian financial option takes the form

$$I_k(S_0,T,K,r,s) = \int_{-\infty}^{\infty} \phi_k(\mathbf{x};\mathbf{I}_k) e^{-rT} \left(\frac{S_0}{k} \sum_{i=1}^{k} e^{\left(r-\frac{s^2}{2}\right) i\Delta + s\sqrt{\Delta}\sum_{j=1}^{i} x_j} - K \right)^+ d\mathbf{x},$$

with $\Delta = T/k$, and where $(Y)^+ \equiv \max(Y,0)$, see Hull (1993). These integrals approximate an expected payoff integral over an interval $[0,T]$ which is discretized with step size Δ, so larger k values provide more accurate approximations. For the purpose of illustrative methods in this section, we consider the approximate computation of $I_{16}(50,1,50,0.05,0.1)$ and $I_{64}(50,1,50,0.05,0.1)$.

Numerical optimization is applied to the function

$$q(\mathbf{x}) = -\frac{\mathbf{x}^t\mathbf{x}}{2} - rT + \log\left(\frac{S_0}{k} \sum_{i=1}^{k} e^{\left(r-\frac{s^2}{2}\right) i\Delta + s\sqrt{\Delta}\sum_{j=1}^{i} x_j} \right)$$

to obtain the mode $\boldsymbol{\mu}$ and modal covariance matrix $\boldsymbol{\Sigma}$, with Cholesky factor $\mathbf{C}$. For $k = 16$, numerical optimization produces an approximate

$$\boldsymbol{\mu} = (0.34413, 0.32397, 0.30351, 0.28294, 0.26214, 0.24118, 0.2199, 0.19855, 0.177, 0.15534, 0.13347, 0.1115, 0.089416, 0.067186, 0.044829, 0.022458)^t$$

with $\mathbf{C}$ approximately equal to the identity matrix. For $k = 64$, numerical optimization produces an approximate

$$\boldsymbol{\mu} = (0.17809, 0.17551, 0.1729, 0.17027, 0.16767, 0.16505, 0.1624, 0.15976, \ldots, 0.020351, 0.017451, 0.01455, 0.01164, 0.0087575, 0.0058308, 0.0029156)^t$$

with $\mathbf{C}$ approximately equal to the identity matrix.

Table 6.5. Some Asian Option Results

	Expected Values	
p's	$I_{16}(50,1,50,0.05,0.1)$	$I_{64}(50,1,50,0.05,0.1)$
10000	1.9181	1.8439
20000	1.9192	1.8473
40000	1.9190	1.8445
Err. Est.	0.003	0.004

The results in Table 6.5 were computed using randomized lattice rules (4.16) for integrals in the form (6.4). The errors given are estimated errors for the last expected value row, using 10 randomly shifted lattice rules. Approximate I_{16} and I_{32} values were also computed with $\boldsymbol{\mu} = \mathbf{0}$, but the results were similar.

Mortgage Backed Securities Computation Example

Another integration problem is given by

$$I_k(C, Q_0, \theta) = \int_{-\infty}^{\infty} \phi_k(\mathbf{x}; \mathbf{I}_k) \left(C \sum_{i=1}^{k} \prod_{j=1}^{i} \frac{1}{1 + Q_0 e^{\theta(\sum_{l=1}^{j} x_l - j\frac{\theta}{2})}} \right) d\mathbf{x},$$

see Fishman (2006, pp. 62–63). These integrals approximate the present value of mortgage acked security assuming k monthly payments of an amount C. The interest rate starts with value Q_0, but changes randomly following a lognormal distribution with volatility parameter θ. For the purpose of illustrative methods in this section, we consider the approximate computation of $I_{120}(1000, 0.07, 0.02)$ ($k = 120$ for 10 year mortgages).

Numerical optimization is applied to the function

$$q(\mathbf{x}) = -\frac{\mathbf{x}^t\mathbf{x}}{2} + \log \left(C \sum_{i=1}^{k} \prod_{j=1}^{i} \frac{1}{1 + Q_0 e^{\theta(\sum_{l=1}^{j} x_l - j\frac{\theta}{2})}} \right)$$

to obtain the mode $\boldsymbol{\mu}$ and modal covariance matrix $\boldsymbol{\Sigma}$, with Cholesky factor $\mathbf{C}$. Numerical optimization provided an approximate

$$\boldsymbol{\mu} = (-0.006176, -0.006043, -0.005918, -0.005818, \ldots, -0.000018)^t$$

with $\mathbf{C}$ approximately equal to the identity matrix.

Table 6.6. Some Mortgage Backed Security Results

	Expected Values	
p's	$I_{120}(1000, .07, .02)$	$\hat{I}_{120}(1000, 0.07, 0.02)$
10000	86196.74	86198.2
20000	86196.76	86196.7
40000	86196.77	86197.5
Err. Est.	0.06	1.5

The results in Table 6.6 were computed using randomized lattice rules (4.16) for integrals in the form (6.4). The errors given are estimated errors for the last expected value row, using 10 randomly shifted lattice rules. The $\hat{I}_{120}$ results are for computations done with $\boldsymbol{\mu} = \mathbf{0}$, and for this example the $\boldsymbol{\mu} \neq \mathbf{0}$ results were clearly better, with significantly smaller standard errors.

A

Description of the R Functions

In this appendix we reproduce the relevant online documentation for the key functions available in the `R` package `mvtnorm`. The complete documentation is available with the library, which can be downloaded from the Comprehensive R Archive Network (CRAN) at www.r-project.org.

`mvnorm`

Description

These functions provide the density function and a random number generator for the multivariate normal distribution with mean equal to `mean` and covariance matrix `sigma`.

Usage

```
dmvnorm(x, mean, sigma, log=FALSE)
rmvnorm(n, mean = rep(0, nrow(sigma)),
        sigma = diag(length(mean)),
        method=c("eigen", "svd", "chol"))
```

Arguments

`x`	Vector or matrix of quantiles. If `x` is a matrix, each row is taken to be a quantile.
`n`	Number of observations.
`mean`	Mean vector, default is `rep(0, length = ncol(x))`.
`sigma`	Covariance matrix, default is `diag(ncol(x))`.
`log`	Logical; if `TRUE`, densities d are given as $\log(d)$.

`method` Matrix decomposition used to determine the matrix root of `sigma`, possible methods are eigenvalue decomposition (`"eigen"`, default), singular value decomposition (`"svd"`), and Cholesky decomposition (`"chol"`).

Examples

```
dmvnorm(x=c(0,0))
dmvnorm(x=c(0,0), mean=c(1,1))

sigma <- matrix(c(4,2,2,3), ncol=2)
x <- rmvnorm(n=500, mean=c(1,2), sigma=sigma)
colMeans(x)
var(x)

x <- rmvnorm(n=500, mean=c(1,2), sigma=sigma,
     method="chol")
colMeans(x)
var(x)
plot(x)
```

---

mvt

Description

These functions provide information about the multivariate t distribution with non-centrality parameter `delta`, covariance matrix `sigma` and degrees of freedom `df`. `dmvt` gives the density and `rmvt` generates random deviates.

Usage

```
rmvt(n, sigma = diag(2), df = 1)
dmvt(x, delta, sigma, df = 1, log = TRUE)
```

Arguments

`x`	Vector or matrix of quantiles. If `x` is a matrix, each row is taken to be a quantile.
`n`	Number of observations.
`delta`	The vector of noncentrality parameters of length `n`.
`sigma`	Covariance matrix, default is `diag(ncol(x))`.
`df`	Degree of freedom as integer.
`log`	Logical; if `TRUE`, densities d are given as $\log(d)$.

Examples

```
dmvt(x=c(0,0), sigma = diag(2))
x <- rmvt(n=100, sigma = diag(2), df = 3)
plot(x)
```

---

pmvnorm

Description

Computes the distribution function of the multivariate normal distribution for arbitrary limits and correlation matrices based on algorithms by Genz and Bretz.

Usage

```
pmvnorm(lower=-Inf, upper=Inf, mean=rep(0, length(lower)),
        corr=NULL, sigma=NULL, algorithm = GenzBretz(),
        ...)
```

Arguments

`lower`	The vector of lower limits of length `n`.
`upper`	The vector of upper limits of length `n`.
`mean`	The mean vector of length `n`.
`corr`	The correlation matrix of dimension `n`.
`sigma`	The covariance matrix of dimension `n`. Either `corr` or `sigma` can be specified. If `sigma` is given, the problem is standardized. If neither `corr` nor `sigma` is given, the identity matrix is used for `sigma`.
`algorithm`	An object of class `GenzBretz` or `Miwa` specifying both the algorithm to be used as well as the associated hyper parameters.
`...`	Additional parameters (currently given to `GenzBretz` for backward compatibility issues).

Details

This program involves the computation of multivariate normal probabilities with arbitrary correlation matrices. It involves both the computation of singular and nonsingular probabilities. The methodology is described in Genz (1992, 1993).

Note that both `-Inf` and `+Inf` may be specified in `lower` and `upper`. For more details see `pmvt`.

The multivariate normal case is treated as a special case of `pmvt` with `df` $= 0$ and univariate problems are passed to `pnorm`.

Multivariate normal density and random numbers are available using `dmvnorm` and `rmvnorm`.

Value

The evaluated distribution function is returned with attributes

`error`	estimated absolute error and
`msg`	status messages.

Examples

```
n <- 5
mean <- rep(0, 5)
lower <- rep(-1, 5)
upper <- rep(3, 5)
corr <- diag(5)
corr[lower.tri(corr)] <- 0.5
corr[upper.tri(corr)] <- 0.5
prob <- pmvnorm(lower, upper, mean, corr)
print(prob)
stopifnot(pmvnorm(lower=-Inf, upper=3,
          mean=0, sigma=1) == pnorm(3))

a <- pmvnorm(lower=-Inf, upper=c(.3,.5), mean=c(2,4),
             diag(2))
stopifnot(round(a,16) ==
          round(prod(pnorm(c(.3,.5),c(2,4))), 16))

a <- pmvnorm(lower=-Inf, upper=c(.3,.5,1), mean=c(2,4,1),
             diag(3))
stopifnot(round(a,16) ==
          round(prod(pnorm(c(.3,.5,1),c(2,4,1))), 16))

# Example from R News paper (original by Genz, 1992):
m <- 3
sigma <- diag(3)
sigma[2,1] <- 3/5
sigma[3,1] <- 1/3
sigma[3,2] <- 11/15
pmvnorm(lower=rep(-Inf, m), upper=c(1,4,2), mean=rep(0, m),
        corr=sigma)

# Correlation and Covariance
a <- pmvnorm(lower=-Inf, upper=c(2,2), sigma = diag(2)*2)
b <- pmvnorm(lower=-Inf, upper=c(2,2)/sqrt(2), corr=diag(2))
stopifnot(all.equal(round(a,5) , round(b, 5)))
```

pmvt

Description

Computes the distribution function of the multivariate t distribution for arbitrary limits, degrees of freedom and correlation matrices based on algorithms by Genz and Bretz.

Usage

```
pmvt(lower=-Inf, upper=Inf, delta=rep(0, length(lower)),
     df=1, corr=NULL, sigma=NULL, algorithm = GenzBretz(),
     ...)
```

Arguments

`lower`	The vector of lower limits of length `n`.
`upper`	The vector of upper limits of length `n`.
`delta`	The vector of noncentrality parameters of length `n`.
`df`	Degree of freedom as integer.
`corr`	The correlation matrix of dimension `n`.
`sigma`	The covariance matrix of dimension `n`. Either `corr` or `sigma` can be specified. If `sigma` is given, the problem is standardized. If neither `corr` nor `sigma` is given, the identity matrix is used for `sigma`.
`algorithm`	An object of class `GenzBretz` defining the hyper parameters of this algorithm.
`...`	Additional parameters.

Details

This program involves the computation of central and noncentral multivariate t-probabilities with arbitrary correlation matrices. It involves both the computation of singular and nonsingular probabilities. The methodology is described in Genz and Bretz (1999, 2002).

For a given correlation matrix `corr`, for short A, say, (which has to be positive semi-definite) and degrees of freedom `df` the following values are numerically evaluated

$$I = K \int s^{df-1} e^{-s^2/2} \Phi(s \cdot lower/\sqrt{df} - delta, s \cdot upper/\sqrt{df} - delta) ds$$

where $\Phi(a, b) = K' \int_a^b \exp(-x'Ax/2)dx$ is the multivariate normal distribution, $K' = 1/\sqrt{|A|)(2\pi)^m}$ and $K = 2^{1-df/2}/Gamma(df/2)$ are

constants and the (single) integral of I goes from 0 to +Inf.

Note that both `-Inf` and `+Inf` may be specified in the lower and upper integral limits in order to compute one-sided probabilities. Randomized quasi-Monte Carlo methods are used for the computations.

Univariate problems are passed to `pt`. If `df` $= 0$, normal probabilities are returned.

Value

The evaluated distribution function is returned with attributes

`error`	estimated absolute error and
`msg`	status messages.

Examples

```
n <- 5
lower <- -1
upper <- 3
df <- 4
corr <- diag(5)
corr[lower.tri(corr)] <- 0.5
delta <- rep(0, 5)
prob <- pmvt(lower=lower, upper=upper, delta=delta,
        df=df, corr=corr)
print(prob)
pmvt(lower=-Inf, upper=3, df = 3, sigma = 1) == pt(3, 3)

# Example from Edwards and Berry (1987)
n <- c(26, 24, 20, 33, 32)
V <- diag(1/n)
df <- 130
C <- c( 1,  1,  1,  0,  0,
       -1,  0,  0,  1,  0,
        0, -1,  0,  0,  1,
        0,  0,  0, -1, -1,
        0,  0, -1,  0,  0)
C <- matrix(C, ncol=5)

### covariance matrix
cv <- C %*% V %*% t(C)

### correlation matrix
dv <- t(1/sqrt(diag(cv)))
cr <- cv * (t(dv) %*% dv)
delta <- rep(0,5)
```

```
myfct <- function(q, alpha) {
  lower <- rep(-q, ncol(cv))
  upper <- rep(q, ncol(cv))
  pmvt(lower=lower, upper=upper, delta=delta, df=df,
       corr=cr, abseps=0.0001) - alpha
}

round(uniroot(myfct, lower=1, upper=5, alpha=0.95)$root, 3)

# compare pmvt and pmvnorm for large df:
a <- pmvnorm(lower=-Inf, upper=1, mean=rep(0, 5),
             corr=diag(5))
b <- pmvt(lower=-Inf, upper=1, delta=rep(0, 5),
          df=rep(300,5), corr=diag(5))
stopifnot(round(a, 2) == round(b, 2))

# correlation and covariance matrix
a <- pmvt(lower=-Inf, upper=2, delta=rep(0,5), df=3,
          sigma = diag(5)*2)
b <- pmvt(lower=-Inf, upper=2/sqrt(2), delta=rep(0,5),
          df=3, corr=diag(5))
attributes(a) <- NULL
attributes(b) <- NULL
stopifnot(all.equal(round(a,3) , round(b, 3)))

a <- pmvt(0, 1,df=10)
attributes(a) <- NULL
b <- pt(1, df=10) - pt(0, df=10)
stopifnot(all.equal(round(a,10) , round(b, 10)))
```

---

qmvnorm

Description

Computes the equicoordinate quantile function of the multivariate normal distribution for arbitrary correlation matrices based on an inversion of the algorithms by Genz and Bretz.

Usage

```
qmvnorm(p, interval = c(-10, 10), tail = c("lower.tail",
        "upper.tail", "both.tails"), mean = 0, corr = NULL,
        sigma = NULL, algorithm = GenzBretz(), ...)
```

Arguments

`p`	Probability.
`interval`	A vector containing the end-points of the interval to be searched by `uniroot`.
`tail`	Specifies which quantiles should be computed. `lower.tail` gives the quantile `x` for which $P[X \leq x] = p$, `upper.tail` gives `x` with $P[X > x] = p$ and `both.tails` leads to `x` with $P[-x \leq X \leq x] = p$.
`mean`	The mean vector of length `n`.
`corr`	The correlation matrix of dimension `n`.
`sigma`	The covariance matrix of dimension n. Either `corr` or `sigma` can be specified. If `sigma` is given, the problem is standardized. If neither `corr` nor `sigma` is given, the identity matrix is used for `sigma`.
`algorithm`	An object of class `GenzBretz` or `Miwa` specifying both the algorithm to be used as well as the associated hyper parameters.
`...`	Additional parameters to be passed to `uniroot`.

Details

Only equicoordinate quantiles are computed, i.e., the quantiles in each dimension coincide. Currently, the distribution function is inverted by using the `uniroot` function which may result in limited accuracy of the quantiles.

Value

A list with four components: `quantile` and `f.quantile` give the location of the quantile and the value of the function evaluated at that point. `iter` and `estim.prec` give the number of iterations used and an approximate estimated precision from `uniroot`.

Examples

```
qmvnorm(0.95, sigma = diag(2), tail = "both")
```

---

qmvt

Description

Computes the equicoordinate quantile function of the multivariate t distribution for arbitrary correlation matrices based on an inversion of the algorithms by Genz and Bretz.

Usage

```
qmvt(p, interval = c(-10, 10), tail = c("lower.tail",
     "upper.tail", "both.tails"), df = 1, delta = 0,
     corr = NULL, sigma = NULL, algorithm = GenzBretz(),
     ...)
```

Arguments

`p`	Probability.
`interval`	A vector containing the end-points of the interval to be searched by `uniroot`.
`tail`	Specifies which quantiles should be computed. `lower.tail` gives the quantile `x` for which $P[X \leq x] = p$, `upper.tail` gives `x` with $P[X > x] = p$ and `both.tails` leads to `x` with $P[-x \leq X \leq x] = p$.
`delta`	The vector of noncentrality parameters of length `n`.
`df`	Degree of freedom as integer.
`corr`	The correlation matrix of dimension `n`.
`sigma`	The covariance matrix of dimension `n`. Either `corr` or `sigma` can be specified. If `sigma` is given, the problem is standardized. If neither `corr` nor `sigma` is given, the identity matrix is used for `sigma`.
`algorithm`	An object of class `GenzBretz` defining the hyper parameters of this algorithm.
`...`	Additional parameters to be passed to `uniroot`.

Details

Only equicoordinate quantiles are computed, i.e., the quantiles in each dimension coincide. Currently, the distribution function is inverted by using the `uniroot` function which may result in limited accuracy of the quantiles.

Value

A list with four components: `quantile` and `f.quantile` give the location of the quantile and the value of the function evaluated at that point. `iter` and `estim.prec` give the number of iterations used and an approximate estimated precision from `uniroot`.

Examples

```
qmvt(0.95, df = 16, tail = "both")
```

B

Description of the MATLAB Functions

In this appendix we reproduce the relevant documentation for the key MATLAB functions available from the Genz website (www.math.wsu.edu/faculty/genz), by following the "Software" link. All functions have introductory comment sections which describe the functions input and output parameters, and some of this comment material is reproduced here.

bvnl :

```
   function p = bvnl( dh, dk, r )
%
%  A function for bivariate normal probabilities.
%  bvnl calculates the probability that x < dh and y < dk.
%    parameters
%      dh 1st upper integration limit
%      dk 2nd upper integration limit
%      r   correlation coefficient
```

tvnl :

```
   function tvn = tvnl( h, r, epsi )
%
%   A function for trivariate normal probabilities;
%    it computes the probability that x(i) < h(i), i = 1-3.
%   h     real array of three upper integration limits
%   r     real array of three correlation coefficients,
%         r should contain the lower left portion of the
%         correlation matrix, r21, r31, r32, in that order.
%   epsi  real required absolute accuracy; maximum
%          accuracy for most computations is 1e-14.
```

tvnls : a simpler but less accurate version of **tvnl**.
qsimvn :

```
function [ p, e ] = qsimvn( m, r, a, b )
%
%  [ P E ] = QSIMVN( M, R, A, B )
%    uses a randomized quasi-random rule with m points to
%    estimate an MVN probability for a positive definite
%    covariance matrix r, with lower integration limits a
%    and upper integration limits b.
%   Probability p is output with error estimate e.
%    Example usage:
%     >> r = [4 3 2 1;3 5 -1 1;2 -1 4 2;1 1 2 5];
%     >> a = -inf*[1 1 1 1 ]'; b = [ 1 2 3 4 ]';
%     >> [ p e ] = qsimvn( 5000, r, a, b ); disp([ p e ])
```

qsimvnv : a vectorized version of **qsimvn**, with similar parameters.
qsilatmvnv : similar to **qsimvnv**, except randomized lattice rules (see Section 4.2.2) are used for the numerical integration.
qscmvn :

```
function [ p, e ] = qscmvn( m, r, a, cn, b )
%
%  [ P E ] = QSCMVN( M, R, A, CN, B )
%    uses a randomized quasi-random rule with m points to
%    estimate an MVN probability for positive semi-definite
%    covariance matrix r, with constraints a < cn*x < b. If
%    r is nxn and cn is kxn, a and b must be kx1.
%   Probability p is output with error estimate e.
%    Example usage:
%     >> r = [4 3 2 1;3 5 -1 1;2 -1 4 2;1 1 2 5];
%     >> a = [ -inf 1 -5 ]'; b = [ 3 inf 4 ]';
%     >> cn = [ 1 2 3 -2; 2 4 1 2; -2 3 4 1 ];
%     >> [p e] = qscmvn( 5000, r, a, cn, b ); disp([ p e ])
```

qscmvnv : a vectorized version of **qscmvn**, with similar parameters.
qsclatmvnv : similar to **qscmvnv**, except randomized lattice rules (see Section 4.2.2) are used for the numerical integration.
mvnlps :

```
function mvnval  = mvnlps( mu, sigma, q, e, r, re )
%
% MVNLPS Multivariate Normal probability for an ellipsoid.
%        MVNVAL = MVNLPS( MU, SIGMA, Q, E, R, RE ) computes
%        an MVN value to relative accuracy RE for an
%        ellipsoid centered at Q with radius R and positive
%        semi-definite ellipsoid matrix E:
%          MVNVAL = PROB( ( X - Q )'E ( X - Q ) < R^2 )
%        SIGMA is a positive definite covariance matrix for
%        a multivariate normal (MVN) density with mean MU.
```

```
%        MU and Q must be column vectors.
%        Example:
%         sg = [ 3 2 1;2 2 1;1 1 1]; mu = [1 -1 0]';
%         e = [4 1 -1; 1 2 1; -1 1 2]; q = [2 3 -2]';
%         p = mvnlps( mu, sg, q, e, 4, 1e-5 ); disp(p)
```

qsimvnef :

```
function [ p, e, ef, efe ] = qsimvnef( m, r, a, b, f )
%
%  [ P E EF EFE ] = QSIMMVNEF( M, R, A, B )
%    uses a randomized quasi-random rule with m points to
%    estimate an MVN expectation for positive definite
%    covariance matrix r, with lower and upper integration
%    limits a b, and expectation function f.
%   Probability MVN p is output with error estimate e,
%    along with expected value ef and error estimate efe.
%      Note: ef approx.= E[f] = I[f]/p = I[f]/I[1], where
%       I[.] denotes the truncated MVN integral.
%    Example:
%     r = [4 3 2 1;3 5 -1 1;2 -1 4 2;1 1 2 5];
%     a = -inf*[1 1 1 1 ]'; b = [ 1 2 3 4 ]';
%     f = inline('x(1)^2*x(2)*x(3)*x(4)','x');
%     [ p e ef efe ] = qsimvnef( 50000, r, a, b, f );
%     disp([ p e ef efe ])
```

qsimvnefv : a vectorized version of **qsimvnef**, with similar parameters.

tvtl :

```
function tvt = tvtl( nu, h, r, epsi )
%
%     A function for computing trivariate normal and
%      t-probabilities; tvtl computes the probability
%      that x(i) < h(i), for i = 1, 2, 3.
%   parameters
%    nu    integer degrees of freedom;
%            use nu = 0 for normal probabilities.
%    h     real array of three upper integration limits
%    r     real array of three correlation coefficients,
%          r should contain the lower left portion of the
%          correlation matrix, r21, r31, r32, in that order.
%    epsi  requested absolute accuracy; maximum accuracy
%           for most computations is approximately 1e-14.
```

qsimvt :

```
function [ p, e ] = qsimvt( m, nu, sigma, a, b )
%
```

```
%  QSIMVT
%    Qsimvt estimates a multivariate t-probability using an
%        m point randomized quasi-random integration method
%    for a t-distribution with
%        nu degrees-of-freedom;
%            Note: nu <= 0 computes normal probabilities.
%        sigma, a positive definite covariance matrix.
%    The probability is computed over a hyper-rectangle with
%        a, lower integration limits and
%        b, upper integration limits.
%      If sigma is nxn then a and b must be nx1.
%    Output for qsimvt is
%        p, the estimated probability, along with
%        e, an absolute error estimate for p.
%    Example:
%     >> r = [4 3 2 1;3 5 -1 1;2 -1 4 2;1 1 2 5];
%     >> a = -inf*[1 1 1 1 ]'; b = [ 1 2 3 4 ]';
%     >> [p e] = qsimvt( 5000, 5, r, a, b ); disp([p e])
```

qscmvt :

```
function [ p, e ] = qscmvt( m, nu, sigma, a, cn, b )
%
%  QSCMVT
%    [ p, e ] = qscmvt( m, nu, sigma, a, cn, b )
%    Qsimvt computes a multivariate t-probability using an
%       m-point randomized quasi-random integration method
%     for a t-distribution with
%       nu degrees-of-freedom;
%           use nu = 0 for normal probabilities.
%       sigma, a positive semi-definite covariance matrix.
%    The probability is computed over an integration region
%     specified a < cn*x < b. If sigma is n x n and
%        cn is k x n, then a and b must be kx1.
%    Output for qscmvt is
%        p, the estimated probability, along with
%        e, an absolute error estimate for p.
```

qscmvtv : a vectorized version of **qscmvt**, with similar parameters.

References

Abrahamson IG (1964) Orthant probabilities for the quadrivariate normal distribution. The Annals of Mathematical Statistics 25:1685–1703 Cited on p. 18.

Albanese C, Seco L (2001) Harmonic analysis in value-at-risk calculations. Matemática Iberoamericano 17:195–219 Cited on p. 15.

Albers W, Kallenberg WC (1994) A simple approximation to the bivariate normal distribution with large correlation coefficient. Journal of Multivariate Analysis 49:87–96 Cited on p. 9.

Ambartzumian R, der Kiureghian A, Ohanian V, Sukiasian H (1998) Multinormal probability by sequential conditioned importance sampling: Theory and application. Probabilistic Engineering Mechanics 13:299–308 Cited on p. 44, 68.

Bansal V, Perkins JD, Pistikopoulos EN (2000) Using mathematical programming to compute singular multivariate normal probabilities. Journal of Statistical Computation and Simulation 67:219–253 Cited on p. 56, 58.

Beckers M, Haegemans A (1992) Comparison of numerical integration techniques for multivariate normal integrals. Technical report, Catholic University of Leuven, Computer Science Department, Belgium Cited on p. 48.

Berens H, Schmid HJ, Xu Y (1995) Multivariate Gaussian cubature formulae. Archiv der Mathematik 64:26–32 Cited on p. 68.

Berntsen J, Espelid TO, Genz A (1991a) An adaptive algorithm for the approximate calculation of multiple integrals. ACM Transactions on Mathematical Software 17:437–451 Cited on p. 52, 60.

Berntsen J, Espelid TO, Genz A (1991b) Algorithm 698: DCUHRE - an adaptive multidimensional integration routine for a vector of integrals. ACM Transactions on Mathematical Software 17:452–456 Cited on p. 52, 60, 86.

Bohrer R, Schervish MJ (1981) An error-bounded algorithm for normal probabilities of rectangular regions. Technometrics 23:297–300 Cited on p. 54.

Bommert K (1999) Das multinomiale Probitmodell. Josef Eul Verlag, Lohmar Cited on p. 43.

Boros E, Prékopa A (1989) Closed form two-sided bounds for probabilities that at least r and exactly r out of n events occur. Mathematical Operations Research 14:317–342 Cited on p. 23.

Börsch-Supan A, Hajivassiliou VA (1993) Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Journal of Econometrics 58:347–368 Cited on p. 44.

Braaten E, Weller G (1979) An improved low-discrepancy sequence for multidimensional Quasi-Monte Carlo integration. Journal of Computational Physics 33:249–258 Cited on p. 49.

Breitung K, Hohenbichler M (1989) Asymptotic approximations for multivariate integrals with an application to multinormal probabilities. Journal of Multivariate Analysis 30:80–97 Cited on p. 27.

Breitung K, Richter WD (1996) A geometric approach to an asymptotic expansion for large deviation probabilities of Gaussian random vectors. Journal of Multivariate Analysis 58:1–20 Cited on p. 27.

Breslaw JA (1994) Evaluation of multivariate normal probability integrals using a low variance simulator. The Review of Economics and Statistics 76:673–682 Cited on p. 43.

Bretz F, Genz A, Hothorn LA (2001) On the numerical availability of multiple comparison procedures. Biometrical Journal 43:645–656 Cited on p. 71, 72.

Bretz F, Hothorn T, Westfall P (2008) Multiple comparison procedures in linear models. In: Brito P (ed) COMPSTAT 2008 - Proceedings in Computational Statistics, Physica Verlag, Heidelberg, pp 423–431 Cited on p. 74.

Bretz F, Hothorn T, Westfall P (2010) Multiple Comparisons Procedures in R. Taylor & Francis, Boca Raton Cited on p. 71, 74.

Brodtkorb P (2006) Evaluating nearly singular multinormal expectations with application to wave distributions. Methodology and Computing in Applied Probability 8:65–91 Cited on p. 58.

Brummelhuis R, Cordoba A, Quintanilla M, Seco L (2002) Principal component value-at-risk. Mathematical Finance 12:23–43 Cited on p. 15.

Bukszár J (2001) Upper bounds for the probability of a union by multitrees. Advances in Applied Probability 33:437–452 Cited on p. 24.

Bukszár J, Prékopa A (2001) Probability bounds with cherry trees. Mathematics of Operations Research 26:174–192 Cited on p. 22.

Bulgren WG, Dykstra RL, Hewett JE (1974) A bivariate t-distribution with applications. Journal of the American Statistical Association 69:525–532 Cited on p. 68.

Bungartz H, Dirnstorfer S (2003) Multivariate quadrature on adaptive sparse grids. Computing 71:89–114 Cited on p. 51, 53.

Caflisch RE (1998) Monte Carlo and Quasi-Monte Carlo methods. Acta Numerica 7:1–49 Cited on p. 46, 59.

Caflisch RE, Morokoff W, Owen AB (1997) Valuation of mortgage backed securities using brownian bridges to reduce effective dimension. Journal of Computational Finance 1:27–46 Cited on p. 45, 59.

Childs DR (1967) Reduction of the multivariate normal integral to characteristic form. Biometrika 54:293–299 Cited on p. 12.
Clark EC (1961) The greatest of a finite set of random variables. Operations Research 9:145–162 Cited on p. 13, 26.
Cools R (1999) Monomial cubature rules since "Stroud": A compilation. Journal of Computational and Applied Mathematics 112:21–28 Cited on p. 51.
Cools R, Kuo F, Nuyens D (2006) Constructing embedded lattice rules for multivariate integration. SIAM Journal on Scientific Computing 28:2162–2188 Cited on p. 48.
Cornish EA (1954) The multivariate t-distribution associated with a set of normal sample deviates. Australian Journal of Physics 7:531–542 Cited on p. 1, 3, 72.
Cox DR, Wermuth N (1991) A simple approximation for bivariate and trivariate normal integrals. International Statistical Review 59:263–269 Cited on p. 9.
Craig P (2008) A new reconstruction of multivariate normal orthant probabilities. Journal of the Royal Statistical Society, Ser B 70:227–243 Cited on p. 18, 19, 20, 39, 41.
Cranley R, Patterson TNL (1976) Randomization of number theoretic methods for multiple integration. SIAM Journal on Numerical Analysis 13:904–914 Cited on p. 48.
Curbera F (1998) Optimal integration of Lipschitz functions with a Gaussian weight. Journal of Complexity 14:122–149 Cited on p. 59.
Curbera F (2000) Delayed curse of dimension for Gaussian integration. Journal of Complexity 16:474–506 Cited on p. 39, 59.
Curnow RN, Dunnett CW (1962) The numerical evaluation of certain multivariate normal integrals. The Annals of Mathematical Statistics 33:571–579 Cited on p. 17, 18.
Dash J (2007) Multivariate integral perturbation teachniques - I (Theory). International Journal of Theoretical and Applied Finance 10:1287–1304 Cited on p. 27.
David FN (1953) A note on the evaluation of the multivariate normal integral. Biometrika 40:458–459 Cited on p. 1, 12.
Davis PJ, Rabinowitz P (1984) Methods of Numerical Integration. Academic Press, New York Cited on p. 17, 51.
Dawson D, Sankoff A (1967) An inequality for probabilities. Proceedings of the American Mathematical Society 18:504–507 Cited on p. 22, 76.
Deák I (1980) Three digit accurate multiple normal probabilities. Numerische Mathematik 35:369–380 Cited on p. 2, 29, 42, 60.
Deák I (1986) Computing probabilities of rectangles in case of multinormal distribution. Journal of Statistical Computation and Simulation 26:101–114 Cited on p. 44.
Deák I (1990) Random Number Generators and Simulation. Akadémiai Kiadó, Budapest Cited on p. 29, 41, 60.

Deák I (2001) Subroutines for computing normal probabilities of sets - Computer experiences. The Annals of Operations Research 100:103–122 Cited on p. 44.

Dellaportas P, Smith AFM (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. Journal of the Royal Statistical Society, Series C 42:443–459 Cited on p. 86.

Dick J, Kuo F (2004) Reducing the cost of component-by-component construction of good lattice rules. Mathematics of Computation 73:1967–1988 Cited on p. 47.

DiDonato AR, Hageman RK (1982) A method for computing the integral of the bivariate normal distribution over an arbitrary polygon. SIAM Journal on Scientific and Statistical Computing 3:434–446 Cited on p. 68.

DiDonato AR, Jarnagin MP, Hageman RK (1980) Computation of the integral of the bivariate normal distribution over convex polygons. SIAM Journal on Scientific and Statistical Computing 1:179–186 Cited on p. 68.

Divgi DR (1979) Calculation of univariate and bivariate normal probability functions. The Annals of Statistics 7:903–910 Cited on p. 8.

Dohmen K (2000) Improved Bonferroni inequalities via union-closed set systems. Journal of Combinatorial Theory, Series A 92:61–67 Cited on p. 24.

Dohmen K (2003) Improved Bonferroni Inequalities via Abstract Tubes, Lecture Notes in Mathematics, vol 1826. Springer-Verlag, Heidelberg Cited on p. 24.

de Doncker E, Ciobanu M, Genz A (1999) Parallel computation of multivariate normal probabilities. Computing Science and Statistics 30:89–93 Cited on p. 69.

de Doncker E, Cucos L, Zanny R, Genz A (2001) Parallel computation of multivariate t-probabilities. In: Tentner A (ed) Proceedings of the High Performance Computing Symposium − HPC 2001, Society for Computer Simulation International, San Diego, pp 129–134 Cited on p. 69.

van Dooren P, de Ridder L (1976) An adaptive algorithm for numerical integration over an n-dimensional rectangular region. Journal of Computational and Applied Mathematics 2:207–217 Cited on p. 52.

Drezner Z (1990) Approximations to the multivariate normal integral. Communications in Statistics − Simulation and Computation 19:527–534 Cited on p. 68.

Drezner Z (1992) Computation of the multivariate normal integral. ACM Transactions on Mathematical Software 18:470–480, correction: **19** (1993) 546 Cited on p. 54, 68.

Drezner Z (1994) Computation of the trivariate normal integral. Mathematics of Computation 62:289–294 Cited on p. 11.

Drezner Z, Wesolowsky GO (1990) On the computation of the bivariate normal integral. Journal of Statistical Computation and Simulation 3:101–107 Cited on p. 9.

Drmota M, Tichy RF (1997) Sequences, Discrepancies and Applications, Lecture Notes in Mathematics, vol 1651. Springer-Verlag, New York Cited on p. 47.

Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50:1096–1121 Cited on p. 72.

Dunnett CW (1989) Algorithm AS 251: Multivariate normal probability integral with product correlation structure. Journal of the Royal Statistical Society, Series C 38:564–579, correction: **42** (1993) 709 Cited on p. 17, 62, 73.

Dunnett CW, Sobel M (1954) A bivariate generalization of Student's distribution with tables for certain special cases. Biometrika 41:153–169 Cited on p. 1, 9, 10, 72.

Escoufier Y (1967) Calculs de probabilités par un méthode de Monte Carlo pour uns variable p-normale. Revue de Statistique Appliquée 15:5–15 Cited on p. 2.

Evans M (1991) Adaptive importance sampling and chaining. In: Flournoy N, Tsutakawa RK (eds) Statistical Numerical Integration, American Mathematical Society, Providence, Rhode Island, Contemporary Mathematics, vol 115, pp 137–143 Cited on p. 85.

Evans M, Swartz T (1988) Monte Carlo computation of some multivariate normal probabilities. Journal of Statistical Computation and Simulation 30:117–128 Cited on p. 12.

Evans M, Swartz T (1995) Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. Statistical Science 10:254–272, discussion: **11** (1996) 54−64 Cited on p. 1, 79, 82.

Evans M, Swartz T (2000) Approximating Integrals via Monte Carlo and Deterministic Methods. Oxford University Press, Oxford Cited on p. 54, 82.

Fang KT, Wang Y (1994) Number-theoretic Methods in Statistics. Chapmann and Hall, London Cited on p. 36, 43.

Fang KT, Hickernell FJ, Niederreiter H (2002) Monte Carlo and Quasi-Monte Carlo Methods 2000. Springer Verlag, Berlin Cited on p. 47.

Faure H (1982) Discrépance de suites associées à un système de numération (en dimension s). Acta Arithmetica 41:337–351 Cited on p. 49.

Fishman G (2006) A First Course in Monte Carlo. Thomson Brooks/Cole Cited on p. 89.

Galambos J, Simonelli I (1996) Bonferroni-type Inequalities with Applications. Springer Verlag, New York Cited on p. 24.

Gassmann HI (1988) Conditional probability and conditional expectation of a random vector. In: Ermoliev Y, Wets RJB (eds) Numerical Techniques for Stochastic Optimization, Springer Verlag, Berlin, pp 237–254 Cited on p. 44, 60.

Gassmann HI (2002) Rectangle probabilities of trivariate normal distributions. Technical report, Dalhousie University, available at http://myweb.dal.ca/gassmann/TRIVAR.HTM Cited on p. 11.

Gassmann HI, Deák I, Szántai T (2002) Computing multivariate normal probabilities: A new look. Journal of Computational and Graphical Statistics 11:920–949 Cited on p. 60, 61.

Genz A (1986) Fully symmetric interpolatory rules for multiple integrals. SIAM Journal on Numerical Analysis 23:1273–1283 Cited on p. 51, 53.

Genz A (1992) Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics 1:141–150 Cited on p. 31, 42, 60.

Genz A (1993) Comparison of methods for the computation of multivariate normal probabilities. Computing Science and Statistics 25:400–405 Cited on p. 48, 52, 60, 61.

Genz A (2004) Numerical computation of rectangular bivariate and trivariate normal and t probabilities. Statistics and Computing 14:151–160 Cited on p. 9, 11, 64, 65, 66.

Genz A, Bretz F (1999) Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. Journal of Statistical Computation and Simulation 63:361–378 Cited on p. 31, 33, 48.

Genz A, Bretz F (2000) Numerical computation of critical values for multiple comparison problems. In: ASA Proceedings of the Statistical Computing Section, pp 84–87 Cited on p. 24, 75, 77.

Genz A, Bretz F (2002) Methods for the computation of multivariate t-probabilities. Journal of Computational and Graphical Statistics 11:950–971 Cited on p. 29, 31, 36, 39, 42, 44, 48, 52, 61.

Genz A, Kahaner DK (1986) The numerical evaluation of certain multivariate normal integrals. Journal of Computational and Applied Mathematics 16:255–258 Cited on p. 20.

Genz A, Kass R (1997) Subregion adaptive integration of functions having a dominant peak. Journal of Computational and Graphical Statistics 6:92–111 Cited on p. 83, 84, 85.

Genz A, Kwong KS (2000) Numerical evaluation of singular multivariate normal distributions. Journal of Statistical Computation and Simulation 68:1–21 Cited on p. 56, 58.

Genz A, Monahan J (1998) Stochastic integration rules for infinite regions. SIAM Journal on Scientific Computing 19:426–439 Cited on p. 59.

Genz A, Monahan J (1999) A stochastic algorithm for high-dimensional integrals over unbounded regions with Gaussian weight. Journal of Computational and Applied Mathematics 112:71–81 Cited on p. 59.

Genz A, Bretz F, Hochberg Y (2004) Approximations to multivariate t integrals with application to multiple comparison procedures. In: Benjamini Y, Bretz F, Sarkar S (eds) New developments in multiple comparison procedures, Institute of Mathematical Statistics, Beachwood, Ohio, IMS Lecture Notes - Monograph Series, vol 47, pp 24–32 Cited on p. 24, 76.

Gerstner T, Griebel G (1998) Numerical integration using sparse grids. Numerical Algorithms 18:209–232 Cited on p. 51, 53.

Geweke J (1989) Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57:1317–1340 Cited on p. 85.

Geweke J (1991a) Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. Computing Science and Statistics 23:571–578 Cited on p. 31, 42, 44.

Geweke J (1991b) Generic, algorithmic approaches to Monte Carlo integration in Bayesian inference. In: Flournoy N, Tsutakawa RK (eds) Statistical Numerical Integration, American Mathematical Society, Providence, Rhode Island, Contemporary Mathematics, vol 115, pp 117–135 Cited on p. 85.

Gibbons RD, Bock RD, Hedekker DR (1987) Approximating multivariate normal orthant probabilities using the Clark algorithm. ONR Technical report, University of Chicago Cited on p. 13.

Gibbons RD, Bock RD, Hedekker DR (1990) Approximating multivariate normal orthant probabilities. ONR Technical report, University of Chicago Cited on p. 13.

Gibson GJ, Glasbey CA, Elston DA (1994) Monte Carlo evaluation of multivariate normal integrals and sensitivity to variate ordering. In: Dimov IT, Sendov B, Vassilevski PS (eds) Advances in Numerical Methods and Applications, World Scientific Publishing, River Edge, pp 120–126 Cited on p. 36, 39.

Glasserman P (2004) Monte Carlo Methods in Financial Engineering. Springer Verlag, New York Cited on p. 59, 80.

Gollwitzer S, Rackwitz R (1987) Comparison of numerical schemes for the multinormal integral. In: Thoft-Christensen P (ed) Reliability and Optimisation of Structural Systems, Springer Verlag, Berlin, Lecture Notes in Engineering, vol 33, pp 157–174 Cited on p. 68.

Golub GH, Van Loan CF (1996) Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore Cited on p. 39, 56.

Goodman LA (1981) Association models and the bivariate normal for contingency tables with ordered categories. Biometrika 68:347–355 Cited on p. 68.

Gupta SS (1963a) Probability integrals of multivariate normal and multivariate t. The Annals of Mathematical Statistics 34:792–828 Cited on p. 10.

Hajivassiliou VA (1993a) Simulation estimation methods for limited dependent variable models. In: Maddala GS, Rao CR, Vinod HD (eds) Handbook of Statistics, Vol. 11: Econometrics, North Holland, Amsterdam, pp 519–543 Cited on p. 42.

Hajivassiliou VA (1993b) Simulating normal rectangle probabilities and their derivatives: The effects of vectorization. The International Journal of Supercomputer Applications 7:231–253 Cited on p. 69.

Hajivassiliou VA, McFadden D, Ruud P (1996) Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and com-

putational results. Journal of Econometrics 72:85–134 Cited on p. 41, 44, 60, 61.

Halton JH (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2:84–90, correction: **2** (1960) 196 Cited on p. 49.

Hammersley JM, Handscomb DC (1965) Monte Carlo Methods. Wiley, New York Cited on p. 41.

Harris B, Soms AP (1980) The use of the tetrachoric series for evaluating multivariate normal probabilities. Journal of Multivariate Analysis 10:252–267 Cited on p. 2, 27.

Hartzel J, Liu IM, Agresti A (2001) Describing heterogeneous effects in stratified ordinal contingency tables, with application to multi-center clinical trials. Computational Statistics and Data Analysis 35:429–449 Cited on p. 1.

Hayter AJ (2006) Recursive integration methodologies with statistical applications. Journal of Statistical Planning and Inference 136:2284–2296 Cited on p. 18.

Hayter AJ, Liu W (1996) A note on the calculation of $Pr\{X_1 < X_2 < \ldots < X_k\}$. The American Statistician 50:365 Cited on p. 18.

Healy MJR (1968) Algorithm AS 6: Triangular decomposition of a symmetric matrix. Journal of the Royal Statistical Society, Series C 17:195–197 Cited on p. 56.

Henery RJ (1981) An approximation to certain multivariate normal probabilities. Journal of the Royal Statistical Society, Series B 43:81–85 Cited on p. 68.

Hickernell FJ (2002) Obtaining $O(N^{-2+\epsilon})$ convergence for lattice quadrature rules. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer Verlag, Berlin, pp 274–289 Cited on p. 48.

Hickernell FJ, Hong HS (1997) Computing multivariate normal probabilities using rank-1 lattice sequences. In: Golub GH, Lui SH, Luk FT, Plemmons RJ (eds) Proceedings of the Workshop on Scientific Computing, Springer Verlag, Singapore, pp 209–215 Cited on p. 48.

Hickernell FJ, Hong HS (2002) Quasi-Monte Carlo methods and their randomizations. In: Chan R, Kwok Y, Yao D, Zhang Q (eds) Proceedings of the IMS Workshop on Applied Probability, International Press, Cambridge, pp 65–88 Cited on p. 49.

Hickernell FJ, Hong HS, L'écuyer P, Lemieux C (2001) Extensible lattice sequences for quasi-Monte Carlo quadrature. SIAM Journal on Scientific Computing 22:1117–1138 Cited on p. 48.

Hlawka E (1961) Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. Annali di Matematica Pura ed Applicata 54:325–333 Cited on p. 45.

Hochberg Y (1994) Approximate and simple multiple comparison procedures based on averaging correlations. Technical Report SOR-94-8, New York University Cited on p. 25.

Hochberg Y, Tamhane AC (1987) Multiple Comparison Procedures. Wiley, New York Cited on p. 1, 25, 71.

Hohenbichler M, Rackwitz R (1983) First-order concepts is system reliability. Structural Safety 1:177–188 Cited on p. 68.

Hohenbichler M, Rackwitz R (1985) A bound and an approximation to the multivariate normal distribution function. Mathematica Japonica 30:821–828 Cited on p. 68.

Hong HS, Hickernell FJ (2003) Algorithm 823: Implementing scrambled digital sequences. ACM Transactions on Mathematical Software 29:95–109 Cited on p. 49, 61.

Hothorn T, Bretz F, Genz A (2001) On multivariate t and Gauss probabilities in R. R Newsletter 1:27–29 Cited on p. 62.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biometrical Journal 50:346–363 Cited on p. 74.

Hsu JC (1992) The factor analytic approach to simultaneous inference in the general linear model. Journal of Computational and Graphical Statistics 1:151–168 Cited on p. 25.

Hsu JC (1996) Multiple Comparisons. Chapman and Hall, London Cited on p. 1, 71, 73.

Hsu JC, Nelson BL (1998) Multiple comparisons in the general linear model. Journal of Computational and Graphical Statistics 7:23–41 Cited on p. 25, 78.

Hull JC (1993) Options, Futures, and Other Derivative Securities. Prentice-Hall, Engelwood Cliffs, New Jersey Cited on p. 80, 88.

Hunter D (1976) An upper bound for the probability of a union. Journal of Applied Probability 13:597–603 Cited on p. 22, 76.

Hutchinson TP (1993) The seventh-root formula for a trivariate normal probability. The American Statistician 47:102–103 Cited on p. 68.

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5:299–314 Cited on p. 62.

Iyengar S (1988) Evaluation of normal probabilities of symmetric regions. SIAM Journal on Scientific and Statistical Computing 9:418–423 Cited on p. 24.

Iyengar S, Tong YL (1989) Convexity properties of elliptically contoured distributions with applications. Sankhyā, Series A 51:13–29 Cited on p. 24.

Jäckel P (2002) Monte Carlo Methods Finance. Wiley, New York Cited on p. 59, 80.

James S (1991) Approximate multinormal probabilities applied to correlated multiple endpoints in clinical trials. Statistics in Medicine 10:1123–1135 Cited on p. 68.

Joe H (1995) Approximations to multivariate normal rectangle probabilities based on conditional expectations. Journal of the American Statistical Association 90:957–964 Cited on p. 26, 62.

Joe S (1990) Randomization of lattice rules for numerical multiple integration. Journal of Computational and Applied Mathematics 31:299–304 Cited on p. 48.

Johnson NL, Kotz S (1970a) Distributions in Statistics: Continuous Univariate Distributions-1. Houghton Mifflin, Boston Cited on p. 7.

Johnson NL, Kotz S (1970b) Distributions in Statistics: Continuous Univariate Distributions-2. Houghton Mifflin, Boston Cited on p. 7.

Johnson NL, Kotz S (1972) Distributions in Statistics: Continuous Multivariate Distributions. John Wiley and Sons, New York Cited on p. 4.

Kamakura WA (1989) The estimation of multinomial probit models: A new calibration algorithm. Transportation Science 23:253–265 Cited on p. 26.

Keane MP (1993) Simulation estimation for panel data models with limited dependent variables. In: Maddala GS, Rao CR, Vinod HD (eds) Handbook of Statistics, Vol. 11: Econometrics, North Holland, Amsterdam, pp 545–571 Cited on p. 42.

Keller A, Heinrich S, Niederreiter H (2008) Monte Carlo and Quasi-Monte Carlo Methods 2006. Springer Verlag, Berlin Cited on p. 47.

Kendall MG (1941) Proof of relations connected with tetrachoric series and its generalization. Biometrika 32:196–198 Cited on p. 1, 27.

Kibble WF (1945) An extension of a theorem of Mehler on Hermite polynomials. Proceedings of the Cambridge Philosophical Society 41:12–15 Cited on p. 1, 27.

Korobov NM (1960) Properties and calculation of optimal coefficients. Doklady Akademii Nauk SSSR 132:1009–1012, english translation: *Soviet Mathematics Doklady* **1** (1960) 696−700 Cited on p. 47.

Kotz S, Nadarajah S (2004) Multivariate t Distributions and Their Applications. Cambridge University Press, Cambridge Cited on p. 1, 4.

Kotz S, Balakrishnan N, Johnson NL (2000) Continuous Multivariate Distributions. Volume 1: Models and Applications. Wiley, New York Cited on p. 1, 68.

Krishnan M (1972) Series representations of a bivariate singly noncentral t distribution. Journal of the American Statistical Association 67:228–231 Cited on p. 68.

Kwong KS (1995) Evaluation of one-sided percentage points of the singular multivariate normal distribution. Journal of Statistical Computation and Simulation 51:121–135 Cited on p. 17.

Kwong KS, Iglewicz B (1996) On singular multivariate normal distribution and its applications. Computational Statistics and Data Analysis 22:271–285 Cited on p. 17.

Langdon MG (1984a) Methods of retrieving choice probability in utility maximizing multiple alternative models. Transportation Research, Series B 18:209–234 Cited on p. 26.

Langdon MG (1984b) Improved algorithms for estimating choice probabilities in the multinomial probit model. Transportation Science 18:267–299 Cited on p. 26.

Larcher G (1998) On the distribution of digital sequences. In: Niederreiter H, Hellekalek P, Larcher G, Zinterhof P (eds) Monte Carlo and Quasi-Monte Carlo Methods 1996, Springer Verlag, New York, Lecture Notes in Statistics, vol 127, pp 109–123 Cited on p. 48.

Lin X, Breslow NE (1996) Analysis of correlated binomial data in logistic-normal models. Journal of Statistical Computation and Simulation 55:133–146 Cited on p. 1.

Lohr SL (1993) Algorithm AS 285: Multivariate normal probabilities of star-shaped regions. Journal of the Royal Statistical Society, Series C 42:576–582 Cited on p. 14, 30, 56, 62.

de Los Reyes JP (1990) Upper bound for singular normal probability integrals by integration over a hypersphere. Communications in Statistics − Simulation and Computation 19:847–861 Cited on p. 58.

Maghsoodloo S, Huang CL (1995) Computing probability integrals of a bivariate normal distribution. InterStat Available at http://interstat.statjournals.net Cited on p. 9.

Marsaglia G (1963) Expressing the normal distribution with covariance matrix $A + B$ in terms of one with covariance matrix A. Biometrika 50:535–538 Cited on p. 16, 17, 18.

Martynov GV (1981) Evaluation of the normal distribution function. Journal of Soviet Mathematics 17:1857–1875 Cited on p. 2.

Maurer W (1983) Bivalent trees and forests or upper bounds for the probability of a union revisited. Discrete Applied Mathematics 6:157–171 Cited on p. 22.

McCann M, Edwards D (1996) A path length inequality for the multivariate-t distribution, with applications to multiple comparisons. Journal of the American Statistical Association 91:211–216 Cited on p. 24.

McCann M, Edwards D (2000) A hybrid method for improved critical points for multiple comparisons. Communications in Statistics − Simulation and Computation 29:703–722 Cited on p. 24.

McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57:995–1026 Cited on p. 44.

McKay MD, Conover WJ, Beckman RJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21:239–245 Cited on p. 60.

Mehler FG (1866) Über die Entwicklung einer Funktion von beliebig vielen Variablen nach Laplace'schen Funktionen höherer Ordnung. Journal für die Reine und Angewandte Mathematik 66:161–176 Cited on p. 1.

Mendell NR, Elston RC (1974) Multifactorial qualitative traits: Genetic analysis and prediction of recurrence risks. Biometrics 30:41–57 Cited on p. 25, 37.

Milton RC (1972) Computer evaluation of the multivariate normal integral. Technometrics 14:881–889 Cited on p. 54.

Miwa T, Hayter AJ, 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. Computational Statistics and Data Analysis 34:17–32 Cited on p. 19.

Miwa T, Hayter AJ, Kuriki S (2003) The evaluation of general non-centred orthant probabilities. Journal of the Royal Statistical Society, Ser B 65:223–234 Cited on p. 18, 39, 41.

Monahan J, Genz A (1997) Spherical-radial integration rules for Bayesian computation. Journal of the American Statistical Association 92:664–674 Cited on p. 29, 59.

Moran PAP (1983) A new expansion for the multivariate normal distribution. Australian Journal of Statistics 25:339–344 Cited on p. 27.

Moran PAP (1984) The Monte Carlo evaluation of orthant probabilities for multivariate normal distributions. Australian Journal of Statistics 26:39–44 Cited on p. 13.

Moskowitz H, Tsai HT (1989) An error-bounded polynomial approximation for bivariate normal probabilities. Communications in Statistics − Simulation and Computation 18:1421–1437 Cited on p. 9.

Naiman DQ, Wynn HP (1992) Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via euler characteristics. The Annals of Statistics 20:43–76 Cited on p. 24.

Naiman DQ, Wynn HP (1997) Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. The Annals of Statistics 25:1954–1983 Cited on p. 24.

Naylor JC, Smith AFM (1982) Applications of a method for the efficient computation of posterior distributions. Journal of the Royal Statistical Society, Series C 31:214–225 Cited on p. 85.

Nelson PR (1991) Numerical evaluations of multivariate normal with correlations $\rho_{ij} = -\alpha_i \alpha_j$. In: Oeztuerk A (ed) The Frontiers of Statistical Scientific Theory and Industrial Applications. Proceedings of ICOSCO-I, American Sciences Press, Columbus, The First International Conference on Statistical Computing, vol II, pp 97–113 Cited on p. 17.

Ni Z, Kedem B (1999) On normal orthant probabilities. Chinese Journal of Applied Probability and Statistics 15:262–275 Cited on p. 13, 29.

Ni Z, Kedem B (2000) Normal orthant probabilities in the equicorrelation case. Journal of Mathematical Analysis and Applications 246:280–295 Cited on p. 13.

Niederreiter H (1972) On a number-theoretical integration method. Aequationes Mathematicae 8:304–311 Cited on p. 47.

Niederreiter H (1988) Low-discrepancy and low-dispersion sequences. Journal of Number Theory 30:51–70 Cited on p. 61.

Niederreiter H (1992) Random Number Generation and Quasi-Monte Carlo Methods. Society of Industrial and Applied Mathematics, Philadelphia Cited on p. 45, 46, 47, 48, 49.

Niederreiter H, Spanier J (2000) Monte Carlo and Quasi-Monte Carlo Methods 1998. Springer Verlag, Berlin Cited on p. 47.

Niederreiter H, Talay D (2006) Monte Carlo and Quasi-Monte Carlo Methods 2004. Springer Verlag, Berlin Cited on p. 47.

Niederreiter H, Xing C (1998) Nets, (t, s)-sequences and algebraic geometry. In: Hellekalek P, Larcher G (eds) Random and Quasi-Random Points Sets, Springer Verlag, New York, Lecture Notes in Statistics, vol 138, pp 267–302 Cited on p. 49, 61.

Niederreiter H, Hellekalek P, Larcher G, Zinterhof P (1998) Monte Carlo and Quasi-Monte Carlo Methods 1996. Springer Verlag, New York Cited on p. 46.

Ninomiya S, Tezuka S (1996) Toward real-time pricing of complex financial derivatives. Applied Mathematical Finance 3:1–20 Cited on p. 59.

Nuyens D, Cools R (2006a) Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel hilbert spaces. Mathematics of Computation 75:903–920 Cited on p. 47.

Nuyens D, Cools R (2006b) Fast component-by-component construction, a reprise for different kernels. In: Niederreiter H, Talay D (eds) Monte-Carlo and Quasi-Monte Carlo Methods 2004, Springer-Verlag, New York, pp 371–385 Cited on p. 47.

Ochi Y, Prentice RL (1984) Likelihood inference in a correlated probit regression model. Biometrika 71:531–543 Cited on p. 1.

Olson JM, Weissfeld LA (1991) Approximation of certain multivariate integrals. Statistics and Probability Letters 11:309–317 Cited on p. 27.

Owen AB (1992) Orthogonal arrays for computer experiments, integration and visualization. Statistica Sinica 2:439–452 Cited on p. 60.

Owen AB (1995) Randomly permuted (t, m, s)-nets and (t, s)-sequences. In: Niederreiter H, Shiue PJS (eds) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer Verlag, New York, Lecture Notes in Statistics, vol 106, pp 299–317 Cited on p. 49, 61.

Owen AB (1997a) Scrambled net variance for integrals of smooth functions. The Annals of Statistics 25:1541–1562 Cited on p. 49.

Owen AB (1997b) Monte Carlo variance of scrambled net quadrature. SIAM Journal on Numerical Analysis 34:1884–1910 Cited on p. 49.

Owen AB (2000) Monte Carlo, Quasi-Monte Carlo and randomized Quasi-Monte Carlo. In: Niederreiter H, Spanier J (eds) Monte Carlo and Quasi-Monte Carlo Methods 1998, Springer Verlag, Berlin, pp 86–97 Cited on p. 46.

Owen DB (1956) Tables for computing bivariate normal probabilities. The Annals of Mathematical Statistics 27:1075–1090 Cited on p. 7, 8, 9.

Owen DB (1965) A special case of a bivariate non-central t distribution. Biometrika 52:437–446 Cited on p. 4.

Owen DB (1985) Orthant probabilities. In: Kotz S, Johnson NL (eds) Encyclopedia of Statistical Sciences, Vol. 6, Wiley, New York, pp 521–523 Cited on p. 11.

Pandey MD (1998) An effective approximation to evaluate multinormal integrals. Structural Safety 20:51–67 Cited on p. 1, 68.

Patefield M, Tandy D (2000) Fast and accurate calculation of Owen's *t*-function. Journal of Statisitcal Software 5, available at http://www.jstatsoft.org Cited on p. 8, 62.

Pearson K (1901) Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. Philosophical Transactions of the Royal Society of London Series A 195:1–47 Cited on p. 1, 27.

Pearson K (1903) Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. Philosophical Transactions of the Royal Society of London Series A 200:1–66 Cited on p. 25.

Plackett RL (1954) A reduction formula for normal multivariate integrals. Biometrika 41:351–360 Cited on p. 1, 9, 10, 43.

Prékopa A (1995) Stochastic Programming. Kluwer, Dordrecht Cited on p. 1.

Ralston A, Rabinowitz P (1978) A First Course in Numerical Analysis. McGraw-Hill, New York Cited on p. 77.

Rice J, Reich T, Cloninger CR, Wette R (1979) An approximation to the multivariate normal integral: Its application to multifactorial qualitative traits. Biometrics 35:451–459 Cited on p. 26.

Richter WD (1994) A geometric approach to the Gaussian law. In: Mammitzsch V, Schneeweiß H (eds) Symposia Gaussiana, Walter de Gruyter, Berlin, pp 25–46 Cited on p. 29.

Richtmyer R (1951) The evaluation of definite integrals and a quasi-Monte Carlo method based on the properties of algebraic numbers. Tech. Rep. LA-1342, Los Alamos Scientific Laboratory, Los Alamos, New Mexico Cited on p. 47, 49.

Ripley BD (1987) Stochastic Simulation. Wiley, New York Cited on p. 41.

Rom D, Sarkar SK (1990) Approximating probability integrals of multivariate normal using association models. Journal of Statistical Computation and Simulation 35:109–119 Cited on p. 68.

Rosenblatt M (1952) Remarks on a multivariate transformation. The Annals of Mathematical Statistics 23:470–472 Cited on p. 31.

Royen T (1987) An approximation for multivariate normal probabilities of rectangular regions. Statistics 18:389–400 Cited on p. 25, 27.

Ruben H (1960) Probability contents in regions under spherical normal distributions I. The Annals of Mathematical Statistics 31:598–618 Cited on p. 13, 14, 15.

Ruben H (1961) Probability contents in regions under spherical normal distributions. III: The bivariate normal integral. The Annals of Mathematical Statistics 32:171–186 Cited on p. 8, 14.

Ruben H (1962) Probability contents in regions under spherical normal distributions. IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. The Annals of Mathematical Statistics 33:552–570 Cited on p. 13, 14.

Rudolfer SM, Watson PC (1993) Evaluation of orthant probabilities for singular bi- and trivariate normal distributions. Journal of Statistical Computation and Simulation 48:219–232 Cited on p. 58.

Sadefo Kamdem J (2005) Value-at-risk and expected shortfall for linear portfolios with elliptically distributed risk factors. International Journal of Theoretical and Applied Finance 8:537–551 Cited on p. 15.

Sadefo Kamdem J, Genz A (2008) Approximation of multiple integrals over hyperboloids with application to a quadratic portfolio with options. Computational Statistics and Data Analysis 52:3389–3407 Cited on p. 15.

Sándor Z, András P (2002) Alternative sampling methods for estimating multivariate normal probabilities. Journal of Econometrics 120:207–234 Cited on p. 49, 60, 61.

Schechter M (1998) Integration over a polyhedron: An application of the Fourier-Motzkin elimination method. American Mathematical Monthly 105:246–251 Cited on p. 40.

Schervish MJ (1984) Algorithm AS 195: Multivariate normal probabilities with error bound. Journal of the Royal Statistical Society, Series C 33:81–94, correction: **34** (1985) 103−104 Cited on p. 36, 54, 60, 61, 62.

Schläfli L (1858) On the multiple integral $\int^n dxdy\ldots dz$ whose limits are $p_1 = a_1x + b_1y + \ldots + h_1z > 0, p_2 > 0, ..., p_n > 0$ and $x^2 + y^2 + \ldots + z^2 < 1$. Quarterly Journal of Pure and Applied Mathematics 2:261–301 Cited on p. 1, 18, 39.

Shaw JEH (1988) A quasirandom approach to integration in Bayesian statistics. The Annals of Statistics 16:895–914 Cited on p. 1, 59, 79.

Sheil WF, O'Muircheartaigh (1977) The distribution of non-negative quadratic forms in normal variables. Journal of the Royal Statistical Society, Series C 26:92–98 Cited on p. 14, 66.

Sheppard WF (1900) On the calculation of the double integral expressing normal correlation. Transactions of the Cambridge Philosophical Society 19:23–26 Cited on p. 1, 9.

Siddiqui MM (1967) A bivariate t distribution. The Annals of Mathematical Statistics 38:162–166, correction: **38** (1967) 1594 Cited on p. 68.

Sloan IH (2002) QMC integration - Beating intractability by weighting the coordinate directions. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer Verlag, Berlin, pp 103–123 Cited on p. 39.

Sloan IH, Joe S (1994) Lattice Methods for Multiple Integration. Clarendon Press, Oxford Cited on p. 47, 48.

Sloan IH, Woźniakowski H (1998) When are Quasi-Monte Carlo algorithms efficient for high dimensional integrals? Journal of Complexity 14:1–33 Cited on p. 45.

Smolyak SA (1963) Quadrature and interpolation formulas for tensor products of certain classes of functions. Doklady Akademii Nauk SSSR 4:240–243 Cited on p. 52.

Sobol IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7:86–112 Cited on p. 49.

Solow AR (1990) A method for approximating multivariate normal orthant probabilities. Journal of Statistical Computation and Simulation 37:225–229 Cited on p. 26.

Somerville PN (1997) Multiple testing and simultaneous confidence intervals: Calculation of constants. Computational Statistics and Data Analysis 25:217–233 Cited on p. 29, 44.

Somerville PN (1998) Numerical computation of multivariate normal and multivariate t probabilities over convex regions. Journal of Computational and Graphical Statistics 7:529–544 Cited on p. 29, 44.

Somerville PN (2001) Numerical computation of multivariate normal and multivariate t probabilities over ellipsoidal regions. Journal of Statistical Software 6, available from http://www.jstatsoft.org Cited on p. 14, 30.

Soong WC, Hsu JC (1998) Using complex integration to compute multivariate normal probabilities. Journal of Computational and Graphical Statistics 6:397–415 Cited on p. 17.

Steck GP, Owen DB (1962) A note on the equicorrelated multivariate normal distribution. Biometrika 49:269–271 Cited on p. 17.

Steen MN, Byrne GO, Gelbard EM (1969) Gaussian quadratures for the integrals $\int_0^\infty \exp(-x^2)f(x)dx$ and $\int_0^b \exp(-x^2)f(x)dx$. Mathematics of Computation 23:661–671 Cited on p. 54.

Stern S (1992) A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60:943–952 Cited on p. 43, 60.

Stroud AH (1971) Approximate Calculation of Multiple Integrals. Prentice-Hall, Engelwood Cliffs, New Jersey Cited on p. 51.

Stuart A, Ord JK (1987) Kendall's Advanced Theory of Statistics, Vol. 1. Arnold, London Cited on p. 27.

Sun HJ (1988a) A general reduction method for n-variate normal orthant probability. Communications in Statistics − Theory and Methods 17:3913–3921 Cited on p. 12.

Sun HJ (1988b) A FORTRAN subroutine for computing normal orthant probabilities of dimensions up to nine. Communications in Statistics − Simulation and Computation 17:1097–1111 Cited on p. 12.

Sun HJ, Asano C (1989) Computing normal orthant probabilities with a Jacobi correlation matrix. Computational Statistics Quarterly 1:27–41 Cited on p. 12.

Szántai T (1986) Evaluation of a special multivariate gamma distribution function. Mathematical Programming Study 27:1–16 Cited on p. 60.

Szántai T (2001) Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function. Annals of Operations Research 100:85–101 Cited on p. 24, 43.

Tang B (1993) Orthogonal array-based latin hypercubes. Journal of the American Statistical Association 88:1392–1397 Cited on p. 60.

Tang LK, Melchers RE (1987) Improved approximation for multinormal integral. Structural Safety 4:81–93 Cited on p. 68.

Terza JV, Welland U (1991) A comparison of bivariate normal algorithms. Journal of Statistical Computation and Simulation 39:115–127 Cited on p. 8.

Tezuka S (2002) Quasi-Monte Carlo - Discrepancy between theory and practice. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer Verlag, Berlin, pp 124–140 Cited on p. 45.

Tomescu I (1986) Hypertrees and Bonferroni inequalities. Journal of Combinatorial Theory, Series B 41:209–217 Cited on p. 22, 23.

Tong YL (1980) Probability Inequalities in Multivariate Distributions. Academic Press, New York Cited on p. 24.

Tong YL (1990) The Multivariate Normal Distribution. Springer Verlag, New York Cited on p. 1, 24.

Tukey JW (1953) The problem of multiple comparisons, unpublished manuscript reprinted in: *The Collected Works of John W. Tukey, Volume 8,* 1994, H. I. Braun (Ed.), Chapman and Hall, New York Cited on p. 73.

van der Vaart HR (1953) The content of certain spherical polyhedra for any number of dimensions. Experientia 9:88–90 Cited on p. 1.

Vijverberg WPM (1996) Monte Carlo evaluation of multivariate student's t probabilities. Economics Letters 52:1–6 Cited on p. 31, 44.

Vijverberg WPM (1997) Monte Carlo evaluation of multivariate normal probabilities. Journal of Econometrics 76:281–307 Cited on p. 42.

Vijverberg WPM (2000) Rectangular and wedge-shaped multivariate normal probabilities. Economics Letters 68:13–20 Cited on p. 42.

Wang M, Kennedy WJ (1990) Comparison of algorithms for bivariate normal probability over a rectangle based on self-validated results form interval analysis. Journal of Statistical Computation and Simulation 37:13–25 Cited on p. 9, 54.

Wang M, Kennedy WJ (1992) A numerical method for accurately approximating multivariate normal probabilites. Computational Statistics and Data Analysis 13:197–210 Cited on p. 27, 54.

Wang O, Kennedy WJ (1997) Application of numerical interval analysis to contain self-validating results for multivariate probabilities in a massively parallel environment. Statistics and Computing 7:163–171 Cited on p. 27, 69.

Wang X, Hickernell FJ (2000) Randomized halton sequences. Mathematical and Computer Modelling 32:887–899 Cited on p. 61.

Wang Y, Hickernell FJ (2002) An historical overview of lattice point sets. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer Verlag, Berlin, pp 158–167 Cited on p. 47.

Wang YJ (1987) The probability integrals of bivariate normal distributions: A contingency table approach. Biometrika 74:1985–1990 Cited on p. 68.

Wang YJ (1997) Multivariate normal integrals and contingency tables with ordered categories. Psychometrika 62:267–284 Cited on p. 68.

Warnock T (1972) Computational investigations of low-discrepancy point sets. In: Zaremba SK (ed) Application of Number Theory to Numerical Analysis, Academic Press, New York, pp 319–343 Cited on p. 49.

Westfall P (1997) Multiple testing of general contrasts using logical constraints and correlations. Journal of the American Statistical Association 92:299–306 Cited on p. 44.

Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y (1999) Multiple Comparisons and Multiple Tests using the SAS System. SAS Institute Inc., Cary, NC Cited on p. 78.

Wijsman RA (1996) New algorithms for the function $T(h, a)$ of Owen, with application to bivariate normal and noncentral t-probabilities. Computational Statistics and Data Analysis 21:261–271 Cited on p. 8.

Worsley KJ (1982) An improved Bonferroni inequality and applications. Biometrika 69:297–302 Cited on p. 22, 76.

Woźniakowski H (2000) Efficiency of Quasi-Monte Carlo algorithms for high dimensional integrals. In: Niederreiter H, Spanier J (eds) Monte Carlo and Quasi-Monte Carlo Methods 1998, Springer Verlag, Berlin, pp 114–136 Cited on p. 39.

Yang ZQ (1998) Comparison of two Gaussian quadratures in multivariate normal integrals. Chinese Journal on Numerical Mathematics and Applications 20:59–69 Cited on p. 54.

Yang ZQ, Zhang CM (1997) Dimension reduction and L_1 approximation for evaluation of multivariate normal integral. Chinese Journal on Numerical Mathematics and Applications 19:82–95 Cited on p. 17.

Zaremba SK (1968) Some applications of multidimensional integration by parts. Annales Polonici Mathematici 21:85–96 Cited on p. 45.

Zellner A (1971) An Introduction to Bayesian Inference in Econometrics. Wiley, New York Cited on p. 31.

Zinterhof P (1969) Einige zahlentheoretische Methoden zur numerischen Quadratur und Interpolation. Sitzungberichte der Österreichischen Akademie der Wissenschaften, Math–Natur II 177:51–77 Cited on p. 47.

Index

Lecture Notes in Statistics

For information about Volumes 1 to 140, please contact Springer-Verlag

141: Peter Müller and Brani Vidakovic (Editors), Bayesian Inference in Wavelet Based Models. xiii, 394 pp., 1999.

142: György Terdik, Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach. xi, 258 pp., 1999.

143: Russell Barton, Graphical Methods for the Design of Experiments. x, 208 pp., 1999.

144: L. Mark Berliner, Douglas Nychka, and Timothy Hoar (Editors), Case Studies in Statistics and the Atmospheric Sciences. x, 208 pp., 2000.

145: James H. Matis and Thomas R. Kiffe, Stochastic Population Models. viii, 220 pp., 2000.

146: Wim Schoutens, Stochastic Processes and Orthogonal Polynomials. xiv, 163 pp., 2000.

147: Jürgen Franke, Wolfgang Härdle, and Gerhard Stahl, Measuring Risk in Complex Stochastic Systems. xvi, 272 pp., 2000.

148: S.E. Ahmed and Nancy Reid, Empirical Bayes and Likelihood Inference. x, 200 pp., 2000.

149: D. Bosq, Linear Processes in Function Spaces: Theory and Applications. xv, 296 pp., 2000.

150: Tadeusz Caliński and Sanpei Kageyama, Block Designs: A Randomization Approach, Volume I: Analysis. ix, 313 pp., 2000.

151: Håkan Andersson and Tom Britton, Stochastic Epidemic Models and Their Statistical Analysis. ix, 152 pp., 2000.

152: David Ríos Insua and Fabrizio Ruggeri, Robust Bayesian Analysis. xiii, 435 pp., 2000.

153: Parimal Mukhopadhyay, Topics in Survey Sampling. x, 303 pp., 2000.

154: Regina Kaiser and Agustín Maravall, Measuring Business Cycles in Economic Time Series. vi, 190 pp., 2000.

155: Leon Willenborg and Ton de Waal, Elements of Statistical Disclosure Control. xvii, 289 pp., 2000.

156: Gordon Willmot and X. Sheldon Lin, Lundberg Approximations for Compound Distributions with Insurance Applications. xi, 272 pp., 2000.

157: Anne Boomsma, Marijtje A.J. van Duijn, and Tom A.B. Snijders (Editors), Essays on Item Response Theory. xv, 448 pp., 2000.

158: Dominique Ladiray and Benoît Quenneville, Seasonal Adjustment with the X-11 Method. xxii, 220 pp., 2001.

159: Marc Moore (Editor), Spatial Statistics: Methodological Aspects and Some Applications. xvi, 282 pp., 2001.

160: Tomasz Rychlik, Projecting Statistical Functionals. viii, 184 pp., 2001.

161: Maarten Jansen, Noise Reduction by Wavelet Thresholding. xxii, 224 pp., 2001.

162: Constantine Gatsonis, Bradley Carlin, Alicia Carriquiry, Andrew Gelman, Robert E. Kass Isabella Verdinelli, and Mike West (Editors), Case Studies in Bayesian Statistics, Volume V. xiv, 448 pp., 2001.

163: Erkki P. Liski, Nripes K. Mandal, Kirti R. Shah, and Bikas K. Sinha, Topics in Optimal Design. xii, 164 pp., 2002.

164: Peter Goos, The Optimal Design of Blocked and Split-Plot Experiments. xiv, 244 pp., 2002.

165: Karl Mosler, Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach. xii, 280 pp., 2002.

166: Hira L. Koul, Weighted Empirical Processes in Dynamic Nonlinear Models, Second Edition. xiii, 425 pp., 2002.

167: Constantine Gatsonis, Alicia Carriquiry, Andrew Gelman, David Higdon, Robert E. Kass, Donna Pauler, and Isabella Verdinelli (Editors), Case Studies in Bayesian Statistics, Volume VI. xiv, 376 pp., 2002.

168: Susanne Rässler, Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches. xviii, 238 pp., 2002.

169: Yu. I. Ingster and Irina A. Suslina, Nonparametric Goodness-of-Fit Testing Under Gaussian Models. xiv, 453 pp., 2003.

170: Tadeusz Caliński and Sanpei Kageyama, Block Designs: A Randomization Approach, Volume II: Design. xii, 351 pp., 2003.

171: D.D. Denison, M.H. Hansen, C.C. Holmes, B. Mallick, B. Yu (Editors), Nonlinear Estimation and Classification. x, 474 pp., 2002.

172: Sneh Gulati, William J. Padgett, Parametric and Nonparametric Inference from Record-Breaking Data. ix, 112 pp., 2002.

173: Jesper Møller (Editor), Spatial Statistics and Computational Methods. xi, 214 pp., 2002.

174: Yasuko Chikuse, Statistics on Special Manifolds. xi, 418 pp., 2002.

175: Jürgen Gross, Linear Regression. xiv, 394 pp., 2003.

176: Zehua Chen, Zhidong Bai, Bimal K. Sinha, Ranked Set Sampling: Theory and Applications. xii, 224 pp., 2003.

177: Caitlin Buck and Andrew Millard (Editors), Tools for Constructing Chronologies: Crossing Disciplinary Boundaries, xvi, 263 pp., 2004.

178: Gauri Sankar Datta and Rahul Mukerjee , Probability Matching Priors: Higher Order Asymptotics, x, 144 pp., 2004.

179: D.Y. Lin and P.J. Heagerty (Editors), Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data, vii, 336 pp., 2004.

180: Yanhong Wu, Inference for Change-Point and Post-Change Means After a CUSUM Test, xiv, 176 pp., 2004.

181: Daniel Straumann, Estimation in Conditionally Heteroscedastic Time Series Models , x, 250 pp., 2004.

182: Lixing Zhu, Nonparametric Monte Carlo Tests and Their Applications, xi, 192 pp., 2005.

183: Michel Bilodeau, Fernand Meyer, and Michel Schmitt (Editors), Space, Structure and Randomness, xiv, 416 pp., 2005.

184: Viatcheslav B. Melas, Functional Approach to Optimal Experimental Design, vii., 352 pp., 2005.

185: Adrian Baddeley, Pablo Gregori, Jorge Mateu, Radu Stoica, and Dietrich Stoyan, (Editors), Case Studies in Spatial Point Process Modeling, xiii., 324 pp., 2005.

186: Estela Bee Dagum and Pierre A. Cholette, Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series, xiv., 410 pp., 2006.

187: Patrice Bertail, Paul Doukhan and Philippe Soulier, (Editors), Dependence in Probability and Statistics, viii., 504 pp., 2006.

188: Constance van Eeden, Restricted Parameter Space Estimation Problems, vi, 176 pp., 2006.

189: Bill Thompson, The Nature of Statistical Evidence, vi, 152 pp., 2007.

190: Jérôme Dedecker, Paul Doukhan, Gabriel Lang, José R. León, Sana Louhichi Clémentine Prieur, Weak Dependence: With Examples and Applications, xvi, 336 pp., 2007.

191: Vlad Stefan Barbu and Nikolaos Limnios, Semi-Markov Chains and Hidden Semi-Markov Models toward Applications, xii, 228 pp., 2007.

192: David B. Dunson, Random Effects and Latent Variable Model Selection, 2008

193: Alexander Meister. Deconvolution Problems in Nonparametric Statistics, 2008.

194: Dario Basso, Fortunato Pesarin, Luigi Salmaso, Aldo Solari, Permutation Tests for Stochastic Ordering and ANOVA: Theory and Applications with R, 2009.

195: Frank Bretz and Alan Genz,, Computation of Multivariate Normal and *t* Probabilities, viii, 124 pp., 2009