ON SPHERICAL MONTE CARLO SIMULATIONS FOR MULTIVARIATE NORMAL PROBABILITIES

HUEI-WEN TENG,* ** *National Central University*

MING-HSUAN KANG,*** *National Chiao Tung University*

CHENG-DER FUH,* *National Central University*

Abstract

The calculation of multivariate normal probabilities is of great importance in many statistical and economic applications. In this paper we propose a spherical Monte Carlo method with both theoretical analysis and numerical simulation. We start by writing the multivariate normal probability via an inner radial integral and an outer spherical integral using the spherical transformation. For the outer spherical integral, we apply an integration rule by randomly rotating a predetermined set of well-located points. To find the desired set, we derive an upper bound for the variance of the Monte Carlo estimator and propose a set which is related to the kissing number problem in sphere packings. For the inner radial integral, we employ the idea of antithetic variates and identify certain conditions so that variance reduction is guaranteed. Extensive Monte Carlo simulations on some probabilities confirm these claims.

Keywords: Spherical; simulation; variance reduction; sphere packings; kissing number; lattice

2010 Mathematics Subject Classification: Primary 11K45; 65C05
  Secondary 65C60; 91G60; 91G70

1. Introduction

Efficient and precise calculation for the multivariate normal probability is of critical importance in many disciplines. To name a few, the multinomial probit model used in econometrics and biometrics has cell probabilities that are negative orthant probabilities. In the financial industry implementation of the CreditMetrics[TM] model, the joint migration probabilities in credit migration model are rectangle probabilities for bivariate normal distributions (cf. Gupton *et al.* (1997)). The estimation of value at risk for risk management considered in Glasserman *et al.* (2000) requires calculation of the multivariate probability of an ellipsoid. In multiple comparisons, multivariate normal probabilities are also considered in Hsu (1996). For more examples and applications; see Genz and Bretz (2009) for a recent summary of multivariate normal distribution and multivariate t-distribution.

Motivated by these applications, we investigate efficient calculation for the multivariate normal probability. That is, for an indicator function $\mathbf{1}_{\mathcal{A}}(x)$ with a support set $\mathcal{A}$ in $\mathbb{R}^d$,

we seek more efficient computation of the following probability:

$$p = \mathbb{P}\{X \in \mathcal{A}\} = \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{A}}(x) \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-(1/2)(x-\mu)^\top \Sigma^{-1}(x-\mu)} \, dx \quad \text{for } \mathcal{A} \subset \mathbb{R}^d, \quad (1)$$

where $^\top$ denotes the transpose, $X = (X_1, \ldots, X_d)^\top$ is a d-dimensional nonsingular multivariate normal distribution with mean vector μ and covariance matrix Σ, denoted by $X \sim N_d(\mu, \Sigma)$, and $\det(\cdot)$ denotes the determinant of a matrix.

Standard approaches of calculating (1) include classical analytic approximation, numerical integration, and the Monte Carlo method. Instead of using analytic approximation and the numerical method (cf. Miwa *et al.* (2003) and Craig (2008)), which are more suitable for low-dimensional cases, in this paper, we study the Monte Carlo method. Although the Monte Carlo method is easy to implement and can overcome the curse of dimensionality. Its convergence rate is rather slow, proportional to $1/\sqrt{M}$, where M is the Monte Carlo sample size (cf. Asmussen and Glynn (2007, p. 69)). Therefore, additional variance reduction methods are required. Typical methods for variance reduction include antithetic variates, Latin hypercube sampling, and primitive Monte Carlo methods (see, for example, Genz (1992), (1993), Hajivassiliou *et al.* (1996), Vijverberg (1997), and Genz and Bretz (2002)). A comparison study of alternative sampling methods can be found in Sándor and András (2004). For a survey on existing methods, see Genz and Bretz (2009).

Monte Carlo methods using a spherical transformation have been studied in the literature. For example, Deák (1980), (2000) used this transformation as the basis for calculating multivariate normal probabilities, and Fang and Wang (1994) proposed a transformation on the unit hypercube to generate points uniformly distributed on the unit sphere. Monahan and Genz (1997) proposed a Monte Carlo simulation method for Bayesian computation, to which the authors used randomized extended simplex design for the spherical integral and Simpson weights for the radial integral. The method proposed by Somerville (2001) can also be regarded as a spherical Monte Carlo method, to which a quadrature rule is used for the radial integral.

In this paper we propose a spherical Monte Carlo method with both theoretical analysis and numerical simulation. There are two aspects in this study. First, because the spherical integral requires generating unit vectors uniformly on the unit sphere, one way to improve the Monte Carlo efficiency is to use a randomly rotated predetermined set of unit vectors. For this purpose, we provide a criterion to select such a set on the unit sphere, which involves the minimal distance of any two points in the set and the cardinality of the set. Among all sets in which the minimal distance of any two points is the same, the desired optimal set is the one with maximal cardinality. Especially when the minimal distance equals 1, finding the desired optimal set is linked to the kissing number problem in sphere packings.

Next, for the radial integral, we offer a variance reduction technique employing the idea of antithetic variates. For this purpose, we categorize a set, which could be a set of basis points on the unit sphere or an event of simulation in $\mathbb{R}^d$ by using the idea of *central symmetry* and *central antisymmetry*. To the best of our knowledge, this seems to be the first attempt to provide sufficient conditions on antithetic variates for a spherical Monte Carlo method. To illustrate the proposed method, simulation studies are given for orthant, rectangle, and ellipsoid probabilities for multivariate normal distributions. The simulation results confirm these claims.

The rest of this paper is organized as follows. In Section 2 we propose the spherical Monte Carlo method with antithetic variates. In Section 3 we link the proposed set of basis points to a sphere packing problem and spherical t-designs, and discuss the practical implementation in high-dimensional cases. Simulation results are given in Section 4, in which we compare

the efficiency of various spherical Monte Carlo estimators, and demonstrate that neither of the proposed spherical Monte Carlo estimator nor the Geweke–Hajivassiliou–Keane (GHK) simulator dominates. Hence, the proposed method is more applicable because it is capable of calculating multivariate normal probabilities of any type of region. We conclude with Section 5. All proofs are deferred to Appendices A and B.

2. The proposed spherical Monte Carlo method

In Section 2.1 we formulate the problem, in Section 2.2 we present the spherical integral, and in Section 2.3 we study the radial integral.

2.1. Problem formulation

Let $\mathbf{1}_{\mathcal{A}}: \mathbb{R}^d \to \mathbb{R}$ be an indicator function with a support set $\mathcal{A}$. Denote the probability density function (PDF) of a d-variate normal random variable $N_d(\mu, \Sigma)$ by

$$\phi(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-(1/2)(x-\mu)^\top \Sigma^{-1}(x-\mu)},$$

where μ is the mean vector and Σ is the variance-covariance matrix. The desired multivariate normal probability of the region $\mathcal{A}$ is an integral of the form in (1).

Let '$\sim$' denote distributed as. Note that a d-variate normal random variable $X \sim N_d(\mu, \Sigma)$ can be expressed by $X = \mu + \Gamma Z$, where Z is a d-variate standard normal random variable, and Γ is the lower triangle matrix such that $\Sigma = \Gamma\Gamma^\top$, the so-called Cholesky decomposition of Σ. By change of variables, $x = \mu + \Gamma z$, we write (1) as

$$p = \int_{\mathbb{R}^d} \mathbf{1}_{\tilde{\mathcal{A}}}(z)\phi(z)\,\mathrm{d}z, \tag{2}$$

where $\phi(\cdot)$ is the standard normal density and $\tilde{\mathcal{A}} = \Gamma^{-1}(\mathcal{A} - \mu)$. With abuse of notation, we denote $\tilde{\mathcal{A}}$ by $\mathcal{A}$ throughout the rest of the paper. It is straightforward to see that the crude Monte Carlo estimator is

$$\hat{p} = \mathbf{1}_{\mathcal{A}}(Z),$$

and the crude Monte Carlo estimator with antithetic variates is

$$\hat{p}_{\mathrm{AT}} = \frac{\mathbf{1}_{\mathcal{A}}(Z) + \mathbf{1}_{\mathcal{A}}(-Z)}{2}.$$

By using the spherical transformation, a point $z \in \mathbb{R}^d$ can be written as $(r, \boldsymbol{u})$, where r is the radius and $\boldsymbol{u}$ is the unit vector of z. Then, $\mathbf{1}_{\mathcal{A}}(z) = \mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u})$ and (2) can be expressed as

$$p = \int_{S^{d-1}} \int_0^\infty \mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u})\left(\frac{1}{\sqrt{(2\pi)^d}} e^{-r^2/2} r^{d-1}\right) \mathrm{d}r\,\mathrm{d}A = \frac{1}{\mathrm{area}(S^{d-1})} \int_{S^{d-1}} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u},$$

where $\mathrm{d}A$ denotes the differential area on the unit sphere S^{d-1}. Denote the χ-distribution with degrees of freedom d by $\chi(d)$. Then the inner radial integral is

$$f(u) = \mathrm{area}(S^{d-1}) \int_0^\infty \mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u})\left(\frac{1}{\sqrt{(2\pi)^d}} e^{-r^2/2} r^{d-1}\right) \mathrm{d}r = \int_0^\infty \mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u}) k_d(r)\,\mathrm{d}r.$$

Here, $\mathrm{area}(S^{d-1}) = 2\sqrt{\pi^d}/\Gamma(d/2)$ is the surface area of S^{d-1}, and $k_d(\cdot)$ is the PDF of $\chi(d)$.

A random unit vector $\boldsymbol{u}$ in $\mathbb{R}^d$ is a vector uniformly distributed on S^{d-1}, written as $\boldsymbol{u} \sim U(S^{d-1})$. Therefore, generating a sample from a standard normal distribution z is equivalent to independently generating a radius $r \sim \chi(d)$ and a unit vector $\boldsymbol{u} \sim U(S^{d-1})$, and setting $z = r\boldsymbol{u}$.

To write (2) using the spherical transformation, an alternative approach is to consider the spherical integral as the innermost integral as in Monahan and Genz (1997). Here we take the radial integral as the innermost integral because the radial integral is of one dimension and its calculation is simple. In the case of a simple region $\mathcal{A}$, such as rectangles, orthants, and ellipsoids, when the unit vector $\boldsymbol{u}$ is fixed, the inner radial integral has a closed-form formula in terms of the cumulative distribution function (CDF) of a χ-distribution; see Deák (2000). For a general region, the inner radial integral can be calculated via Monte Carlo methods or numerical methods; see Davis and Rabinowitz (1984). Numerical quadrature methods produce biased estimators in general and are beyond the scope of this paper. Here we focus only on Monte Carlo methods.

The outer spherical integral using Monte Carlo methods requires generating a random unit vector $\boldsymbol{u}$. A straightforward method to sample $\boldsymbol{u}$ is to generate d-independent standard normal random variables to have a vector in $\mathbb{R}^d$ and then normalize the vector by its length. A more efficient algorithm to sample $\boldsymbol{u}$ can be found in Fang and Wang (1994) by generating just $(d-1)$ random numbers.

2.2. Variance reduction on the spherical integral

One way to obtain an efficient spherical Monte Carlo estimator for the spherical integral is taking the average value of a randomly rotated predetermined finite set of unit vectors, denoted by V. More precisely, let $O(d)$ be the orthogonal group of $\mathbb{R}^d$ consisting of all $d \times d$ matrices $\boldsymbol{D}$ satisfying that $\boldsymbol{D}\boldsymbol{D}^{\top}$ equals the identity matrix. There is a unique probability measure on $O(d)$, which is invariant under left multiplication by all elements of $O(d)$ and a random orthogonal matrix $\boldsymbol{T}$, which is uniformly distributed on $O(d)$ with respect to this probability measure, denoted by $\boldsymbol{T} \sim U(O(d))$. Then randomly rotating a set V simply means to apply a random orthogonal matrix $\boldsymbol{T}$ to every element of V.

The standard algorithm to generate a random orthogonal matrix is described as follows; see Heiberger (1978). First, generate a random $d \times d$ matrix whose entries are independently standard normal random variables. Then, apply the Gram–Schmidt method to the column vectors to obtain the desired random orthogonal matrix. More efficient algorithms to generate a random orthogonal matrix using only $(d-1)(d+2)/2$ standard normal random variables can be found in Stewart (1980), Diaconis and Shahshahani (1987), and Anderson *et al.* (1987).

One important feature of the random orthogonal matrix is that if we fix a unit vector and let a random orthogonal matrix act on it, then the resulting vector is uniformly distributed on S^{d-1}. In other words, for a continuous function $h(\boldsymbol{u})$ on S^{d-1} and any unit vector $\boldsymbol{v} \in S^{d-1}$, we have

$$\frac{1}{\text{area}(S^{d-1})} \int_{S^{d-1}} h(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u} = \int_{O(d)} h(\boldsymbol{T}\boldsymbol{v}) \, \mathrm{d}\boldsymbol{T}. \tag{3}$$

Here, $\mathrm{d}\boldsymbol{T}$ is the unique left-invariant probability measure on $O(d)$ as mentioned above.

A spherical Monte Carlo estimator using V is

$$\hat{p}^V = \frac{1}{|V|} \sum_{\boldsymbol{v} \in V} \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{T}\boldsymbol{v}), \tag{4}$$

where the r_vs are independent and identically distributed as $\chi(d)$ for $v \in V$, $T \sim U(O(d))$, and the r_vs and T are independent.

When the innermost radial integral can be calculated explicitly, i.e. $f(u)$ can be expressed in terms of the CDF of $\chi(d)$, an estimator using V is

$$\hat{p}_*^V = \frac{1}{|V|} \sum_{v \in V} f(Tv),$$

where $T \sim U(O(d))$. By using the fact of conditioning, indeed, $f(Tv)$ equals $\mathbb{E}[\mathbf{1}_{\mathcal{A}}(r_v, Tv) \mid T]$, and, thus, enjoys smaller variance compared with $\hat{p}^V$.

To have an efficient simulation, the crucial step involves the selection of V to minimize the variance of $\hat{p}^V$. For this purpose, we propose a criterion of determining the desired set V and in Section 3 provide a feasible solution that relates to the maximal kissing number problem in sphere packings.

To characterize the variance of the estimator for the spherical integral for a finite subset V of the unit sphere S^{d-1}, we need to define some notation as follows. First, denote $d(\cdot, \cdot)$ as the usual Euclidean distance function on $\mathbb{R}^d$, and let

$$d_{\min}(V) = \min\{d(v, v^\top) \mid v, v^\top \in V, v \neq v^\top\}$$

be the minimal distance of any two points in V. For a region $\mathcal{A}$ in S^{d-1}, define the diameter of $\mathcal{A}$ as

$$\operatorname{diam}(\mathcal{A}) = \sup\{d(x, x^\top) \mid x, x^\top \in \mathcal{A}\}.$$

It is easy to see that when $\operatorname{diam}(\mathcal{A}) < d_{\min}(V)$, the intersection of V and A is either the empty set or a single point.

Let π be the normalized measure on S^{d-1} induced from the Lebesgue measure on $\mathbb{R}^d$ so that $\pi(S^{d-1}) = 1$. Note that π is a probability measure and is invariant under the action of $O(d)$. For a region $\mathcal{A}$ in S^{d-1}, define an estimator on $O(d)$ as

$$g_{\mathcal{A}}^V(T) = \sum_{v \in V} \frac{1}{|V|} \mathbf{1}_{\mathcal{A}}(Tv), \tag{5}$$

where $T \sim U(O(d))$. Suppose that $\operatorname{diam}(\mathcal{A}) < d_{\min}(V)$. Let TV denote the set consisting of Tv for all $v \in V$. Since orthogonal matrices preserve the Euclidean distance function, we have $d_{\min}(TV) = d_{\min}(V)$ and $|\mathcal{A} \cap TV| = 1$ or 0. On the other hand, the estimator $g_{\mathcal{A}}^V(T)$ is unbiased. Therefore, we have

$$g_{\mathcal{A}}^V(T) = \begin{cases} \dfrac{1}{|V|} & \text{with probability } |V|\pi(\mathcal{A}), \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\operatorname{diam}(\mathcal{A}) < d_{\min}(V)$ implies $\pi(\mathcal{A}) < 1/|V|$. Moreover,

$$\operatorname{var}(g_{\mathcal{A}}^V(T)) = \left(\frac{1}{|V|}\right)^2 |V|\pi(\mathcal{A}) - \pi^2(\mathcal{A}) = \frac{\pi(\mathcal{A})}{|V|} - \pi^2(\mathcal{A}). \tag{6}$$

If $\operatorname{diam}(\mathcal{A}) \geq d_{\min}(V)$, we can decompose $\mathcal{A}$ and obtain the following result.

Theorem 1. *Let $\mathcal{A}$ be a region of S^{d-1}, which can be written as disjoint union of the subregions $\mathcal{A}_1, \ldots, \mathcal{A}_N$ so that $\mathrm{diam}(\mathcal{A}_i) < d_{\min}(V)$ for all $i = 1, \ldots, N$. Then the variance of $g_{\mathcal{A}}^V(\boldsymbol{T})$ satisfies the following inequality:*

$$\mathrm{var}(g_{\mathcal{A}}^V(\boldsymbol{T})) \leq \pi(\mathcal{A}) - \pi^2(\mathcal{A}) + \left(\frac{N}{|V|} - 1\right)\pi(\mathcal{A}).$$

The proof of Theorem 1 is given in Appendix A.

When N is the minimal number such that the decomposition condition in Theorem 1 holds, we denote the upper bound in Theorem 1 by $C(V, \mathcal{A})$ for short. Note that $C(V, \mathcal{A})$ reduces to (6), the exact variance, when $N = 1$ in Theorem 1. In the general case, if $N < |V|$, $C(V, \mathcal{A})$ is less than $\pi(\mathcal{A}) - \pi^2(\mathcal{A})$, the variance of the crude Monte Carlo estimator $\hat{p}$. Next by using Theorem 1, we immediately have the following corollary.

Corollary 1. *For a given two finite subsets V and $V^\top$ of S^{d-1}. If one of the following two conditions holds:*

(i) *$|V| = |V^\top|$ and $d_{\min}(V) > d_{\min}(V^\top)$;*

(ii) *$d_{\min}(V) = d_{\min}(V^\top)$ and $|V| > |V^\top|$,*

then $C(V, \mathcal{A}) \leq C(V^\top, \mathcal{A})$ for any region $\mathcal{A}$ in S^{d-1}.

Fix a given region $\mathcal{A}$, the upper bound $C(V, \mathcal{A})$ depends only on the minimal distance between any two points in V, and the cardinality of the set V. Corollary 1 suggests two approaches to minimize the upper bound of the variance of the estimator $\hat{p}^V$ defined in (4).

Case 1: Among all sets V with the same cardinality, the set with maximal $d_{\min}(V)$ minimizes $C(V, \mathcal{A})$.

Case 2: Among all sets V with the same $d_{\min}(V)$, the set with maximal cardinality minimizes $C(V, \mathcal{A})$.

Feasible approaches to construct such a set V in either case are available. Based on case 1, for $|V| \leq 130$ in three to five dimensions, sets, and known maximal $d_{\min}(V)$ are reported in Sloane. Other than the method based on case 1, here we provide a solution based on case 2.

Note that in the case of $d_{\min}(V) = 2$, V must have exactly two points, since the distance between any two unit vectors is less than or equal to 2, and the equality holds only when they form an antipodal pair. This shows that the method of antithetic variates is optimal under the above criterion for $d_{\min}(V) = 2$. In the case of $d_{\min}(V) = 1$, constructing a set with maximal cardinality is related to the kissing number problem in sphere packings. Details of this linkage will be discussed in Section 3.1.

2.3. Variance reduction on the radial integral

The efficiency of simulators may be improved by using antithetic variates. A straightforward application of antithetic variates on the unit sphere can be found in Deák (1980). Here we provide an alternative approach by applying the idea of antithetic variates for a set of points on the unit sphere. For this purpose, we first introduce the concepts of central symmetry and central antisymmetry. A set V is called *centrally symmetric* if for all $\boldsymbol{v}$ in V, $-\boldsymbol{v}$ also lies in V. In this case, we can decompose V as the disjoint union of V^+ and $-V^+$. Here, V^+ consists of all vectors in V whose first nonzero coordinates are positive. It is easy to see that the property of centrally symmetric is preserved under the action of $O(d)$. Motivated by the idea of antithetic

variates, we propose a spherical Monte Carlo estimator using a centrally symmetric set V on S^{d-1} with antithetic variates:

$$\hat{p}_{\text{AT}}^V = \frac{1}{|V|} \sum_{\boldsymbol{v} \in V^+} \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{T}\boldsymbol{v}) + \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, -\boldsymbol{T}\boldsymbol{v}), \tag{7}$$

where the $r_{\boldsymbol{v}}$s are independent and identically distributed as $\chi(d)$ for $\boldsymbol{v} \in V$, $\boldsymbol{T} \sim U(O(d))$, and the $r_{\boldsymbol{v}}$s and $\boldsymbol{T}$ are independent. We remark that the proposed set V generated by the shortest nonzero vectors of a lattice described in Section 3 is centrally symmetric, therefore $\hat{p}_{\text{AT}}^V$ in (7) is well defined.

It is known that certain conditions are required in Cartesian coordinates to ensure antithetic variates enjoy lower variance; see Ross (2013). Likewise, an additional condition is needed in the spherical case. A set $\mathcal{A}$ is called *centrally antisymmetric* if $\mathcal{A}$ and $(-\mathcal{A})$ are disjoint. It is easy to see that the property of centrally antisymmetric is also preserved under the action of $O(d)$. When a set is neither centrally symmetric nor centrally antisymmetric, it is called a mixed set. As a result, we have the following theorem.

Theorem 2. *If the set V on S^{d-1} is centrally symmetric and the event $\mathcal{A}$ in $\mathbb{R}^d$ is centrally antisymmetric, then* $\text{var}(\hat{p}_{\text{AT}}^V) \leq \text{var}(\hat{p}^V)$.

The proof of Theorem 2 is given in Appendix B.

In Theorem 2 we show that the variance of $\hat{p}_{\text{AT}}^V$ can be reduced when V is centrally symmetric and $\mathcal{A}$ is centrally antisymmetric. As will be shown in Section 4.5, for certain central antisymmetric sets, although $\text{var}(\hat{p}_{\text{AT}}^V)$ equals $\text{var}(\hat{p}^V)$ as a mathematical fact, $\hat{p}_{\text{AT}}^V$ is preferred because it requires fewer independently drawn radii and thus enjoys lower computational cost compared with $\hat{p}^V$. For demonstration purposes, in Section 4.5 we provide an explicit central antisymmetric set, for which $\hat{p}_{\text{AT}}^V$ enjoys smaller variance than $\hat{p}^V$.

3. Practical implementation

In Section 3.1 we explore the link of an optimal V based on case 2 in Corollary 1 to the kissing number problem in sphere packings, and then make a connection to spherical t-designs. In Section 3.2 we discuss some possible approaches in high-dimensional cases.

3.1. The linkage to the kissing number problem in sphere packings

For a sphere packing, a kissing number is defined as the number of nonoverlapping spheres that can be arranged such that they each touch another given sphere. Here, all spheres must have the same radius. The kissing number problem seeks the maximal possible kissing number in a sphere packing. In Figure 1 we illustrate two sphere arrangements in $\mathbb{R}^2$ with a kissing number of 4 and 6, respectively. Moreover, the latter has the maximal kissing number in two dimensions.

Now for a finite subset V of S^{d-1} with $d_{\min}(V) = 1$, we can construct an associated sphere packing from V as follows. Consider all spheres centred at $\boldsymbol{v} \in V$ with radius $\frac{1}{2}$ and one sphere centred at the origin with radius $\frac{1}{2}$. It is easy to see that the number $|V|$ equals the kissing number of this sphere packing, and, as a result, to find a V with maximal $|V|$ with $d_{\min}(V) = 1$ is exactly the same as finding a sphere packing producing the maximal kissing number.

The next question is how to obtain a sphere arrangement with maximal kissing number. This is a difficult question in geometry with partial solutions, to which we can directly utilize existing results. Note that most of the known maximal kissing numbers in sphere packings comes from the study of lattices. Given a basis of $\mathbb{R}^d$, a *lattice* L is defined as the set of all integral linear
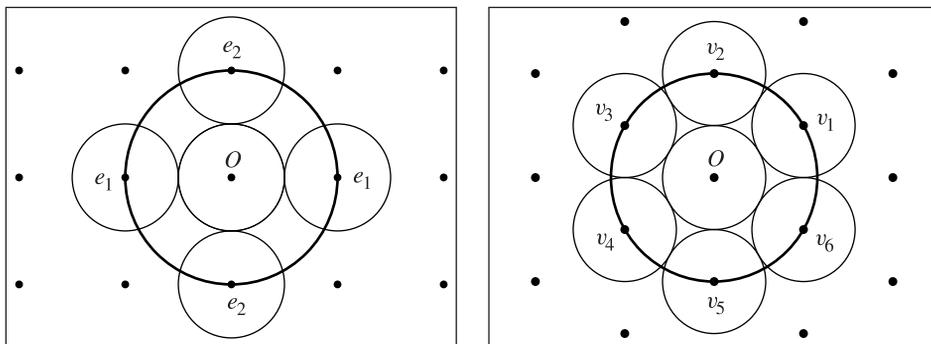
FIGURE 1: Lattice $\mathbb{Z}^2$ and the corresponding sphere arrangement with a kissing number of 4 (*left*). Lattice A_2 and the corresponding sphere arrangement with a kissing number of 6 (*right*).

combinations of the basis. Let $d_{\min}$ be the shortest distance among all pairs of points in L. Then we can obtain a sphere arrangement by setting spheres with radius $\frac{1}{2}d_{\min}$ centred at all points of L. In this case, every sphere kisses the same number of spheres. Especially for the sphere centred at the origin, the number of spheres kissing it equals the number of shortest nonzero vectors in L. In short, we simply call the shortest nonzero vectors by shortest vectors in the rest of the paper. To obtain V from a lattice L, we simply collect all the shortest vectors and normalize them to unit vectors, and denote it by V_L.

For example, the integer lattice $\mathbb{Z}^d$ is generated by the standard basis $\{e_1, \ldots, e_d\}$ and consists of all integer points. In this case, the kissing number of $\mathbb{Z}^d$ equals $2d$, and the set of shortest vectors is

$$V_{\mathbb{Z}^d} = \{\pm e_1, \ldots, \pm e_d\}.$$

Another example considers the lattice A_2, which is generated by the basis $\boldsymbol{v}_1 = (\sqrt{3}/2, \frac{1}{2})$ and $\boldsymbol{v}_2 = (0, 1)$. Moreover, we obtain

$$V_{A_2} = \left\{\left(\pm\frac{\sqrt{3}}{2}, \pm\frac{1}{2}\right), (0, \pm 1)\right\}.$$

Note that the kissing number of A_2 is 6, which is indeed the maximal kissing number in two dimensions. In Figure 1 we depict these two lattices and associated sphere arrangements in two dimensions.

In general, it is difficult to prove if a kissing number is maximal in any arbitrary dimension; see Conway and Sloane (1999). For d from 2 to 8, lattices that produce the known maximal kissing number are the lattices A_2, A_3, D_4, D_5, E_6, E_7, and E_8. These lattices are called root lattices, which comes from semisimple Lie algebra and these symbols stand for the name of the corresponding Lie algebra. Explicit construction of these lattices can be found in Conway and Sloane (1999). We can thus form the desired V from these lattices by taking the normalized shortest vectors in these lattices.

It is worth mentioning that the set V_L above also forms a spherical t-design for some $t > 0$. That is

$$\frac{1}{\text{area}(S^{d-1})} \int_{S^{d-1}} g(\boldsymbol{u}) \, d\boldsymbol{u} = \frac{1}{|V_L|} \sum_{\boldsymbol{v} \in V_L} g(\boldsymbol{v})$$

TABLE 1: The lattice $L_d^{\max}$, the cardinality of its normalized shortest vectors V, and the associated t of V as a spherical t-design in d-dimensions.

							d		
	2	3	4	5	6	7	8	16	24
$L_d^{\max}$	A_2	A_3	D_4	D_5	E_6	E_7	E_8	Barnes–Wall lattice	Leech lattice
$|V|$	6	12	24	40	72	126	240	4320	196 560
t	5	3	5	3	5	5	7	7	11

for all real-valued continuous functions $g(\boldsymbol{u})$ on the unit sphere, which are restrictions of polynomial functions of degree less than or equal to t in $\mathbb{R}^d$. This means if the function $g(\boldsymbol{u})$ is a polynomial function of degree less than or equal to t, its integral over the unit sphere S^{d-1} is simply an average of function values of these points. To some extent, elements in V_L behave as a quadrature point that allows the integral of polynomials equal to a finite sum of polynomial values. On the other hand, if the function $g(\boldsymbol{u})$ can be approximated by a polynomial well, then its integral can be approximated as an average of these functional values. Therefore, choosing V_L as the sample set provides a good approximation of the desired integral. In other words, under the multivariate normal probability density, the spherical Monte Carlo estimator $g^{V_L}(\boldsymbol{T}) := \sum_{\boldsymbol{v} \in V_L} (1/|V_L|) g(\boldsymbol{T}\boldsymbol{v})$ produces 0 variance for estimating a polynomial function of degree less than or equal to t in $\mathbb{R}^d$, and, hence, enjoys a small variance if $g(\cdot)$ can be well approximated by a polynomial of degree less than or equal to t in $\mathbb{R}^d$. In Table 1 we list the cardinality of each V_L from the corresponding lattices in d-dimensions and the associated t for the proposed V_L as a spherical t-design.

3.2. High-dimensional cases

In principle, our proposed method is feasible in higher dimensions once a suitable V is selected. Theorem 1 suggests the use of V with the maximal cardinality given $d_{\min}(V) = 1$. The construction of such a V relates to the maximal kissing number problem in sphere packings, and can be accomplished via a lattice in some cases. The lattice with the maximal presently known kissing number, denoted by $L_d^{\max}$, in dimension up to 40 can be found in Nebe and Sloane. For demonstration purposes, we report results in dimension from 2 to 8 in Section 4.4.

Nevertheless, some difficulties arise in constructing V from the lattice $L_d^{\max}$. Not all of the optimal lattices are well-studied lattices so that their shortest vectors cannot be constructed generically. Moreover, the size of V may be too large for practical implementation in high-dimensional cases. For example, the maximal kissing number is 196 560 in 24 dimensions, which seems to be formidable for practical implementation.

In high-dimensional cases, a simple (but not optimal based on our criterion in Theorem 1) approach is to use a certain family of lattices. For example, we consider three lattices, including $\mathbb{Z}_d$, D_d, and A_d in this paper, which are defined as follows. Firstly, $\mathbb{Z}_d$ is generated by the standard basis $e_1, \ldots, e_d$ of $\mathbb{R}^d$ and $V_{\mathbb{Z}_d} = \{\pm e_1, \ldots, \pm e_d\}$. Secondly, D_d is generated by $(e_1 - e_2), (e_2 - e_3), \ldots, (e_{d-1} - e_d)$, and $(e_{d-1} + e_d)$. Thirdly, V_{D_d} consists of all permutations of $(1/\sqrt{2})(\pm 1, \pm 1, 0, \ldots, 0)$ and there are $2d(d-1)$ such vectors. Finally, identify $\bar{\mathbb{R}}^d$ as the hyperplane $x_1 + \cdots + x_{d+1} = 0$ in $\mathbb{R}^{d+1}$. Then A_d is generated by $(e_1 - e_2), (e_2 - e_3), \ldots, (e_d - e_{d+1})$, so that V_{A_d} consists of all permutations of $(1/\sqrt{2})(1, -1, 0, \ldots, 0) \in \bar{\mathbb{R}}^{d+1}$ and there are $d(d+1)$ such vectors. Therefore, spherical packings generating by lattices $\mathbb{Z}_d$, A_d, and D_d have kissing numbers of $2d$, $d(d+1)$, and $2d(d-1)$, respectively.

TABLE 2: The cardinality of V arising from various lattices, variance ratio (VR), and penalized variance ratio (PVR) by cardinality of $\hat{p}_*^V$ for calculating $\mathbb{P}\{Z \in [-\infty, 1]^d\}$ in d-dimensions. The Monte Carlo sample size is 100.

	Cardinality				VR				PVR-cardinality			
d	$\mathbb{Z}_d$	A_d	D_d	$L_d^{\max}$	$\mathbb{Z}_d$	A_d	D_d	$L_d^{\max}$	$\mathbb{Z}_d$	A_d	D_d	$L_d^{\max}$
16	32	272	480	4320	116.3	840.9	1248.3	40 650.8	3.6	3.1	2.6	9.4
24	48	600	1104	1.9×10^5	146.9	1296.7	1972.0	1.5×10^6	3.1	2.2	1.8	7.8

An additional advantage of using these lattices is that the associated V of these three families of lattices can be constructed generically, and this fact makes using a specific family of lattices more desirable in practical implementation. As a remark, the spherical estimator $\hat{p}_*^{V_{D_d}}$ in our framework equals the orthonormalized-2 estimator in Deák (2000), albeit this connection was not revealed in the original paper.

In the following demonstration, we focus just on the spherical estimator $\hat{p}_*^V$ with V arising from a lattice, because $\hat{p}_*^V$ enjoys lower variances than $\hat{p}^V$ and $\hat{p}_{AT}^V$ when the inner radial integral can be explicitly computed.

To compare the efficiency of $\hat{p}_*^V$ with V arising from various lattices, we report variance ratio and penalized variance ratios. The variance ratio is defined as the variance of a crude Monte Carlo estimator divided by the variance of an estimator of interest. Therefore, an estimator with variance ratio greater than 1 is more efficient than the crude Monte Carlo estimator. To account for a specific penalized factor, such as the cardinality of an associated V, the number of independent random variables, or the computational time, we define the penalized variance of an estimator as the product of its variance and a penalized factor of interest. Similarly, we define the penalized variance ratio of an estimator as the penalized variance of the crude Monte Carlo estimator divided by the penalized variance of an estimator of interest.

In Table 2 we list cardinalities of V arising from four lattices, including $\mathbb{Z}_d$, A_d, D_d, and the proposed $L_d^{\max}$ in 16 and 24 dimensions. We summarize variance ratios and penalized variance ratios by cardinality to compromise the effect of cardinality. The quantity of interest is $p = \mathbb{P}(Z \in [-\infty, 1]^d)$. The proposed lattices $L_d^{\max}$ based on Theorem 1 are the Barnes–Wall lattice in 16 dimensions and the Leech lattice in 24 dimensions, respectively. The explicit construction of the shortest vectors of these lattices can be found in Conway and Sloane (1999).

It is clear that the cardinality of the proposed $V_{L_d^{\max}}$ is dramatically greater than the other three counterparts. The variance reduction by using $V_{L_d^{\max}}$ is substantial. As for the penalized variance ratio by cardinality, the spherical Monte Carlo estimator employing $V_{L_d^{\max}}$ remains the most competitive one at a factor of about three to four.

4. Simulation studies

In Section 4.1 we describe the simulation settings we use for comparison and in Section 4.2 we discuss the computational cost of each estimator. In Section 4.3 we study the minimum sample size required to achieve the desired precision level. In Section 4.4 we compare the efficiency of the proposed estimators in terms of variance ratios and penalized variance ratios. Moreover, in Section 4.5 we discuss the conditions required on the region $\mathcal{A}$ under which $\hat{p}_{AT}^V$ is more efficient than $\hat{p}^V$. Finally, in Section 4.6 we compare our estimators with the GHK multivariate normal simulator.

4.1. Simulation settings

Similar to the simulation settings used in Vijverberg (1997), we consider three types of covariance model: the identity covariance matrix, the one-factor model, and the AR(1) model. For the one-factor model with parameter ρ, the covariant matrix is set as $\Sigma = (\rho_{ij})$ with $\rho_{ij} = 1$ if $i = j$ and $\rho_{ij} = \rho$ if $i \neq j$. For the AR(1) model with parameter ρ, the covariant matrix is set as $\Sigma = (\rho_{ij})$ with $\rho_{ij} = \rho^{|i-j|}$ for all i and j. For the one-factor model and the AR(1) model, four different values of ρ, $-0.1, 0.1, 0.2$, and 0.3 are used. All these values ρ produce positive definite covariance matrices.

For the region $\mathcal{A}$, we consider rectangles, orthants, and ellipsoids. For each region shape we further propose sets having a feature other than central symmetry. Because orthants are unlikely to be centrally symmetric, there are a total of eight types of region. Among each type, three explicit sets are given in an order such that a set having the smallest area is listed first.

For rectangles, we include $[-a, a]^d$ for $a = 0.5, 1, 2$ as three centrally symmetric sets, $[0, a]^d$ for $a = 1, 2, 3$ as three centrally antisymmetric sets, and $[3, 1] \times [-1, 1]^{d-1}$, $[-1, 2]^d$ and $[-1.5, 2.5]^d$ as three mixed sets. For orthants, we include $[-\infty, a]^d$ for $a = -2, -1, 0$ as three centrally symmetric sets, and $[-\infty, 2] \times [-\infty, 0.5]^{d-1}$, $[-\infty, a]^d$ for $a = 1, 2$ as three mixed sets. For ellipsoids, we include $\{x \in \mathbb{R}^d : x_1^2 + \cdots + x_d^2 \leq a\}$ for $a = 1, 2, 3$ as three centrally symmetric sets, $\{x \in \mathbb{R}^d : (x_1 - 2)^2 + x_2^2 + \cdots + x_d^2 \leq a\}$ for $a = 1, 2, 3$ as three centrally antisymmetric sets, and $\{x \in \mathbb{R}^d : (x_1 - 1)^2 + x_2^2 + \cdots + x_d^2 \leq a\}$ for $a = 1.5, 2.5, 3.5$ as three mixed sets.

Note that the property of centrally antisymmetric is preserved under changing coordinates via a linear transformation. Therefore, when the region $\mathcal{A}$ is centrally antisymmetric, $\tilde{\mathcal{A}} = \Gamma^{-1}(\mathcal{A})$ remains centrally antisymmetric. Here we perform simulations in 2 to 8 dimensions.

4.2. Computational cost

Since the quantity of interest in this paper is the multivariate normal probability, and the integrand consists of an indicator function, we restrict the computational cost to be the smallest possible number of independent random numbers required to generate one realization for the estimator. It is clear that the computational cost of $\hat{p}$ and $\hat{p}_{\text{AT}}$ are both equal to d.

Recall that in Section 2.2 the computational cost to generate a $d \times d$ orthogonal matrix is $(d + 2)(d - 1)/2$. Because a realization of $\hat{p}^V$ requires us to generate one random orthogonal matrix and $|V|$ independent radii, the computational cost of $\hat{p}^V$ is $(d + 2)(d - 1)/2 + |V|$. Note that the cardinality of the proposed V is even since it is centrally symmetric.

Let V^+ consist of all vectors in V whose first nonzero coordinate is positive, and V is decomposed as the disjoint union of V^+ and $-V^+$. Without loss of generality, we assume that $V^+ = \{v_1, \ldots, v_{|V|/2}\}$.

Because generating a realization of $\hat{p}_{\text{AT}}^V$ requires a generation of one random orthogonal matrix and $|V|/2$ radii, the computational cost of $\hat{p}_{\text{AT}}^V$ is $(d + 2)(d - 1)/2 + |V|/2$. Finally, because an explicit formula of the innermost radial integral exists and no radius is required, the computational cost of $\hat{p}_*^V$ is simply $(d + 2)(d - 1)/2$.

4.3. Minimum sample size

We conduct an experiment in order to provide the minimum Monte Carlo sample size to achieve a predetermined precision level. For this purpose, it is common to select a Monte Carlo sample size M so that the length of the associated $(1 - \alpha)\%$ confidence interval is less than a predetermined level k. Then the smallest sample size M needs to satisfy $2z_{\alpha/2}\sqrt{\sigma^2/M} \leq k$, or,

TABLE 3: Averaged minimum sample sizes for calculating rectangle and orthant probabilities in d-dimensions such that the length of the associated 95% confidence interval is less than $p/10$.

	d						
	2	3	4	5	6	7	8
$\hat{p}$	873	1 879	4 031	8 752	19 410	44 226	103 886
$\hat{p}^V$	1 820	1 913	2 128	2 775	3 505	4 734	6 013
$\hat{p}^V_{AT}$	1 905	2 063	2 354	3 225	4 116	5 413	6 867
$\hat{p}^V_*$	127	254	380	658	861	1 157	1 319

equivalently, $M \geq 4z^2_{\alpha/2}\sigma^2/k^2$, where the quantile $z_{\alpha/2}$ satisfies $\mathbb{P}(Z \leq z_{\alpha/2}) = \alpha/2$ with a standard normal random variable Z, and σ^2 is the variance of the estimator.

Similar to Ross (2013), suppose that we are interested in calculating p and the length of the 95% confidence interval is required to be less than $p/10$. Furthermore, for ease of comparison, we simply consider the identity covariance matrix and rectangle and orthant probabilities, so that the true probability is theoretically known and the smallest sample size for the crude Monte Carlo can be derived exactly. For spherical Monte Carlo estimators, we estimate the minimum sample size iteratively with an initial sample size of 50; see Ross (2013).

In Table 3 we summarize the averaged minimum sample size required to achieve the desired precession over all rectangle and orthant probabilities as described in Section 4.1 for the crude Monte Carlo estimator and various spherical Monte Carlo estimators. Numerical results in Table 3 show that the crude Monte Carlo estimator requires quite a large sample size when the dimension grows. In contrast, three spherical Monte Carlo estimators require a drastically smaller sample size. It is even more transparent in the case of $\hat{p}^V_*$, to which the reduction of the sample size using $\hat{p}^V_*$ is substantial compared with the crude Monte Carlo estimator.

4.4. Efficiency of various estimators

To compare the efficiency of the proposed estimators fairly, in addition to Section 4.3, in Table 4 we summarize the averaged variance ratios and penalized variance ratios for various estimators used in calculating multivariate normal probabilities over various regions and covariance structures as outlined in Section 4.1. Overall, in terms of variance ratios, the efficiency of $\hat{p}^V_*$ is the highest, followed by the other two spherical Monte Carlo estimators $\hat{p}^V$ and $\hat{p}^V_{AT}$, and finally by the crude Monte Carlo estimator with antithetic variates $\hat{p}_{AT}$. Spherical Monte Carlo estimators $\hat{p}^V$, $\hat{p}^V_{AT}$, and $\hat{p}^V_*$ produce higher averaged variance ratios in higher dimensions, whereas $\hat{p}_{AT}$ produces variance ratios of about 2 in all dimensions. In addition, variance ratios of $\hat{p}^V_*$ are dramatically larger in all cases. This numerical evidence suggests that $\hat{p}^V_*$ is preferred, if the innermost radial integral $f(\boldsymbol{u})$ can be calculated explicitly.

The estimator $\hat{p}^V_*$ remains and enjoys the highest penalized variance ratios by computational cost. Penalized variance ratios by computational cost of $\hat{p}^V_{AT}$ and $\hat{p}_{AT}$ increase slightly in higher dimensions. Penalized variance ratios by computational cost of $\hat{p}_{AT}$ are about 2 in all dimensions. Although variance ratios of $\hat{p}^V_{AT}$ are slightly smaller than those of $\hat{p}^V$ in most cases, penalized variance ratios by computational cost of $\hat{p}^V_{AT}$ are larger. These numerical results suggest that $\hat{p}^V_{AT}$ is preferred accounting for the expense of drawing independent random variables.

TABLE 4: Averaged variances ratios (VR) and penalized variance ratios (PVR) of various estimators for calculating $p = \mathbb{P}\{X \in \mathcal{A}\}$ over various regions in d-dimensions. The Monte Carlo sample size is 10 000 for $\hat{p}$ and $\hat{p}_{AT}$, and is 1 000 for $\hat{p}^V$, $\hat{p}^V_{AT}$, and $\hat{p}^V_*$.

d	VR				PVR-cost				PVR-time			
	$\hat{p}_{AT}$	$\hat{p}^V$	$\hat{p}^V_{AT}$	$\hat{p}^V_*$	$\hat{p}_{AT}$	$\hat{p}^V$	$\hat{p}^V_{AT}$	$\hat{p}^V_*$	$\hat{p}_{AT}$	$\hat{p}^V$	$\hat{p}^V_{AT}$	$\hat{p}^V_*$
Rectangle region												
2	1.8	11.5	10.0	3487.4	1.8	2.9	4.0	3487.4	2.7	0.2	0.2	37.3
3	1.6	19.3	15.9	1965.4	1.6	3.4	4.3	1179.3	2.4	0.4	0.4	12.1
4	1.6	36.3	29.2	1413.3	1.6	4.4	5.6	628.2	3.7	1.3	1.2	11.5
5	1.6	55.5	44.2	1728.8	1.6	5.1	6.5	617.4	5.7	3.8	3.8	29.4
6	1.6	97.1	75.2	4255.7	1.6	6.3	8.1	1276.7	12.7	14.1	13.8	117.5
7	1.6	163.9	124.8	10885.9	1.6	7.5	9.7	2822.3	30.5	50.5	47.8	525.2
8	1.6	308.1	234.2	34907.0	1.6	9.0	12.1	7978.7	65.7	172.2	157.2	2359.5
Orthant region												
2	2.8	14.8	13.9	565.7	2.8	3.7	5.6	565.7	9.5	0.4	0.4	7.5
3	2.5	27.3	24.0	238.9	2.5	4.8	6.6	143.3	10.1	1.0	1.1	3.5
4	2.4	54.4	48.4	327.5	2.4	6.6	9.2	145.6	13.5	3.5	3.8	4.2
5	2.3	83.7	76.5	371.9	2.3	7.8	11.3	132.8	18.8	9.8	10.6	5.4
6	2.2	143.0	133.2	569.7	2.2	9.3	14.3	170.9	29.4	26.2	31.5	8.8
7	2.1	238.2	224.5	903.0	2.1	10.9	17.5	234.1	50.4	73.9	97.6	14.2
8	2.1	441.0	417.7	1883.1	2.1	12.8	21.6	430.4	110.9	172.6	221.2	26.1
Elliptical region												
2	2.4	10.4	7.5	3.0×10^8	2.4	2.6	3.0	3.0×10^8	0.6	15.6	13.1	1.5×10^8
3	2.2	20.0	13.8	2.7×10^8	2.2	3.5	3.8	1.6×10^8	0.5	26.9	21.3	6.0×10^7
4	2.0	38.1	26.7	4.2×10^7	2.0	4.6	5.1	1.9×10^7	0.6	53.4	42.0	6.3×10^6
5	1.8	59.9	43.2	2.6×10^6	1.8	5.5	6.4	9.1×10^5	0.8	102.6	79.9	6.9×10^5
6	1.8	102.7	76.9	6.3×10^7	1.8	6.7	8.2	1.9×10^7	1.3	272.6	200.4	2.3×10^7
7	1.7	168.4	129.5	2.0×10^8	1.7	7.7	10.1	5.1×10^7	2.4	734.6	517.7	1.4×10^8
8	1.7	309.7	241.7	1.1×10^{10}	1.7	9.0	12.5	2.6×10^9	5.4	2069.2	1415.6	1.3×10^{10}

In terms of penalized variance ratios by computational time, in a similar fashion, $\hat{p}_*^V$ outperforms the other three estimators, followed by spherical Monte Carlo estimators $\hat{p}^V$ and $\hat{p}_{AT}^V$, and then by the crude Monte Carlo estimator with antithetic variates $\hat{p}_{AT}^V$.

Our simulation studies show that $\hat{p}_*^V$ works particularly well in calculating ellipsoid probabilities. In fact, when the covariance matrix is the identity matrix and the ellipsoid is a circle centred at the origin, $f(\boldsymbol{u})$ is identical for all $\boldsymbol{u}$, and, hence, $\hat{p}_*^V$ is a zero-variance estimator. Because of this fact, we exclude the identity covariance matrix in averaging variance ratios and penalized variance ratios for calculating ellipsoid probabilities. When $\mathcal{A}$ slightly deviates from a circle centred at the origin and the covariance matrix is no longer the identity matrix, $f(\boldsymbol{u})$ produces quite close values at different $\boldsymbol{u}$, and this exemplifies a feature that $\hat{p}_*^V$ is a dramatically efficient estimator for calculating certain ellipsoid probabilities.

4.5. Antithetic variates

To investigate how the property of central antisymmetry affects the results, in Table 5 we summarize the averaged variance ratios and penalized variance ratios over all centrally antisymmetric sets and covariance matrices as outlined in Section 4.1. In these cases, values of variance ratios of $\hat{p}_{AT}^V$ and $\hat{p}^V$ are quite close. This provides numerical evidence that the property of central antisymmetry does matter when incorporating the idea of antithetic variates.

Note that due to an interesting feature in these centrally antisymmetric sets, averaged variance ratios and penalized variance ratios of $\hat{p}^V$ and $\hat{p}_{AT}^V$ in Table 5 are indeed the same. To explain why, let $\mathcal{A}$ currently be one of these four sets. Given a $\boldsymbol{v} \in V$, if $r\boldsymbol{v}$ belongs in $\mathcal{A}$, $-r^\top \boldsymbol{v}$ does not belong in $\mathcal{A}$ for all $r^\top$. As a result, a constituent of the spherical estimator $\hat{p}_{AT}^V$, $\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{T}_{\boldsymbol{v}}) + \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, -\boldsymbol{T}\boldsymbol{v})$ and a constituent of the spherical estimator $\hat{p}^V$, $\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}1}, \boldsymbol{T}_{\boldsymbol{v}}) + \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}2}, -\boldsymbol{T}\boldsymbol{v})$, both become 0 or 1. This leads to the conclusion that both estimators have the same variance. Therefore, the inequality in Theorem 2 is indeed an equality for these centrally antisymmetric regions.

To provide a reasonable region showing that $\hat{p}_{AT}^V$ enjoys a lower variance, consider a region $\mathcal{A} = [-1, -\frac{1}{2}] \times [-2, 2]^{d-1} \bigcup [0, \frac{1}{2}] \times [-2, 2]^{d-1}$. Clearly, this region is centrally antisymmetric, and, in addition, when using the antithetic variate, a constituent of estimator $\hat{p}_{AT}^V$, $\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{T}_{\boldsymbol{v}}) + \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, -\boldsymbol{T}\boldsymbol{v})$, would take a value of 0 or 1. On the contrary, when no antithetic variate is used, a constituent of the estimator $\hat{p}^V$, $\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}1}, \boldsymbol{T}_{\boldsymbol{v}}) + \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}2}, -\boldsymbol{T}\boldsymbol{v})$, would take a value of 0, 1, or 2. Because these two estimators are unbiased, the latter would have a larger variance. In Table 5 we demonstrate the desired results by showing that the averaged variance ratios of $\hat{p}_{AT}^V$ are larger than those of $\hat{p}^V$. Although the improvement is of a moderate scale, accounting for the computational cost, $\hat{p}_{AT}^V$ is preferred in terms of both penalized variance ratios by computational cost and by computational time.

4.6. Comparison with the GHK simulator

To demonstrate the applicability of the proposed method, we compare our method with the celebrated GHK method. The GHK simulator exploits the fact that a multivariate normal distribution function can be expressed as the product of sequentially conditioned univariate normal distribution functions, which can be easily and accurately evaluated. Moreover, based on a comprehensive simulation study, Hajivassiliou *et al.* (1996) concluded that the GHK simulator is the most reliable and accurate method to simulate multivariate normal orthant and rectangle probabilities. Here we focus just on $\hat{p}_*^V$ because the inner radial integral can be calculated explicitly in these cases. Furthermore, because the GHK simulator produces exact

TABLE 5: Averaged variance ratios (VR) and penalized variance ratios (PVR) of $\hat{p}^V$ and $\hat{p}_{AT}^V$ for calculating $p = \mathbb{P}\{X \in \mathcal{A}\}$ in d-dimensions. The Monte Carlo sample size is 1 000.

	VR		PVR-cost		PVR-time	
d	p^V	p_{AT}^V	p^V	p_{AT}^V	p^V	p_{AT}^V
Centrally antisymmetric region						
2	17.9	17.1	4.5	6.8	1.5	0.8
3	29.0	26.5	5.1	7.2	3.3	2.2
4	53.2	49.2	6.4	9.4	6.2	7.4
5	78.2	72.0	7.2	10.6	17.9	19.2
6	128.6	121.8	8.4	13.1	53.3	62.4
7	205.5	199.9	9.4	15.6	174.3	221.0
8	373.9	365.6	10.9	18.9	1025.8	1229.7
$\mathcal{A} = [-1, -\frac{1}{2}] \times [-2, 2]^{d-1} \bigcup [0, \frac{1}{2}] \times [-2, 2]^{d-1}$						
2	7.9	15.1	2.0	6.0	2.0	4.3
3	20.2	28.3	3.6	7.7	5.0	7.7
4	34.8	56.3	4.2	10.7	8.2	14.4
5	63.8	95.2	5.9	14.0	13.3	22.6
6	111.6	161.2	7.3	17.3	22.6	36.5
7	211.0	281.6	9.7	21.9	36.1	53.4
8	400.3	517.1	11.6	26.7	42.1	66.5

values in the case of independent multivariate normal distributions, we exclude the identity matrix but perform simulations on the other eight covariance structures.

To easily compare the performance of our proposed method with the GHK simulator, we report the relative efficiency, which is defined as the variance of the GHK simulator divided by the variance of our estimator $\hat{p}_*^V$. Similarly, we report the penalized relative efficiency corrected by computational cost and computational time. We remark that the GHK simulator has a computational cost of $(d-1)$ by its construction. In addition, we use the base 10 logarithmic scale on the averaged variance ratios and penalized variance ratios to report moderate values of measurements. In Figure 2 we depict the averaged relative efficiency and averaged penalized relative efficiency for calculating probabilities of five groups of regions in d-dimensions.

When the relative efficiency or the penalized relative efficiency in base 10 logarithmic scale is greater than 0, the proposed spherical Monte Carlo estimator is preferred because it produces an estimator with small variance or penalized variance. Otherwise, the GHK simulator is preferred. As mentioned in Section 4.1, we run simulations for five types of regions: (A) rectangle and centrally symmetric, (B) rectangle and centrally antisymmetric, (C) rectangle and mixed, (D) orthant and centrally antisymmetric, and (E) orthant and mixed. For each type of regions, we further consider 'three' explicit sets, ordered by the their areas from the smallest to the largest. For example, for the type of region (A), they are rectangle and centrally symmetric. And we further consider 'three' explicit sets $[-a, a]^d$ for $a = \frac{1}{2}, 1, 2$, which are ordered by their areas from the smallest to the largest in d-dimensions. In Figure 2, for each type of regions, numerical results are averaged across three types of covariance structures (identity, one-factor, and AR(1)). Numerical results for three sets within each type of regions are delineated in circles (set of the smallest area), diamonds (set of the second smallest area), and squares (set of the largest area).

Note that the performance of these two estimators varies significantly with various parameter
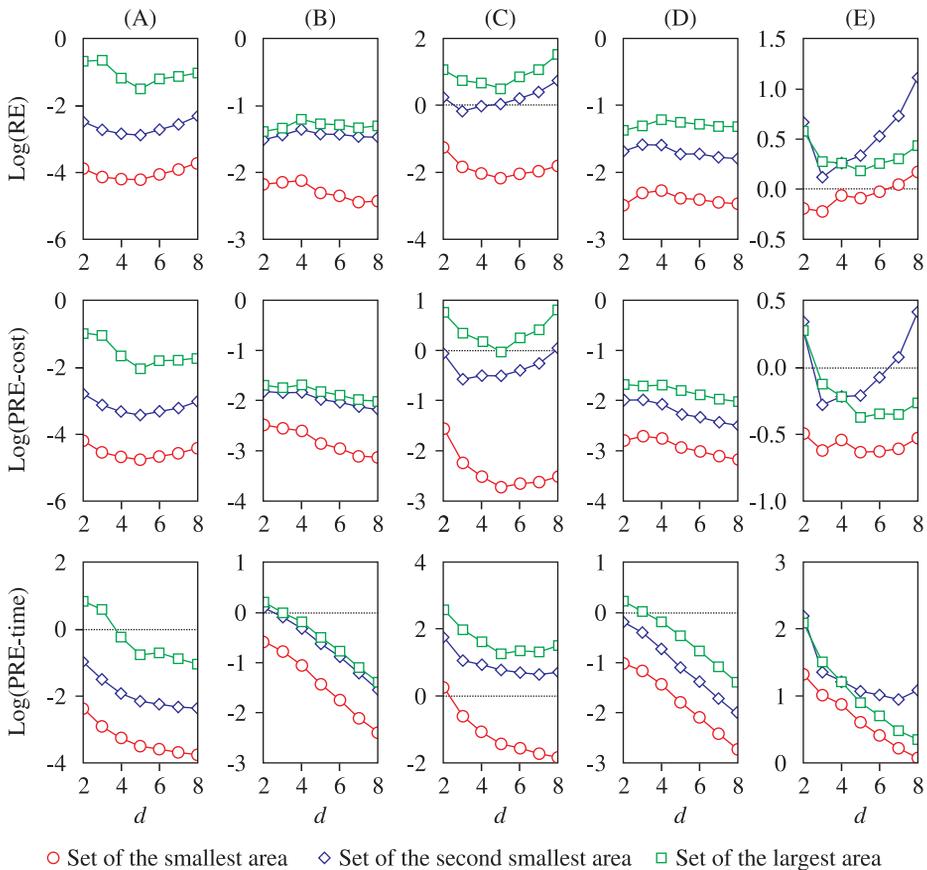
FIGURE 2: Averaged relative efficiency (RE) and penalized relative efficiency (PRE) of $\hat{p}_*^V$ over the GHK simulator in base 10 logarithmic scale against d-dimensions in calculating multivariate normal probabilities over various regions: (A) rectangle and centrally symmetric, (B) rectangle and centrally antisymmetric, (C) rectangle and mixed, (D) orthant and centrally antisymmetric, and (E) orthant and mixed. Numerical results for three sets within each type of regions are delineated in circles (set of the smallest area), diamonds (set of the second smallest area), and squares (set of the largest area). The Monte Carlo sample size is 1 000.

settings. Second, in general, we see that the data points shown as circles have lower values than those shown as squares against various dimensions. This feature suggests that the GHK simulator is favoured in calculating probabilities of relatively smaller regions. This is likely because the GHK simulator draws truncated standard normal random variables, and involves the idea of importance sampling.

A further investigation finds that, in calculating rectangle probabilities, the GHK simulator performs particularly well in calculating probabilities of centrally symmetric and centrally antisymmetric sets, whereas our method performs better in several mixed regions. This is particularly the case in higher dimensions. Similarly, in calculating orthant probabilities, the GHK simulator performs better in calculating probabilities of centrally antisymmetric regions; whereas $\hat{p}_*^V$ performs better in several mixed regions particularly in higher dimensions.

When incorporating penalized factors such as computational cost and computational time, similar conclusions hold that $\hat{p}_*^V$ is a competitive method to the GHK simulator. From this simulation study, we summarize that our method is a competitive method to the GHK simulator for calculating orthant and rectangle probabilities, because neither dominates the other in terms of relative efficiencies and penalized relative efficiencies. Under what conditions an estimator outperforms the other depends heavily on the property of the region, the dimension, and the size of the probability of interest. Nevertheless, our method is applicable to calculate multivariate normal probabilities of any type of regions.

5. Conclusion

Clearly, the proposed spherical Monte Carlo method is quite promising for calculating multivariate normal probabilities. To explain the superior performance of the proposed integration method, we examine the differences in integrating the radial and spherical parts. To reduce the variance for the spherical integral, we randomly rotate a set of points on the unit sphere, which is constructed from a sphere packing with the maximal kissing number. Moreover, we employ the idea of antithetic variates and show that it can have further variance reduction under certain conditions. Simulation studies confirm our theoretical results that the proposed method provides substantial variance reduction in some cases.

The success of the proposed spherical Monte Carlo estimator for calculating multivariate normal probabilities suggests that the method could be extended to accommodate more general cases, such as the calculation of probabilities based on elliptical distributions, Dirichlet distributions, and even high-dimensional integrals. For financial applications, it is expected that our method can be extended to option pricing and correlated default probabilities evaluation among others. A further challenge involves, for example, financial industry implementation of the CreditMetrics model and the computation of incremental risk charge under the Basel III regulatory framework, under which time dependence is considered. To accomplish these goals, we need to utilize the proposed spherical method with certain features of the distribution of interest, or, choose basis points for the spherical integration based on case 1. The above discussions pave a way to a fruitful area for further research.

Appendix A. Proof of Theorem 1

The proof of Theorem 1 depends on the calculation of the variance in (5). To this end, we first consider the covariance of any two random variables appearing in (5). Let $\mathcal{A}$ and $\mathcal{A}^\top$ be two disjoint regions of S^{d-1} satisfying that $\mathrm{diam}(\mathcal{A})$ and $\mathrm{diam}(\mathcal{A}^\top)$ are both less than $d_{\min}(V)$. Because the covariances of $g_{\mathcal{A}}^V(\boldsymbol{T})$ and $g_{\mathcal{A}^\top}^V(\boldsymbol{T})$ are difficult to calculate, we would like to find an upper bound of the covariance instead. Since $g_{\mathcal{A}}^V(\boldsymbol{T})$ and $g_{\mathcal{A}^\top}^V(\boldsymbol{T})$ are simple step functions, we have

$$
\begin{aligned}
\mathbb{E}[g_{\mathcal{A}}^V(\boldsymbol{T}) g_{\mathcal{A}^\top}^V(\boldsymbol{T})] &= \mathbb{E}\left[\sum_{\boldsymbol{v} \in V} \frac{1}{|V|} \mathbf{1}_{\mathcal{A}}(\boldsymbol{T}\boldsymbol{v}) \sum_{\boldsymbol{v}^\top \in V} \frac{1}{|V|} \mathbf{1}_{\mathcal{A}^\top}(\boldsymbol{T}\boldsymbol{v}^\top) \right] \\
&= \frac{1}{|V|^2} \sum_{\boldsymbol{v} \in V} \sum_{\boldsymbol{v}^\top \in V} \mathbb{E}[\mathbf{1}_{\mathcal{A}}(\boldsymbol{T}\boldsymbol{v}) \mathbf{1}_{\mathcal{A}^\top}(\boldsymbol{T}\boldsymbol{v}^\top)] \\
&= \frac{1}{|V|^2} \sum_{\boldsymbol{v}, \boldsymbol{v}^\top \in V} \mathbb{P}\{\boldsymbol{T}\boldsymbol{v} \in \mathcal{A}, \boldsymbol{T}\boldsymbol{v}^\top \in \mathcal{A}^\top\}.
\end{aligned}
$$

Since all points v in V are predetermined and Tv is simply a rotation on the unit sphere. Therefore,

$$\sum_{v, v^\top \in V} \mathbb{P}\{Tv \in \mathcal{A}, Tv^\top \in \mathcal{A}^\top\} \leq \sum_{v, v^\top \in V} \min(\mathbb{P}\{Tv \in \mathcal{A}\}, \mathbb{P}\{Tv^\top \in \mathcal{A}^\top\}).$$

Recall that when v is fixed, Tv is uniformly distributed on S^{d-1}, which implies that $\mathbb{P}\{Tv \in \mathcal{A}\} = \pi(\mathcal{A})$, $\mathbb{P}\{Tv^\top \in \mathcal{A}^\top\} = \pi(\mathcal{A}^\top)$, and

$$\mathbb{E}[g_{\mathcal{A}}^V(T) g_{\mathcal{A}^\top}^V(T)] \leq \frac{1}{|V|^2} \min(|V|\pi(\mathcal{A}), |V|\pi(\mathcal{A}^\top)) \leq \frac{1}{2|V|}(\pi(\mathcal{A}) + \pi(\mathcal{A}^\top)).$$

Hence, $\text{cov}(g_{\mathcal{A}}^V(T), g_{\mathcal{A}^\top}^V(T)) \leq (1/2|V|)(\pi(\mathcal{A}) + \pi(\mathcal{A}^\top)) - \pi(\mathcal{A})\pi(\mathcal{A}^\top)$. Then, we have

$$
\begin{aligned}
\text{var}(g_{\mathcal{A}}^V(T)) &= \text{var}\left(\sum_{j=1}^N g_{\mathcal{A}_j}^V\right) \\
&= \sum_{j=1}^N \text{var}(g_{\mathcal{A}_j}^V) + \sum_{i \neq j} \text{cov}(g_{\mathcal{A}_i}^V, g_{\mathcal{A}_j}^V) \\
&\leq \sum_{j=1}^N \frac{\pi(\mathcal{A}_j)}{|V|} - \pi^2(\mathcal{A}_i) + \sum_{i \neq j}\left(\frac{1}{2|V|}(\pi(\mathcal{A}_i) + \pi(\mathcal{A}_j)) - \pi(\mathcal{A}_i)\pi(\mathcal{A}_j)\right) \\
&= \frac{1}{|V|}\pi(\mathcal{A}) - \pi^2(\mathcal{A}) + \frac{1}{2|V|}\sum_{i \neq j}(\pi(\mathcal{A}_i) + \pi(\mathcal{A}_j)) \\
&= \frac{1}{|V|}\pi(\mathcal{A}) - \pi^2(\mathcal{A}) + \frac{1}{2|V|}\sum_{j=1}^N 2(N-1)\pi(\mathcal{A}_i) \\
&= \frac{1}{|V|}\pi(\mathcal{A}) - \pi^2(\mathcal{A}) + \frac{N-1}{|V|}\pi(\mathcal{A}) \\
&= \pi(\mathcal{A}) - \pi^2(\mathcal{A}) + \left(\frac{N}{|V|} - 1\right)\pi(\mathcal{A}).
\end{aligned}
$$

Appendix B. Proof of Theorem 2

In order to prove Theorem 2, we first note that

$$
\begin{aligned}
&|V|(\text{var}(\hat{p}_{\text{AT}}^V) - \text{var}(\hat{p}^V)) \\
&\quad = \sum_{v \in V^+} \text{cov}(\mathbf{1}_{\mathcal{A}}(r_v, Tv), \mathbf{1}_{\mathcal{A}}(r_v, -Tv)) - \text{cov}(\mathbf{1}_{\mathcal{A}}(r_v, Tv), \mathbf{1}_{\mathcal{A}}(r_{-v}, -Tv)). \quad (8)
\end{aligned}
$$

If $\mathcal{A}$ is centrally antisymmetric then for any $\chi(d)$-distributed independent random variables r_1, r_2, and $\chi(d)$-distributed random variable r, we have

$$
\begin{aligned}
&\text{cov}(\mathbf{1}_{\mathcal{A}}(r_1, u), \mathbf{1}_{\mathcal{A}}(r_2, -u)) - \text{cov}(\mathbf{1}_{\mathcal{A}}(r, u), \mathbf{1}_{\mathcal{A}}(r, u)) \\
&\quad = \mathbb{E}[\mathbf{1}_{\mathcal{A}}(r_1, u)\mathbf{1}_{\mathcal{A}}(r_2, -u)] - \mathbb{E}[\mathbf{1}_{\mathcal{A}}(r, u)\mathbf{1}_{\mathcal{A}}(r, -u)] \\
&\quad = \mathbb{E}[\mathbf{1}_{\mathcal{A}}(r_1, u)\mathbf{1}_{\mathcal{A}}(r_2, -u)] \\
&\quad \geq 0,
\end{aligned}
$$

which implies that

$$\mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u})) \leq \mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r_1, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r_2, -\boldsymbol{u})). \tag{9}$$

Followed by (3), we can write (8) as

$$\sum_{\boldsymbol{v} \in V^+} \mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, -\boldsymbol{u})) - \mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r_{\boldsymbol{v}}, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r_{-\boldsymbol{v}}, -\boldsymbol{u}))$$

$$= \frac{|V|}{2}(\mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r, -\boldsymbol{u})) - \mathrm{cov}(\mathbf{1}_{\mathcal{A}}(r_1, \boldsymbol{u}), \mathbf{1}_{\mathcal{A}}(r_2, \boldsymbol{u}))),$$

which is less than or equal to 0 by (9).

Acknowledgements

References

ANDERSON, T. W., OLKIN, I. AND UNDERHILL, L. G. (1987). Generation of random orthogonal matrices. *SIAM J. Sci. Statist. Comput.* **8,** 625–629.

ASMUSSEN, S. AND GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

CONWAY, J. H. AND SLOANE, N. J. A. (1999). *Sphere Packings, Lattices and Groups*, 3rd edn. Springer, New York.

CRAIG, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *J. R. Statist. Soc. B* **70,** 227–243.

DAVIS, P. J. AND RABINOWITZ, P. (1984). *Methods of Numerical Integration*, 2nd edn. Academic Press, Orlando, FL.

DEÁK, I. (1980). Three digit accurate multiple normal probabilities. *Numer. Math.* **35,** 369–380.

DEÁK, I. (2000). Subroutines for computing normal probabilities of sets—computer experiences. *Ann. Operat. Res.* **100,** 103–122.

DIACONIS, P. AND SHAHSHAHANI, M. (1987). The subgroup algorithm for generating uniform random variables. *Prob. Eng. Inf. Sci.* **1,** 15–32.

FANG, K.-T. AND WANG, Y. (1994). *Number-Theoretic Methods in Statistics* (Monogr. Statist. Appl. Prob. **51**). Chapman & Hall, London.

GENZ, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graphical Statist.* **1,** 141–150.

GENZ, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Comput. Sci. Statist.* **25,** 400–405.

GENZ, A. AND BRETZ, F. (2002). Methods for the computation of multivariate *t* probabilities. *J. Comput. Graphical Statist.* **11,** 950–971.

GENZ, A. AND BRETZ, F. (2009). *Computation of Multivariate Normal and t Probabilities* (Lecture Notes Statist. **195**). Springer, Dordrecht.

GLASSERMAN, P., HEIDELBERGER, P. AND SHAHABUDDIN, P. (2000). Variance reduction techniques for estimating value-at-risk. *Manag. Sci.* **46,** 1349–1364.

GUPTON, G. M., FINGER, C. C. AND BHATIA, M. (1997). *CreditMetrics – Technical Document*. Morgan, New York.

HAJIVASSILIOU, V., MCFADDEN, D. AND RUUD, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. *J. Econometrics* **72,** 85–134.

HEIBERGER, R. M. (1978). Algorithm AS 127: generation of random orthogonal matrices. *J. R. Statist. Soc. C* **27,** 199–206.

HSU, J. C. (1996). *Multiple Comparisons*. Chapman & Hall, London.

MIWA, T., HAYTER, A. J. AND KURIKI, S. (2003). The evaluation of general non-centred orthant probabilities. *J. R. Statist. Soc. B* **65,** 223–234.

MONAHAN, J. AND GENZ, A. (1997). Spherical-radial integration rules for Bayesian computation. *J. Amer. Statist. Assoc.* **92,** 664–674.

Nebe, G. and Sloane, N. J. A. A catalogue of lattices. Available at http://www.math.rwth-aachen.de/∼Gabriele.Nebe/ LATTICES/.

Ross, S. M. (2013). *Simulation*, 5th edn. Elsevier, Amsterdam.

Sándor, Z. and András, P. (2004). Alternative sampling methods for estimating multivariate normal probabilities. *J. Econometrics* **120,** 207–234.

Sloane, N. J. A. Spherical Codes. Available at http://neilsloane.com/packings/index.html#I.

Somerville, P. N. (2001). Numerical computation of multivariate normal and multivariate *t* probabilities over ellipsoidal regions. *J. Statist. Software* **6,** 10pp.

Stewart, G. W. (1980). The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numerical Anal.* **17,** 403–409.

Vijverberg, W. P. M. (1997). Monte Carlo evaluation of multivariate normal probabilities. *J. Econometrics* **76,** 281–307.