Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

GENERATING "DEPENDENT" QUASI-RANDOM NUMBERS

Shane G. Henderson

Department of Industrial and
Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117, U.S.A.

Belinda A. Chiera
Roger M. Cooke

Department of Mathematics
Delft University of Technology
THE NETHERLANDS

ABSTRACT

Under certain conditions on the integrand, quasi-Monte Carlo methods for estimating integrals (expectations) converge faster asymptotically than Monte Carlo methods. Motivated by this result we consider the generation of quasi-random vectors with given marginals and given correlation matrix. We extend the "Normal To Anything" (NORTA) method, introduced by Cario and Nelson, to this context, and term the extension the "Quasi-Random to Anything" (QUARTA) method.

1 INTRODUCTION

We present a new approach for computing integrals (expectations) of the form $Eg(X)$, for some function g, and a class of random vectors X. This problem arises in a host of applications. For example, in stochastic linear programming, X represents certain random input data to a linear program, and the function g gives the optimal objective value of the linear program (Infanger 1994). In stochastic activity networks, X represents the random task durations on the arcs of the network, and g reflects the length of the longest path between two specified nodes.

In both of these applications, X takes the form of a d-dimensional vector of real-valued random variables. If i.i.d. replicates of X can be generated, then the Monte Carlo method may be used to estimate $Eg(X)$.

If the d components of X are modeled as independent random variables, then univariate generation techniques may be applied to each of the components independently to generate X. However, the assumption of independent components may be an unreasonable one for many applications (Infanger 1994, Cario and Nelson 1997).

There are many models for specifying multivariate random vectors with correlated components and marginal distributions from a single parametric family; see Devroye (1986) and Johnson (1987) for surveys. There are fewer methods for dealing with the case where the marginal distributions do not come from a common family.

Specifying the distributions of such random vectors can be an onerous task, let alone developing variate generation algorithms. It is natural then, to simply specify the marginal distributions of the components of X, together with their covariance matrix. This approach does not necessarily uniquely specify the distribution of X. However, it is far easier to specify this data than to specify a full multivariate distribution. Furthermore, there are methods for generating random vectors with specified marginals and covariance matrix.

The extremal distributions method of Hill and Reilly (1994) can be applied in this case, but practically speaking, it appears to be limited in applicability to low-dimensional ($d \leq 4$ say) random vectors.

Cario and Nelson (1997) describe the "Normal to Anything" (NORTA) method, which easily scales to high-dimensional random vectors. The basic idea is to begin with a random vector Z with a multivariate normal distribution, and transform Z to yield a random vector X with the desired marginals and correlation structure. Cario and Nelson (1997) gave structural results that establish the feasibility of a numerical approach to determining the correlation structure of Z which induces the required correlation structure of X. They traced the origins of the NORTA method back to Mardia (1970) who looked at transformations of bivariate normal random variables, and Li and Hammond (1975), who looked at random vectors where all marginals have densities (with respect to Lebesgue measure). Iman et al (1981) and Iman and Conover (1982) implemented a joint normal transform procedure, where the variables Z and X have the same rank correlation structure. Their approach is essentially the NORTA method with a different method for choosing the correlation matrix of Z. Clemen and Reilly (1999) use the NORTA method, attempting to ensure a given rank correlation in the output. They employ an explicit formula for the rank correlations of multivari-

ate normal random variables to determine the appropriate correlation structure of Z.

It might be conceived that the NORTA method could be used to generate random vectors with arbitrary marginals and an arbitrary *feasible* covariance matrix (feasible in the sense that a random vector with the specified marginals and covariance matrix exists). Unfortunately, this is not the case, as Ghosh and Henderson (2000) show that there are sets of marginals with feasible covariance matrix that cannot be generated with the NORTA procedure. This fact was noted, although not rigorously established, in both Li and Hammond (1975) and Clemen and Reilly (1999). However, the method can be adjusted to generate a random vector X with the required marginals, and a covariance matrix that is "close" to the desired covariance matrix. Clemen and Reilly (1999) give one such adjustment, and Ghosh and Henderson (2000) give another. We outline how Ghosh and Henderson (2000) do this in Section 2.

In this paper, we use the NORTA procedure as a tool to assist in computing $Eg(X)$. A Monte Carlo approach to computing $Eg(X)$ generates i.i.d. replicates $X(1), \ldots, X(n)$ of X, and computes

$$\alpha_{\mathrm{mc}}(n) = \frac{1}{n}\sum_{i=1}^{n} g(X(i)).$$

If $E[g(X)]^2 < \infty$, then α_n satisfies the central limit theorem

$$\sqrt{n}(\alpha_n - Eg(X)) \Rightarrow \sigma N(0,1)$$

as $n \to \infty$, where $\Rightarrow$ denotes weak convergence, $N(0,1)$ denotes a standard normal random variable, and $\sigma^2 = \mathrm{var}(g(X))$. Hence α_n converges at rate $n^{-1/2}$ to $Eg(X)$, independent of the dimension d of X.

Alternatively, numerical integration techniques may be employed to estimate $Eg(X)$. We may write

$$Eg(X) = \int_S g(x)\pi(dx),$$

where π is the distribution of X, and $S \subseteq \mathbb{R}^d$ is the support of π. This integral can be transformed into one on the unit hypercube in d dimensions with respect to Lebesgue measure. Numerical integration techniques can then be applied to estimate $Eg(X)$. In particular, the (deterministic) points $u(1), u(2), \ldots, u(n)$ within the unit hypercube might be chosen, and an approximation of $Eg(X)$ computed via

$$\frac{1}{n}\sum_{i=1}^{n} h(u(i)), \tag{1}$$

where the function h depends on the transformation of the integral over S to one over the unit hypercube.

It is known (Niederreiter 1992, p. 32) that under certain conditions on g, sequences $u = u(1), u(2), \ldots$ exist for which the error in (1) decreases at most at rate $n^{-1}(\log n)^d$. This rate is (asymptotically) faster than the rate $n^{-1/2}$ exhibited by the Monte Carlo method.

Such sequences are termed quasi-random number (QRN) sequences, and because they are designed to "uniformly" fill the unit hypercube, we will say that they are quasi-random numbers with a uniform distribution on the unit hypercube. However, it should be noted that QRN sequences are specifically designed to be deterministic, and not share certain properties with an i.i.d. sequence (unlike pseudo-random numbers). Consequently some care must be exercised when speaking about the distribution of such a sequence of points.

It is then reasonable to ask whether it is possible to generate QRN sequences with a nonuniform distribution. Gentle (1998, Chapter 2) surveyed a number of methods used to transform uniformly distributed univariate (quasi-)random number sequences to nonuniform distributions. Further, the issue of directly sampling (quasi-)random numbers from specific distributions including univariate and multivariate distributions, as well as over geometric objects, has also been explored (Gentle, 1998, Chapter 3). Chiera and Cooke (2000) have recently extended this problem by looking at the generation of QRN sequences for Markov trees with diagonal band copulae. A joint distribution is determined by one-dimensional marginals and rank correlations on a tree whose nodes are the one dimensional marginals, together with a maximum entropy condition. This latter condition is always consistent and ensures that realizations are Markov; that is, they possess a conditional independence structure given by the tree, considered as an (undirected) belief net (Meeuwissen, 1993, Meeuwissen and Bedford 1997, Cooke 1997).

But why are nonuniform QRN sequences of interest? The answer is that it is possible to estimate $Eg(X)$ using an estimator of the form

$$\alpha_{\mathrm{qmc}}(n) = \frac{1}{n}\sum_{i=1}^{n} g(x(i)),$$

where the quasi-random sequence $\{x(i)\}$ is chosen "to have distribution π" (we will formalize the notion of the distribution of a deterministic sequence of points in Section 3). There is certainly potential value in such an approach.

To see why, suppose that $g(x) = c$, a constant. Then clearly, $Eg(X) = c$. Using the estimator α_{qmc} yields the exact solution with a single point $x(1)$. If the change of variables technique mentioned earlier were to be used, then in general, the integrand over $[0,1)^d$ will not be constant, and a single integrand evaluation will not yield the exact value of the expectation. While this example is highly

artificial, it nevertheless motivates the use of quasi-random numbers with distribution π.

Given that we have an acceptable definition of quasi-random numbers with a given distribution π say, we then need a method for generating them. To do so directly appears to be a rather formidable task, as the support of the distribution π and so forth must be taken into account. However, there are a host of methods for generating quasi-random numbers in the unit hypercube (see Gentle, 1998, for a survey of these methods). So instead we transform quasi-random numbers with a uniform distribution in the unit hypercube to quasi-random numbers with the desired distribution π. Basically, some form of inversion method needs to be applied, and the NORTA method is one such method. Therefore, we propose to take quasi-random numbers, and transform them into the desired distribution using the NORTA method. We will refer to this process as the "quasi-random to anything", or QUARTA method. (It should be noted that NORTA can only generate a restricted class of distributions, and is not a completely general method.)

To demonstrate the potential of QUARTA, we will provide two elementary examples that show that improved accuracy over pure Monte Carlo is possible with (approximately) the same amount of computation.

The remainder of this paper is structured as follows. In Section 2, we review the NORTA method. We discuss some of the properties of the method, and mention some pertinent results from Ghosh and Henderson (2000). In Section 3 we define what we mean by quasi-random numbers with a given distribution. Then, in Section 4, we outline the QUARTA method, which is basically a variant of the NORTA method. Finally, in Section 5, we provide two examples of the application of the QUARTA method for estimating expectations (integrals).

2 THE NORTA METHOD

Suppose that we wish to generate i.i.d. replicates of an $\mathbb{R}^d$-valued random vector $X = (X_1, \ldots, X_d)$, with marginal distribution functions

$$F_i(\cdot) = P(X_i \leq \cdot), i = 1, \ldots, d,$$

and correlation matrix

$$\Sigma_X = (\Sigma_X(i, j) : 1 \leq i, j \leq d),$$

with $\Sigma_X(i, j) = \text{cor}(X_i, X_j)$. We require that $E(X_i^2) < \infty$ for $i = 1, \ldots, d$, so that the correlation matrix Σ_X is defined, but otherwise impose no conditions on the marginal distribution functions $F_i, i = 1, \ldots, d$. We assume that Σ_X is *feasible* for the given marginals, in that a random vector exists with the specified marginals and correlation structure.

Cario and Nelson (1997) described the NORTA procedure for solving this problem, which works as follows.

1. Generate an $\mathbb{R}^d$ valued normal random vector $Z = (Z_1, \ldots, Z_d)$ with mean vector 0 and covariance matrix $\Sigma_Z = (\Sigma_Z(i, j) : 1 \leq i, j \leq d)$, where $\Sigma_Z(i, i) = 1$ for $i = 1, \ldots, d$. Then each Z_i is a standard normal random variate (mean 0 and variance 1). We will further specify Σ_Z shortly.
2. Compute the vector $X = (X_1, \ldots, X_d)$ via

$$X_i = F_i^{-1}(\Phi(Z_i)), \tag{2}$$

for $i = 1, \ldots, d$, where Φ is the distribution function of a standard normal random variable, and

$$F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}. \tag{3}$$

At the conclusion of this procedure, X will have the prescribed marginal distributions, because $\Phi(Z_i)$ is uniformly distributed on $(0, 1)$, and so $F_i^{-1}(\Phi(Z_i))$ will have the required marginal distribution. Note that this algorithm requires the calculation of $\Phi(z)$ for many z. While this is not possible in closed form, fast numerical algorithms are available to perform the calculation to high accuracy (Abramowitz and Stegun, 1964, Chapter 26).

It is easy to generate multivariate normal random vectors. See p. 480 of Law and Kelton (2000), for example.

In a preprocessing step, the correlation matrix Σ_Z is chosen in an attempt to ensure that X will have the prescribed correlation matrix Σ_X. Determining the matrix Σ_Z is the principal difficulty in applying the NORTA method, and we now explain how this may be done.

As in Cario and Nelson (1997), define the function $c_{ij}(z) = \text{cor}(X_i, X_j)$, where X_i and X_j are defined via (2) and $\text{cor}(Z_i, Z_j) = z$. We would like to choose $\Sigma_Z(i, j)$ so that $c_{ij}(\Sigma_Z(i, j))$ matches the desired correlation $\Sigma_X(i, j)$. Cario and Nelson (1997) established the following structural result for the function c_{ij}.

Theorem 1 *The quantity $c_{ij}[\Sigma_Z(i, j)]$ is nondecreasing in $\Sigma_Z(i, j)$, and the minimum (resp. maximum) possible correlation between X_i and X_j (for the given marginal distribution functions F_i and F_j) is achieved by taking $\Sigma_Z(i, j) = -1$ (resp. $+1$). Furthermore, if there exists some $\epsilon > 0$ such that*

$$E|X_i X_j|^{1+\epsilon} < \infty$$

for all values $-1 \leq \Sigma_Z(i, j) \leq 1$, then c_{ij} is a continuous function of $\Sigma_Z(i, j) \in [-1, 1]$.

Theorem 1 allows one to perform an efficient numerical search for values $\Lambda_Z(i, j)$ that yield

$$c_{ij}(\Lambda_Z(i, j)) = \Sigma_X(i, j) \text{ for } i < j. \tag{4}$$

We take $\Lambda_Z(i,i) = 1$ for $i = 1, \ldots, d$. The values $\Lambda_Z(i,j)$ for $i > j$ can then be chosen to ensure that the matrix Λ_Z is symmetric. Unless otherwise stated, we henceforth assume that a solution to (4) exists.

Let the matrix Λ_Z satisfy (4). If Λ_Z is not positive semidefinite, then it is not a valid correlation matrix, and there is no normal random vector Z with correlation matrix Λ_Z. In particular, Λ_Z cannot be used within a NORTA procedure to generate X.

However, it may still be possible to generate X via a NORTA transformation. If (4) does not have a unique solution, then a second matrix Λ'_Z may exist that satisfies (4) and that is positive semidefinite.

Theorem 2 below basically shows that a solution to (4) is unique when all the marginals have densities (with respect to Lebesgue measure) that are positive everywhere. We conjecture that the solution is, in fact, unique for arbitrary marginals. For a proof of this result, see the appendix.

Theorem 2 *Suppose that for $i = 1, \ldots, d$, F_i has a density f_i that is positive everywhere, i.e., $f_i(x) > 0$ for all $x \in \mathbf{R}$. If Λ_Z solves (4), then Λ_Z is unique.*

We immediately obtain the following corollary.

Corollary 3 *Suppose that it is possible to generate a random vector X with prescribed marginals and correlation matrix using a NORTA transformation. If the conditions of Theorem 2 hold, then the matrix Λ_Z found using a perfectly accurate numerical search procedure will be symmetric and positive semidefinite.*

Proof: The assumption that a NORTA transformation exists that yields the required correlations ensures that there is a positive semidefinite correlation matrix Λ that solves (4). Under the conditions stated, the solution to (4) is unique, and so $\Lambda_Z = \Lambda$ and is positive semidefinite.

The significance of this corollary is basically that if NORTA *can* work, then it *will* work for the distributions characterized in Theorem 2.

So then, *can* NORTA work, i.e. does a NORTA transformation exist for any set of marginals and feasible correlation matrix? Li and Hammond (1975) suggested the following counterexample to this important question.

Let X_1, X_2 and X_3 be 3 uniformly distributed random variables on $(0, 1)$ with correlation matrix

$$\Sigma_X = \begin{pmatrix} 1 & -0.4 & 0.2 \\ -0.4 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix}.$$

Li and Hammond quote the formula

$$\Lambda_Z(i,j) = 2\sin\left(\frac{\pi}{6}\Sigma_X(i,j)\right) \tag{5}$$

for the (unique) matrix Λ_Z that solves (4) (see Kruskal 1958 for a proof). The (unique) matrix Λ_Z resulting from these computations is not positive semidefinite.

Of course, this is only a counterexample if the random vector $X = (X_1, X_2, X_3)$ exists. Li and Hammond (1975) did not show this, but Ghosh and Henderson (2000) have since shown, using linear programming techniques, that indeed such a random vector can be constructed. Therefore, there are sets of marginal distributions with a feasible covariance matrix that NORTA cannot reproduce.

Suppose we take the position that we wish to use NORTA to generate a random vector with the prescribed marginals, and a covariance matrix that is, at least approximately, the required covariance matrix. Ghosh and Henderson (2000) describe a semidefinite programming approach that can assist in this regard. The method may be summarized as follows.

1. Use a numerical search procedure as described in Cario and Nelson (1997) to determine a symmetric matrix Λ_Z such that (4) is satisfied.
2. If Λ_Z is positive semidefinite, then one can proceed directly with the NORTA procedure.
3. If not, then we wish to find another matrix Σ_Z that is "close" in some sense to Λ_Z. So minimize $d(\Sigma_Z, \Lambda_Z)$ subject to the constraint that Σ_Z is positive semidefinite, where d is some measure of distance.
4. Use the matrix Σ_Z as the correlation matrix of Z within the NORTA procedure.

With a suitable choice of distance function d, the optimization in Step 3 above can be formulated as a semidefinite programming problem; see Ghosh and Henderson (2000) for one choice of d, or Alfakih and Wolkowicz (2000) for another. Efficient algorithms are available for solving such problems; see Wolkowicz, Saigal, and Vandenberghe (2000). The random vectors generated with the NORTA procedure using Σ_Z will have the correct marginal distributions, but will most likely have a different covariance matrix from that desired. However, the continuity established in Theorem 1 suggests that the covariance matrix will differ only slightly from that desired if Σ_Z is "close" to Λ_Z. The numerical examples given in Ghosh and Henderson (2000) suggest that this is usually the case.

3 QUASI-RANDOM VECTORS

We need to be somewhat careful in defining what is meant by "dependent" QRN sequences, as unlike pseudo-random number sequences, QRN sequences are specifically designed *not* to mimic the properties of i.i.d. sequences. Thus it does not, apriori, make sense to refer to the "distribution" of a QRN sequence.

QRN sequences are explicitly designed not to exhibit the clustering and gaps that are representative of an i.i.d. sequence. It is exactly this property that leads to faster convergence than the Monte Carlo method in estimating

integrals (expectations). To be able to measure this clustering/gap effect, one often speaks of the *discrepancy* of a particular point set. We begin with a discussion of the classical notion of discrepancy on the unit hypercube, as adapted from Niederreiter (1992).

Let $u = \{u(k)\}$ be a sequence of vectors defined within the unit hypercube $[0, 1)^d$ in d dimensions. For a given set $B \subseteq \mathbb{R}^d$, let

$$A_n(B; u) \triangleq \sum_{k=1}^{n} I(u(k) \in B)$$

be the number of vectors $u(k)$ from $u(1), \ldots, u(n)$ contained in the set B. Let $\mathcal{B}$ be a nonempty family of Lebesgue measurable subsets of $\mathbb{R}^d$. A general definition of the discrepancy D_n of the first n terms of the sequence u is

$$D_n(\mathcal{B}; u) \triangleq \sup_{B \in \mathcal{B}} \left| \frac{A_n(B; u)}{n} - \lambda_d(B) \right|, \tag{6}$$

where $\lambda_d(\cdot)$ is Lebesgue measure on $\mathbb{R}^d$.

If the class of sets $\mathcal{B}$ is taken to be all sets of the form

$$B = \prod_{i=1}^{d} [0, a_i)$$

for $a_i \in [0, 1)$, then the above definition yields the *star* discrepancy $D_n^*(u)$ of the sequence u (p. 14, Niederreiter 1992). The significance of this definition lies in its use in establishing a bound on the error in an estimate of an integral using the sequence of points u. In particular, the following well-known result is known as the Koksma-Hlawka inequality.

Theorem 4 (Koksma-Hlwaka Inequality): *Let g be a real-valued function defined on $[0, 1]^d$ and suppose that g has bounded variation $V(g)$ in the sense of Hardy and Krause (see p. 19 of Niederreiter 1992 for a definition). Then for any sequence of vectors $u = \{u(k)\}$ in $[0, 1)^d$, we have*

$$|\alpha(n) - \alpha| \leq V(g) D_n^*(u),$$

where

$$\alpha \triangleq \int_{[0,1]^d} g(x)\, dx < \infty,$$

and for a given sequence of points $u = \{u(k)\}$ with $u(k) \in [0, 1)^d$,

$$\alpha(n) = \frac{1}{n} \sum_{k=1}^{n} g(u(k)).$$

It is known (p. 32, Niederreiter 1992) that sequences u exist with the property that

$$D_n^*(u) = O(n^{-1} (\log n)^d),$$

so that the error in $\alpha(n)$ is at most of order $n^{-1}(\log n)^d$. This is a faster asymptotic rate than that obtained by the Monte Carlo method ($n^{-1/2}$), which motivates the use of quasi-random numbers for numerical integration. It should be noted however, that these asymptotic results may not be representative of the sample sizes n used in practice. In other words, for moderate n, it may be that the Monte Carlo method yields lower error than can be obtained through the use of a quasi-random sequence; see Kocis and Whiten (1997). This effect is especially pronounced in higher dimensions d.

If a sequence u of vectors has a star discrepancy that converges to 0 as $n \to \infty$, then we can view the sequence of vectors as an analogue of a sample from the uniform distribution on $[0, 1)^d$. It is very natural to generalize the notion of discrepancy to a non-uniform distribution on more general domains. Indeed, more general distributions on the unit hypercube have already been considered in the literature (Niederreiter, Tichy and Turnwald 1990).

Let π be a probability distribution on some set $S \subseteq \mathbb{R}^d$. Let $\mathcal{B}$ denote a class of π-measurable sets in S. For a given sequence of points $u = \{u(k)\}$ with $u(k) \in \mathbb{R}^d$, define the π-discrepancy of u to be

$$D_n^{\pi}(\mathcal{B}; u) = \sup_{B \in \mathcal{B}} \left| \frac{A_n(B; u)}{n} - \pi(B) \right|. \tag{7}$$

Note the similarity of (7) to (6). In particular, the goal of this definition is to characterize the property that the sequence u is an analogue of the probability distribution π on S. Niederreiter, Tichy and Turnwald (1990) studied a more general version of this definition where weights could be assigned to each of the points in the sequence u, but restricted the domain S to be the unit hypercube.

We will say that u is a quasi-random sequence of numbers with distribution π if

$$D_n^{\pi}(\mathcal{B}; u) \to 0$$

as $n \to \infty$, where $\mathcal{B}$ is the class of all sets of the form

$$\prod_{i=1}^{n} (a_i, b_i),$$

where $-\infty \leq a_i < b_i \leq \infty$, $i = 1, \ldots, d$.

Our goal is to use such sequences to estimate expectations of the form $Eg(X)$, where X is distributed according to π.

So how can a sequence for which the π-discrepancy rapidly converges to 0 be computed? Our goal in the next section is to show how, for the class of distributions that can be obtained through a NORTA transformation, to use a quasi-random sequence of points in $[0,1)^d$ to obtain a quasi-random sequence of points with the desired distribution.

4 THE QUARTA METHOD

We would like to be able to generate quasi-random vectors with given marginals and given feasible correlation matrix (where the exact meaning of this statement is given in the previous section). We do so by extending the NORTA method, and term the method QUARTA, which is intended to be mnemonic for "quasi-random to anything".

Suppose that $(u(n) : n \geq 1)$ is a quasi-random sequence of d-dimensional vectors in the unit hypercube $[0,1)^d$ with independent components (recall from the previous section that there is an appropriate notion of "independence" in this setting where the vectors are actually deterministic). We wish to transform these vectors into a quasi-random sequence $(x(n) : n \geq 1)$ of d-dimensional random vectors with marginals F_i and feasible correlation matrix Σ_X. We will transform $u(i)$ into $x(i)$ for $i \geq 1$. Note that in the following procedure the index i on the $u(i)$'s and $x(i)$'s has been dropped for ease of readability.

1. Identify a correlation matrix Σ_Z that yields (at least approximately) the appropriate correlation matrix Σ_X exactly as in the Monte Carlo version of NORTA. Compute R, a Cholesky factor of Σ_Z, so that $\Sigma_Z = R^T R$.
2. Transform u into y, where y has normal marginals, via $y_j = \Phi^{-1}(u_j)$.
3. Set $z = Ry$, so that z is standard multivariate normal with correlation matrix Σ_Z.
4. Compute x via a NORTA transformation of z, i.e., set $x_j = F_j^{-1}(\Phi(z_j))$ for $j = 1, \ldots, d$.

The time-consuming Step 1 above need only be done once, and then the required dependent quasi-random vectors can be rapidly generated (Step 2 can be performed very quickly; see Marsaglia, Zaman and Marsaglia 1994). These quasi-random vectors are analogous to i.i.d. random vectors with a distribution π say, with the required marginals and (at least approximately) the required covariance matrix.

When all of the marginal distribution functions F_i have densities f_i with respect to Lebesgue measure, it is a simple matter to specify the distribution π using the "change of variables technique"; see p. 408 of Apostol (1969) for example. When this is not the case, one can of course still specify π, but not in such a nice form.

The primary use for vectors distributed according to π is in numerical integration, and more specifically, in calculating an expected value with respect to the distribution π. In particular, one can approximate $\alpha \stackrel{\triangle}{=} \int_{\mathbf{R}^d} g(x)\pi(dx)$ via

$$\alpha_n \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n} g(x(i)),$$

where each vector $X(i)$ is generated according to the procedure outlined above. We will give two examples of such a calculation in Section 5.

5 NUMERICAL EXAMPLES

In this section we present two examples where we wish to compute $\alpha \stackrel{\triangle}{=} Eg(X)$ for some function g and some random vector X. Both examples are contrived, but serve to demonstrate the potential applicability of the ideas presented in this paper. We will compare the error of an estimator α_{mc} based on a pure Monte Carlo approach to one obtained using our proposed QMC methodology α_{qmc}. It is easy to assess the error in a pure Monte Carlo experiment using the sample variance. However, it is more difficult to assess the error using QMC methods.

In both of our examples, we will use an approach suggested by Cranley and Patterson (1976) for assessing the error in the QMC approach. In the QMC method, $Eg(X)$ is estimated via $n^{-1}\sum_{k=1}^{n} g(x(k))$ for some deterministic sequence of points $\{x(k)\}$. The sequence $\{x(k)\}$ is, in turn, based upon a deterministic sequence $\{u(k)\}$ of points in the d-dimensional unit hypercube $[0,1)^d$. The error may be assessed by randomly shifting the sequence $\{u(k)\}$ several times, each time computing an estimate of $Eg(X)$. The resulting estimates are i.i.d. and unbiased, and consequently the error may be assessed. The procedure is as follows.

1. Select a run length m, and for each $i = 1, ..., m$ perform Steps 2, 3 and 4 below.
2. Generate a random vector U that is uniformly distributed in $[0,1)^d$.
3. Compute $\tilde{u}(k) \stackrel{\triangle}{=} (u(k) + U) \bmod 1$, where the mod operation is performed componentwise, $k = 1, \ldots, n$.
4. Compute

$$Y_i = n^{-1}\sum_{k=1}^{n} g(\tilde{x}(k)),$$

   where the $\tilde{x}(k)$'s are obtained from the $\tilde{u}(k)$'s using a NORTA transformation.
5. Compute the sample mean α_{qmc} and sample variance v_m of $(Y_1, \ldots, Y_m)$.
6. Compute a confidence interval for α, given by $\alpha_{\text{qmc}} \pm z\sqrt{v_m/m}$, where z is chosen from normal tables to ensure the required confidence level.

An assessment of the error in the estimator α_{qmc} is the confidence interval halfwidth $z\sqrt{v_m/m}$.

To enable a fair comparison, the pure Monte Carlo estimator α_{mc} should be based on the same number of function evaluations, namely mn.

We used Sobol' sequences as implemented in Press et al. (1992) to generate the required quasi-random sequences for both examples.

Our first example is low-dimensional, and in such situations, other methods for numerical integration such as quadrature schemes are typically preferred to either Monte Carlo or quasi-Monte Carlo approaches. However, the example serves as a useful first demonstration that there is potentially value in the quasi-Monte Carlo approach.

Example 1 Suppose that $X = (X_1, X_2)$, where each X_i is exponentially distributed with mean 1, and the correlation between X_1 and X_2 is 0.4. We wish to compute $E(X_1 + X_2)$ using both standard Monte Carlo methods, and quasi-Monte Carlo methods. It is useful, for comparison purposes, to note that the exact answer is 2.

The ARTAGEN software package (Cario and Nelson 1997) was used to determine the appropriate normal correlation (0.4464) required in the NORTA procedure to arrive at a correlation of 0.4 between 2 exponential random variables. In our experiments, we first took $n = 128$ (a quasi-random sequence of length 128), and $m = 100$.

The resulting confidence interval for $E(X_1 + X_2)$ was 2.000 ± 0.007. In contrast, a confidence interval generated using the standard Monte Carlo method (again generating the X_i's using NORTA) with 12800 realizations gave a confidence interval of 2.02 ± 0.03. Thus, the quasi-random approach reduces variance by a factor of $(0.03/0.007)^2 \approx 18$ over standard Monte Carlo. (This variance reduction factor was recomputed several times, and each time was of the same order of magnitude, namely approximately 20.) It is, of course, important to also consider the time required to achieve these results. These results were each obtained in approximately 2 minutes of computation using MATLAB on a Sun Sparc 2. A more precise reporting of the times is not relevant, since no attempt was made to optimize the code for either of the implementations, and a more precise comparison depends on both the implementation and computer architecture. What is important is that the variance reduction reported above essentially "comes for free", in the sense that both methods require approximately the same amount of computation.

In fact, we expect that the variance reduction factor above will increase with n, since the quasi-random estimator is expected to converge at rate $n^{-1}(\log n)^2$, while the standard Monte Carlo estimator is expected to converge at rate $n^{-1/2}$. To gauge the rate of convergence of the quasi-random estimator, in Figure 1 we plotted the log of the confidence interval halfwidth as determined by the Cranley Patterson procedure versus n. The resulting graph clearly shows a slope of approximately -1, indicating that the rate of convergence is approximately of the order n^{-1}. Of course, we would not expect a term of the order $(\log n)^2$ to show up in such a plot.

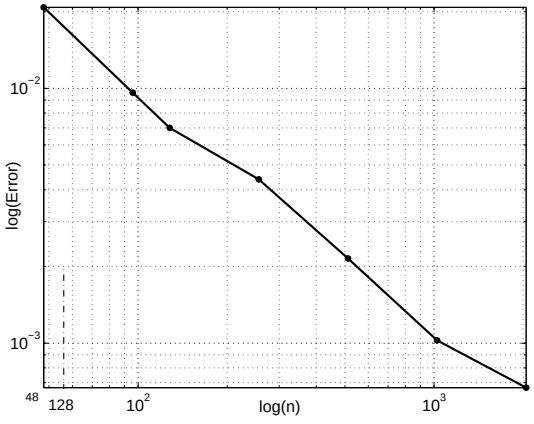

Figure 1: Log-log Plot of Confidence Interval Halfwidth Versus n for the Quasi-random Estimator in Example 1

This first example certainly lends support to the notion that the use of quasi-random numbers can lead to computational improvements in Monte Carlo calculations involving the use of the NORTA method. Our second example reinforces this notion through a more interesting application.

Example 2 Suppose that we wish to compute the expected length of the longest path in a stochastic activity network as shown in Figure 2. We assume that the time required to complete task (arc) j in the network is exponentially distributed with mean μ_j, and that the correlation matrix of the task durations is Σ_X. Specifically, we set

$$\mu = (10, 5, 12, 11, 5, 5)',$$

and

$$\Sigma_X = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.3 & 0 & 0 \\ & 1 & 0.5 & 0 & 0.3 & 0 \\ & & 1 & 0 & 0.5 & 0.3 \\ & & & 1 & 0.1 & 0.5 \\ & & & & 1 & 0.3 \\ & & & & & 1 \end{pmatrix}.$$

(Only the upper half of this positive semidefinite matrix is specified as it is symmetric.) The correlation matrix Σ_Z which yields Σ_X after a NORTA transformation was again obtained using the ARTAGEN software. The matrix Σ_Z is obtained from Σ_X by simply replacing all 0.5's with 0.54656, all 0.3's with 0.34208, and all 0.1's with 0.11936. The resulting Σ_Z is positive definite. We again took $m = 100$ and $n = 128$.

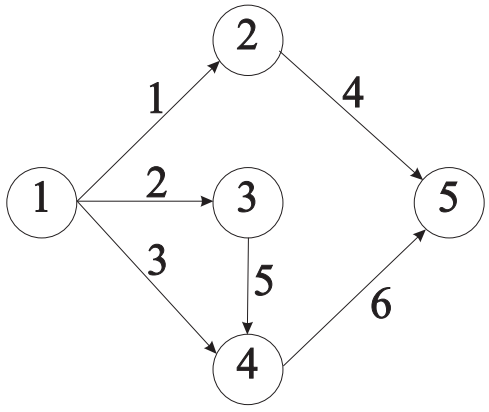

Figure 2: Stochastic Activity Network Example with Arc Labels as Shown

Once again, both the quasi-Monte Carlo approach and the standard Monte Carlo approach took approximately the same time to compute. The resulting quasi-Monte Carlo confidence interval for the expected length of the longest path was 26.52 ± 0.09. The corresponding standard Monte Carlo confidence interval was 26.6 ± 0.3. We see that the quasi-Monte Carlo estimator reduces variance from the standard estimator by a factor of $(0.3/0.09)^2 \approx 11$.

Furthermore, as in Example 1, we expect that the quasi-random estimator converges at a rate that is close to linear. Figure 3 below lends credence to this supposition.

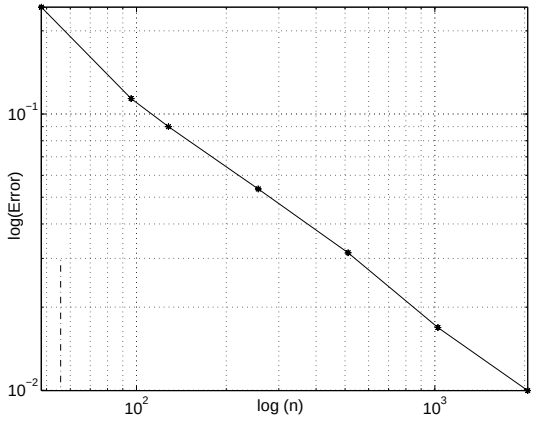

Figure 3: Log-log Plot of Confidence Interval Halfwidth Versus n for the Quasi-random Estimator in Example 2

In both of these examples the quasi-Monte Carlo estimator outperforms the standard Monte Carlo estimator, as we might have expected. These results reinforce the notion that the use of dependent quasi-random numbers can lead to useful efficiency improvements over estimators based on pseudo-random numbers, and the log-log plots above suggest that the improvements can be expected to grow without bound as the runlength increases.

Appendix

Before proving Theorem 2, we first give two preliminary results.

Lemma 5 *Suppose that X is a nondegenerate random variable, and that $g_1 : \mathbb{R} \to \mathbb{R}$ and $g_2 : \mathbb{R} \to \mathbb{R}$ are continuous, strictly increasing functions. If $Eg_i^2(X) < \infty$ for $i = 1, 2$, then $cov(g_1(X), g_2(X)) > 0$, i.e., the covariance is* strictly *positive.*

Proof: For $i = 1, 2$, define $g_i^{-1}(\cdot)$ as in (3). We have that

$$\begin{aligned}
&\operatorname{cov}(g_1(X), g_2(X)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(g_1(X) \le x, g_2(X) \le y) - \\
&\quad P(g_1(X) \le x)P(g_2(X) \le y)\, dx\, dy \qquad (8) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(X \le \min\{g_1^{-1}(x), g_2^{-1}(y)\}) - \\
&\quad P(g_1(X) \le x)P(g_2(X) \le y)\, dx\, dy \qquad (9) \\
&\ge 0.
\end{aligned}$$

The first equality (8) above is given in Whitt (1976), who attributes the result to Hoeffding (1940). The inequality follows since the integrand in (9) is given by $P(g_1(X) \le x)P(g_2(X) > y)$ if $g_1^{-1}(x) \le g_2^{-1}(y)$, and $P(g_1(X) > x)P(g_2(X) \le y)$ otherwise.

It remains to show that the above inequality is, in fact, strict. Define

$$V = \{v : P(X \le v)P(X > v) > 0\}$$

to be the set of values v such that X has positive probability of being both less than or equal to v, and greater than v. Note that V is an interval of strictly positive length, by our assumption that X is nondegenerate. Hence $g_i(V) = \{x : x = g_i(v), v \in V\}$ is an interval of strictly positive length, for $i = 1, 2$. Choose $v_1 \in V$, $v_2 \in V$ with $v_1 < v_2$.

Now select $w_1, w_2 \in [v_1, v_2]$ with $w_1 < w_2$, and define $x = g_1(w_1)$, and $y = g_2(w_2)$. Observe that

$$\begin{aligned}
P(g_1(X) \le x)P(g_2(X) > y) &= P(X \le w_1)P(X > w_2) \\
&> 0,
\end{aligned}$$

and that this holds for all $w_1 < w_2$ with $v_1 \le w_1 < w_2 \le v_2$. Equivalently, this holds for all x, y with

$$g_1(v_1) \le x < g_1(v_2) \text{ and } g_2(g_1^{-1}(x)) < y \le g_2(v_2). \qquad (10)$$

Observe that (10) is a set of positive Lebesgue measure (in 2 dimensions), and it immediately follows that the covariance (9) is strictly positive.

We also need the following result, which strengthens Lemma 2 of Cario and Nelson (1997).

Lemma 6 *Let (Z_1, Z_2) have a standard bivariate normal distribution with $cor(Z_1, Z_2) = \rho_1$. Let (N_1, N_2) have a standard bivariate normal distribution with $cor(N_1, N_2) = \rho_2 > \rho_1$. Let g_1, g_2 be continuous and strictly increasing, and suppose that $Eg_i(N)^2 < \infty$, where N has a standard normal distribution, for $i = 1, 2$. Then*

$$Eg_1(N_1)g_2(N_2) > Eg_1(Z_1)g_2(Z_2).$$

The proof of this result uses the result of Lemma 5, and is virtually identical to that of Theorem 1 of Cario and Nelson (1997). It is therefore omitted.

Proof of Theorem 2: Let $1 \leq i, j \leq d$, and set $g_1 = F_i^{-1}(\Phi(\cdot))$, and $g_2 = F_j^{-1}(\Phi(\cdot))$. Then g_1 and g_2 satisfy the conditions of Lemma 6, and so if X_i and X_j are generated via the NORTA method from Z_i and Z_j, we immediately see that EX_iX_j is a strictly increasing function of the correlation ρ between Z_i and Z_j. Hence the covariance between X_i and X_j is strictly increasing in ρ, and so (4) can have at most one solution.

REFERENCES

Abramowitz, M., and I. A. Stegun, eds. 1964. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* National Bureau of Standards, Washington D.C.

Alfakih, A. and H. Wolkowicz. 2000. Matrix completion problems. In *Handbook of Semidefinite Programming: Theory, Algorithms and Applications.* H. Wolkowicz, R. Saigal, L. Vandenberghe, eds, 533–545. Kluwer, Boston.

Apostol, T. M. 1967. *Calculus volume 1*, 2nd ed. Wiley, New York.

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A guide to simulation*. 2d ed. New York: Springer-Verlag.

Cario, M. C. and B. L. Nelson. 1997. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing* 10:72–81.

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Chiera, B.A. and Cooke, R.M. 2000. Dependent quasirandom numbers. In: Cottam, M.P., Harvey, D.W., Pape, R.P. and J. Tait (Eds), *Foresight and Precaution*, Volume 2, Balkema, Rotterdam, 1519-1523.

Clemen R. T., and T. Reilly. 1999. Correlations and copulas for decision and risk analysis. *Management Science* 45:208–224.

Cooke, R. 1997. Markov and entropy properties of tree- and vine-dependent variables. *Proceedings of the Section on Bayesian Statistical Science American Statistical Association,* Alexandria, VA, 166-175.

Cranley, R., and T. N. L. Patterson. 1976. Randomization of number theoretic methods for multiple integration. *SIAM Journal of Numerical Analysis.* 13:904–914.

Devroye, L. 1986. *Non-uniform random variate generation.* Springer-Verlag, New York.

Gentle, J.E. 1998. *Random number generation and monte carlo method*, Springer-Verlag, New York.

Ghosh, S. and S. G. Henderson. 2000. Random vectors with specified marginals and covariance matrix, and chessboard distributions. *In Preparation.*

Hill, R. R., and C. H. Reilly. 1994. Composition for multivariate random vectors. In *Proceedings of the 1994 Winter Simulation Conference*, 332–339, J. D. Tew, S. Manivannan, D. A. Sadowsky, A. F. Seila, eds. IEEE, Piscataway New Jersey.

Hoeffding, W. 1940. Masstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universitat Berlin* 5:179–233.

Iman, R., Helton, J. and J. Campbell. 1981. An approach to sensitivity analysis of computer models: Part I and II, *J. of Quality Technology*, 13(4):174-183,232-240.

Iman, R. and W. Conover. 1982. A distribution-free approach to inducing rank correlation among input variables, *Communications in Statistics - Simulation and Computation*, 11 (3):311-334.

Infanger, G. 1994. *Planning under uncertainty: Solving large scale stochastic linear programs.* Boyd and Fraser.

Johnson, M. E. 1987. *Multivariate statistical simulation.* Wiley, New York.

Kocis, L., and W. J. Whiten. 1997. Computational investigations of low-discrepancy sequences. *ACM Trans. on Math. Software.* 23(2):266-294.

Kruskal, W. 1958. Ordinal measures of association. *Journal of the American Statistical Association* 53:814–861.

Li, S. T., and J. L. Hammond. 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man, and Cybernetics.* 5:557–561.

Mardia, K. V. 1970. A translation family of bivariate distributions and Fréchet's bounds. *Sankhya.* A32:119–122.

Marsaglia, G., A. Zaman, and J. C. W. Marsaglia. 1994. Rapid evaluation of the inverse of the normal distribution function. *Statistics and Probability Letters.* 19:259–266.

Meeuwissen, A. 1993. *Dependent random variables in uncertainty analysis.* PhD dissertation, Dept. of Mathematics, Delft University of Technology, Report 94-28, Delft.

Meeuwissen, A. and T. Bedford. 1997. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *J. of Statistical Computation and Simulation*, 57(1-4):143-175.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NSF Regional Conference Series in Applied Mathematics 63. SIAM Philadelphia Pennsylvania.

Niederreiter, H., R. F. Tichy, and G. Turnwald. 1990. An inequality for differences of distribution functions. *Archiv Der Mathematik.* 54:166-172.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and B.P. Flannery. 1992. *Numerical recipes in FORTRAN 77: The art of scientific computing*, 2nd edition. Cambridge University Press.

Whitt, W. 1976. Bivariate distributions with given marginals. *Annals of Statistics* 4:1280–1289.

Wolkowicz, H., R. Saigal, L. Vandenberghe, eds. 2000. *Handbook of semidefinite programming: Theory, algorithms and applications.* Kluwer, Boston.

AUTHOR BIOGRAPHIES

SHANE G. HENDERSON is an assistant professor in the Department of Industrial and Operations Engineering at the University of Michigan. He was a lecturer in the Department of Engineering Science at the University of Auckland from 1997 to 1999. His research interests include discrete-event simulation, queueing theory and scheduling problems.

BELINDA A. CHIERA is a post-doctoral researcher at the Department of Mathematics of the Delft University of Technology. Her interests include the generation of quasirandom number sequences for high dimensional distributions with dependence, environmental modelling, scheduling problems and cryptology. She received her bachelors and doctorate degrees from the University of South Australia.

ROGER M. COOKE is a Professor of Applications of Decision Theory in the Department of Mathematics of the Delft University of Technology. His interests focus on mathematical aspects of risk analysis, including high dimensional multivariate techniques, uncertainty and sensitivity analysis, dependence modelling, design of reliability data bases and expert judgment methods. He received his bachelors and doctorate degrees from Yale University.