Applications of Scrambled Low Discrepancy Sequences To Exotic Options

Ken Seng Tan
Depart. of Stat. & Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: 519 885-1211 x 6713
Fax: 519 746-1875
Email: kstan@uwaterloo.ca

Phelim P. Boyle
Centre for Advanced Studies in Finance
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: 519 885-1211 x 6513
Fax: 519 888-7562
Email: pboyle@uwaterloo.ca

Abstract

This paper deals with a recent modification of the Monte Carlo method known as quasi random Monte Carlo. Under this approach, one uses specially selected deterministic sequences rather than random sequences as in Monte Carlo. These special sequences are known as low discrepancy sequences and have the property that they tend to be evenly dispersed throughout the unit cube. For many applications in finance,the use of low discrepancy sequences seems to provide more accurate answer than random sequences. One of the main drawbacks of the use of low discrepancy sequences is that there is no obvious method of computing the standard error of the estimate. This means that in performing the calculations, there is no clear termination criterion for the number of points to use. We address this issue here and consider a modified version of Owen's technique for overcoming this problem. We test these procedures using a high dimensional example of a derivative security. The exact price of this security can also be calculated very simply and so we have a benchmark against which to test our calculations. We find that our procedures give promising results and we are able to obtain standard errors that indicate the accuracy of our methods.

Keywords: Monte Carlo Simulation; Quasi-random Sequences; Faure Sequences; Randomized Nets; Numerical Finance; Exotic Option Valuation.

1 Introduction

The Monte Carlo simulation method is a powerful and flexible approach for providing numerical solutions to a large class of complex problems. In recent years, the Monte Carlo approach has been extensively used in the finance and investment area. Initially, the applications were mainly concerned with calculations related to the pricing of complex financial instruments and the computation of related hedging parameters. Examples of such instruments are mortgage-backed securities and various complex exotic options. More recently, Monte Carlo methods have been used to estimate the distribution of returns of entire portfolios. Applications include the calculation of credit risks and market risks and value at risk computations. Monte Carlo methods are also useful in several applications in actuarial science: examples include the simulation of asset and liability returns for asset-liability management purposes as well as in dynamic solvency testing.

We now discuss briefly why the Monte Carlo method is useful in the investment and finance area. In modern financial economics, security prices are modeled as stochastic processes to reflect future uncertainty. The current price of a security can be represented as the expected value of the future payouts on the security. This follows from the assumption of no-arbitrage. The expectation is taken with respect to a probability measure that is induced by the current price system. If we normalize the securities, the revised prices become martingales under this probability measure; hence it is often called the equivalent martingale measure. If a complex financial instrument has a payout that depends on the prices of several underlying securities or a payout that depends on the price path of an existing security, then its price can be written as a multi-dimensional integral. There are many different types of financial instruments of this nature. In some cases, the dimensions are quite large; for example, under mortgage-backed securities, the number of dimensions is as high as 360.

For high dimensional problems, the Monte Carlo method has strong advantages over alternative numerical integration schemes. In a recent survey article, Boyle, Broadie and Glasserman (1996) describe the method as follows:

> *Monte Carlo becomes increasingly attractive compared to other methods of numerical integration as the dimension of the prob-*

lem increases. Consider the integral of the function $f(x)$ over the [s]-dimensional unit hypercube. The simple (or crude) Monte Carlo estimate of the integral is equal to the average value of the function f over n points selected at random from unit hypercube. From the strong law of numbers this estimate converges to the true value of the integral as n tends to infinity. In addition, the central limit theorem assures us that the standard error of the estimate tends to zero as $1/\sqrt{n}$. Thus the error convergence rate is independent of the dimension of the problem and this is the dominant advantage of the method over classical numerical integration approaches. The only restriction on the function f is that it should be square integrable and this is a relatively mild restriction.

One disadvantage of the standard Monte Carlo method is that in some cases, notably for large scale problems, the rate of convergence is very slow. Different methods of speeding up the convergence have been proposed. These techniques are known as variance reduction techniques. For a summary, see Boyle, Broadie and Glasserman (1996). Recently, so-called quasi-Monte Carlo methods[1] or the low discrepancy (LD) methods have been used in finance applications. These methods rely on the use of specially selected deterministic sequences instead of random sequences. These deterministic sequences have the property that they are well dispersed throughout the unit cube and are known as low discrepancy sequences. The monograph by Niederreiter (1992) provides an excellent discussion of these sequences. Applications of low discrepancy sequences to finance problems are discussed by Boyle, Broadie and Glasserman (1996), Caflisch and Morokoff (1996), Joy, Boyle and Tan (1996), Ninomiya and Tezuka (1996) and Paskov and Traub (1995). There are differences in the relative efficiency of (LD) methods versus standard Monte Carlo for the typical finance problems and for other more general applications. In the case of finance problems, the LD methods appear to outperform standard Monte Carlo for some high dimensional problems. For example, in mortgage-backed security applications with dimensions as high as 360, Paskov and Traub (1995) report good results using Sobol' sequences while Ninomiya and Tezuka (1996) conclude generalized Niederreiter sequences are superior. For more general integrals - which can be non smooth or highly periodic - the superiority of low discrepancy algorithms vanishes for dimensions around 30 or even lower. For a comparison

of the two approaches in the case of more general integrands see Bratley, Fox and Niederreiter (1992) and Morokoff and Caflisch (1994),(1995). The advantages of LD methods for finance applications appears to stem from the smoothness of the integrand in many applications and the fact that the effective dimension in finance applications is sometimes lower than the actual dimension. However it is not alway's the case that the classical LD methods dominate standard Monte Carlo for finance applications. The relative effectiveness depends on several factors including the nature of the integrand and the properties of the sequence used to evaluate it.

Until recently a major drawback of the classical LD approach has been the absence of a reliable practical error bound. Even though there exists a deterministic upper bound, this theoretical bound significantly overestimates the actual error in practice. This is in contrast to the crude Monte Carlo method for which the standard error of the estimate is readily available. The purpose of the present paper is to investigate a modification of the technique proposed by Owen (1995) for overcoming this problem. We use a particular type of derivative security for this analysis. This security has the advantage that its price has an exact closed-form solution no matter how large the dimensions of the problem.

2 Basic Concepts Related to Discrepancies

The concept of discrepancy is useful in discussing how well points are dispersed throughout a region. In this section we describe this concept. Suppose we have an s-dimensional unit hypercube and a set of points scattered throughout this region. The discrepancy can be thought of as the greatest absolute difference between the continuous uniform probability and the discrete uniform probability, taken over all possible subcubes of $[0,1)^s$ containing the origin. Let V be a subcube of $[0,1)^s$ containing the origin and let $v(V)$ be its volume. The discrepancy D_N^* of the sequence $\{\boldsymbol{X}_i\}$ of N points is defined as

$$D_N^* = \sup_{V \in [0,1)^s} \left| \frac{\#\text{ of points in } V}{N} - v(V) \right|.$$

The expectation of the discrepancy of a random sequence can be shown to be bounded by $(\log(\log N))N^{-1/2}$. It is possible to construct sequences for which the discrepancy is smaller than that of a random sequence. Sequences

satisfying such criteria are known as low discrepancy sequences. Typically, an infinite sequence is said to be quasi-random or low discrepancy if the discrepancy bound satisfies

$$D_N^* \geq c_s \frac{(\log N)^s}{N},$$

where c_s is a constant that depends only on the dimension s. Examples of these sequences are given by Halton (1960), Sobol' (1967), Faure (1982) and Niederreiter (1987). In this paper, we only consider Faure sequences.

The importance of discrepancy can be seen from the integration error bound given by the Koksma-Hlawka inequality, which states that

$$\left| \int_{[0,1)^s} f(\boldsymbol{X}) d\boldsymbol{X} - \frac{1}{N} \sum_{i=1}^{N} \hat{f}(\boldsymbol{X}_i) \right| < V(f) \cdot D_N^*$$

where f is a function of bounded variation in the sense of Hardy and Krause, $\{\boldsymbol{X}_i\}$ is a sequence of N points in $[0,1)^s$ with discrepancy D_N^*, and $\hat{f}(\boldsymbol{X}_i)$ is the simulated function value. In essence, the Koksma-Hlawka inequality separates the integration error bound into two components. These correspond to the smoothness of the integrand and the uniformity of the sequence used in evaluating the function. For the same function f, the Koksma-Hlawka inequality implies that the sequence with the smaller discrepancy results in smaller error bounds. This suggests we should use a sequence with the smallest possible discrepancy in evaluating the function in order to achieve the smallest possible error bound. Low discrepancy methods strive to attain this goal. The Koksma-Hlawka inequality asserts that if we use low discrepancy sequences as the integration nodes in Monte Carlo integration, the absolute integration error is $\mathcal{O}((\log N)^s N^{-1})$. For large N, this rate of convergence is considerably faster than the standard Monte Carlo methods whose error bound is $\mathcal{O}(N^{-1/2})$.

The Koksma-Hlawka inequality provides theoretical justification for the application of low discrepancy sequences in Monte Carlo integration. In practice, it poses several difficulties. First, the two factors D_N^* and $V(f)$ are extremely hard to compute. Second, even if these factors can be estimated, the bound is not sharp and is only correct asymptotically. As argued by Caflisch and Morokoff (1994), the number of points, N, at which the low discrepancy sequence starts to exhibit asymptotic behaviour grows exponentially with dimension s. In actual applications, we generally will be quite

far away from the asymptotic behaviour and hence there is need for a more feasible way of characterizing the error bound. This issue is addressed in this paper.

Many of the low discrepancy sequences cited above can be considered as special cases of (t, m, s)-nets and (t, s)-sequences. We now give a brief introduction to these concepts. See Niederreiter (1992, Chapter 4) for detailed discussion.

An *elementary interval* in base b is an interval E in $[0,1)^s$ of the form

$$E = \prod_{j=1}^{s} \left[\frac{a_j}{b^{d_j}}, \frac{(a_j+1)}{b^{d_j}} \right)$$

with $d_j \geq 0, 0 \leq a_j < b^{d_j}$ and a_j, d_j are integers. An elementary interval E is thus a subinterval of the unit-cube $[0,1)^s$ whose j-th axis has length $1/b^{d_j}$. When we divide the j-th axis into b^{d_j} equal slices and repeat the division for other axes, the subinterval obtained is the elementary interval having volume $b^{-\sum_{j=1}^{s} d_j}$.

Let $0 \leq t \leq m$ be integers. A (t, m, s)-net in base b is a finite sequence with b^m points from $[0,1)^s$ such that every elementary interval in base b of volume b^{t-m} contains exactly b^t points of the sequence.

An infinite sequence of points $\{\boldsymbol{X}_i\} \in [0,1)^s$ is a (t, s)-sequence in base b if for all $k \geq 0$ and $m > t$, the finite sequence $\boldsymbol{X}_{kb^m}$, ..., $\boldsymbol{X}_{(k+1)b^m-1}$ forms a (t, m, s)-net in base b.

Sobol' (1967) describes how to construct (t, m, s)-nets and (t, s)-sequences in base 2. Faure (1982) provides a construction of the $(0, m, s)$-net and $(0, s)$-sequence in a prime base greater than or equal to s. Niederreiter (1987) generalizes the construction of the Sobol' sequence to arbitrary bases and the Faure sequence to bases that are of prime power greater than or equal to s.

To understand how the uniformity of a sequence is maintained when the sequence satisfies the net property, let us consider the Faure sequence. Joy, Boyle and Tan (1995) and Fox (1986) describe how to construct Faure sequences. For ease of illustration, we only consider the Faure sequence in 2-dimensions so that the points can be plotted on a graph and hence the distribution of the points in $[0,1)^2$ can be seen. Since the 2-dimensional Faure sequence corresponds to a (t, s)-sequence with $t = 0, s = 2$ and $b = 2$, this

implies that for all $k \geq 0$ and $m > 0$, the finite sequence

$$\boldsymbol{X}_{kb^m}, \ \ldots, \boldsymbol{X}_{(k+1)b^m-1} \tag{1}$$

is a $(0, m, 2)$-net in base b. More specifically, consider a partition of a 2-dimensional Faure sequence with $k = 8$ and $m = 3$ in sequence (1). The resulting sequence is therefore a (0,3,2)-net in base 2 with $2^3 = 8$ elements. The theory of nets guarantees that every elementary interval in base 2 (or rectangle in this case) with area 2^{-3} contains only one point of this subsequence. The rectangles of interest are of the form

$$\left[\frac{a_1}{2^{d_1}}, \frac{(a_1+1)}{2^{d_1}}\right) \times \left[\frac{a_2}{2^{d_2}}, \frac{(a_2+1)}{2^{d_2}}\right) \tag{2}$$

in $[0,1)^2$ with integers a_j, d_j such that $d_j \geq 0$ and $0 \leq a_j < 2^{d_j}$ for $j = 1, 2$. Another constraint on d_1 and d_2 can be derived by recognizing that for this subsequence, the elementary interval has area 2^{-3} and since the rectangle produced from (2) has area $2^{-(d_1+d_2)}$, we have

$$2^{-(d_1+d_2)} = 2^{-3},$$

for integers $d_1, d_2 \geq 0$. This implies

$$d_1 + d_2 = 3.$$

Since d_1 and d_2 are integers, solving above equation yields 4 sets of solutions, namely (0,3), (3,0), (1,2) and (2,1), where the first coordinate refers to d_1 while the second coordinate refers to d_2. It is clear from the above analysis that, in general there is no unique way of characterizing the elementary intervals in base b. Figure 1 demonstrates the 4 possible representation of the elementary intervals in base 2 corresponding to the 4 sets of solution of (d_1, d_2). In this figure and the subsequent figures, the horizontal and vertical axes represent, respectively, the first and second dimension of the point. A remarkable feature is that irrespective of how the elementary interval is developed, as long as the elementary intervals satisfy the necessary condition, each elementary interval will contain exactly one point, as it should.

One compelling advantage of a (t, s)-sequence is that points can be subsequently added without distorting the uniformity of the sequence. This is in contrast to other techniques such as lattice rules or stratifying sampling

methods where the number of points have to be preset. To see how the uniformity of the sequence is maintained when more points are added, let us consider introducing an additional 8 points to our existing sequence in the above example. The result is shown in Figure 2 with the "diamond"-shape points represent the first set of the 8 points while the "+"-shape points correspond to the subsequent set of 8 points. With 16 points in the sequence, the area of the elementary interval reduces to 2^{-4}. Each revised rectangle still contains only a single point. This should not be surprising since the sequence of 16 points in fact is a (0,4,2)-net in base 2 ($k = 4$ and $m = 4$ in (1)). The 3 panels in this figure again demonstrate that there is in general no unique way of characterizing the elementary intervals.

The above phenomenon can be explained as follows: consider the lower right panel of Figure 1. Suppose each of the rectangles is cut into 2 identical squares with area 2^{-4}. This is achieved by joining 2 horizontal lines from $(0, 0.25)$ to $(1, 0.25)$ and from $(0, 0.75)$ to $(1, 0.75)$. Sixteen identical squares are produced but only 8 of them contain a point. If points are to be added subsequently while maintaining the overall uniformity, the natural positions for these newcomers are those squares without any points. Consequently, each empty square is successively filled up by the newly added points. When exactly 8 points are added so that each of the 16 squares has exactly one point, a $(0, 4, 2)$-net in base 2 is formed as shown in Figure 2.

A similar phenomenon holds when the dimension of the sequence is greater than two. For a s-dimensional Faure sequence of b^m points , each s-dimensional elementary interval of volume b^{-m} contains only a single point. Since it is not possible to plot the points in $[0, 1)^s$, for $s > 2$ or 3, we use a simpler approach by examining the two-dimensional orthogonal projection of the low discrepancy points. We are implicitly assuming that if a sequence is uniformly dispersed in $[0, 1)^s$, then any two-dimensional orthogonal projections should also be uniformly dispersed. We do not prove this assertion but it seems to provide at least a necessary conclusion. For instance, Figure 3 plots the first and second coordinates of a 7-dimension Faure sequence with 6859 points. These points appear to be uniformly dispersed throughout the unit-square. Figure 4 provides a similar comparison except that random points are generated. From the graph, one can see that the random points tend to cluster and tend to have gaps. These are typical features of random points.

Even though the Faure sequence in high dimensions still maintains low discrepancy, undesirable features exist when we focus on their orthogonal

projections. In Figure 5 we plot nine pairs of the orthogonal projections of the 6859 ($= 19^3$) points from a 19-dimension Faure sequence. These nine pairs were selected at random. By merely increasing the dimensions from seven to nineteen, the orthogonal projection reveals an interesting phenomenon of the Faure points. The graph clearly suggests that undesirable correlation exists between these Faure points. Such patterns have also been pointed out by Morokoff and Caflisch (1994) and Boyle, Broadie and Glasserman (1996). As argued by Morokoff and Caflisch (1994), the anomaly is due to the large prime base used to generate the Faure points and is explained by Figure 6. The grid shows subsets of the elementary intervals with volume 19^{-3} for the subsets of the points in the first panel of Figure 5. According to the (t, m, s)-net property, the Faure sequence in base 19 with 6859 points must have exactly one point in each elementary interval of volume 19^{-3}. This is confirmed by Figure 6. The points between successive elementary intervals, on the other hand, form a regular pattern and are not uniformly distributed within these intervals. This behaviour leads to the highly correlated structure as shown in Figure 5. Such patterns, however, are not unique to Faure sequences. Braaten and Weller (1979) pointed out the existence of such problems in Halton sequence and Morokoff and Caflisch (1994) examined this phenomenon in various types of low discrepancy sequences.

3 Owen's Randomization Technique

The idea of combining Monte Carlo and LD methods has been proposed by several authors. Cranley and Patterson (1976) first introduce the randomization technique to number theoretic method. Braaten and Weller (1979) randomly permute the Halton sequence. Joe (1990) randomizes lattice rules. Faure (1992) provides an optimal permutation for a one-dimensional low discrepancy sequence. In this paper, we consider the randomization technique recently proposed by Owen (1995). This technique can be summarized as follows: let $\{\boldsymbol{A}\}$ be a (t, m, s)-net or a (t, s)-sequence in base b. Let $\boldsymbol{A}_i = (A_i^1, A_i^2, ..., A_i^s)$ denote the i-th term in the sequence. Each component of $\boldsymbol{A}_i$ can be expressed in its base b representation as

$$A_i^j = \sum_{k=1}^{\infty} a_{ijk}\, b^{-k}$$

where $0 \leq a_{ijk} < b$ for all i, j, k.

A scrambled version of $\{\boldsymbol{A}\}$ is a sequence $\{\boldsymbol{X}\}$ with components $\boldsymbol{X}_i = (X_i^1, ..., X_i^s)$ such that

$$X_i^j = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$$

where $x_{ijk} = \pi(\cdot)$ and π is a random permutation function on the digits $\{0, 1, ..., b-1\}$. Owen (1995) shows that the scrambled sequence $\{\boldsymbol{X}\}$ is also a (t, m, s)-net and (t, s)-sequence. The permutation functions π are rather complex and is described in detail in Owen (1995) and Hickernell (1996). The following description follows from Hickernell (1996). The π are chosen so that the resulting scrambled net satisfies the following conditions:

1. For any $\boldsymbol{X}_i \in \{\boldsymbol{X}\}$, each digit x_{ijk} is uniformly distributed on the set $\{0, 1, ..., b-1\}$.

2. For any two points $\boldsymbol{X}_i, \boldsymbol{X}_{i'} \in \{\boldsymbol{X}\}$, the corresponding pair $(X_i^j, X_{i'}^j)$, for $j = 1, \ldots, s$ are mutually independent.

3. Suppose $\boldsymbol{A}_i, \boldsymbol{A}_{i'} \in \{\boldsymbol{A}\}$ for $i \neq i'$, let $\boldsymbol{X}_i, \boldsymbol{X}_{i'} \in \{\boldsymbol{X}\}$ be the corresponding points in the scrambled net. Suppose further that in base b expansion, the components A_i^j and $A_{i'}^j$ share the same first k_{ij} digits but their $k_{ij} + 1$st digits are different; i.e. $a_{ijk} = a_{i'jk}$ for $k = 1, \ldots, k_{ij}$ and $a_{ijk_{ij}+1} \neq a_{i'jk_{ij}+1}$, then

   - $x_{ijk} = x_{i'jk}$, for $k = 1, \ldots k_{ij}$.
   - the random vector $(x_{ijk_{ij}+1}, x_{i'jk_{ij}+1})$ is uniformly distributed on the set $\{(n, n') : n \neq n'; n, n' = 0, 1, \ldots, b-1\}$.
   - $x_{ijk_{ij}+2}, x_{ijk_{ij}+3}, \cdots, x_{i'jk_{ij}+2}, x_{i'jk_{ij}+3}, \cdots$ are mutually independent.

Condition 1 implies that the marginal probability distribution of any point $\boldsymbol{X}_i \in \{\boldsymbol{X}\}$ is uniform on $[0, 1)^s$ while Condition 2 ensures that $(X_{i1}, X_{i2}, \ldots, X_{is})$ are mutually independent. These two conditions are typically satisfied by any random sequence. For a sequence to preserve low discrepancy even after scrambling, the correlation between different points in $\{\boldsymbol{X}\}$ must be maintained. This property is guaranteed by Condition 3. In the case of a random sequence, $x_{ij1}, x_{ij2}, \ldots, x_{i'j1}, x_{i'j2}$ are mutually independent whenever $\boldsymbol{X}_i \neq \boldsymbol{X}_{i'}$, instead of Condition 3.

After randomization, each individual point in the scrambled net has a uniform distribution in $[0,1)^s$. The sample variance of the estimate can the be estimated statistically by replications.

3.1 Computational Issues

To implement Owen's randomization technique, one is faced with two problems. The first problem is that the expansion of X_j must be truncated at some finite $K_{\max}$. One choice suggested by Owen is to take $K_{\max}$ large enough so that $b^{-K_{\max}}$ is small compared to the error committed in truncating the expansion. An alternate choice is take $K_{\max} = M$ if there are at most B^M points will ever be used.

The second practical issue is the memory storage problem. To scramble the k-th digit in the j-th dimension, the random permutation π depends not only on a_{ijk}, but also on the realized values of the $k-1$ digits $\{a_{ij1}, \ldots, a_{ijk-1}\}$. This implies that the random permutation π is state-dependent. Hence, scrambling an s-dimensional net requires $s(1 + b + b^2 + \ldots + b^{K_{\max}-1}) = s\frac{b^{K_{\max}}-1}{b-1}$ permutations. If b storage locations are used to store a permutations, then $sb\frac{b^{K_{\max}}-1}{b-1}$ storage locations are required. In other words, such a technique becomes impractical for high dimension s, large base b or large $K_{\max}$.

In this paper, we consider two simplified version of Owen's randomization techniques. Both of these randomization procedures maintain the essential features of Owen's technique and yet are feasible for high dimensional scrambling. The first technique requires only sb storage locations while the second technique requires $bK_{\max}$ storage locations. For ease of reference, we refer these two proposed modifications as Randomized A and Randomized B respectively.

3.2 Orthogonal Projections

We now compare the effects of the randomization techniques by examining the orthogonal projections. We apply the randomized technique B to the same sets of points reported in Figure 5. Similar to Figure 6, Figure 7 shows subsets of the elementary intervals after randomization. The randomized procedure effectively destroys the regular structure displayed in Figure 6 to

the extent that each elementary interval still contains a single point. Figure 8 gives the same set of the orthogonal projections for the randomized Faure points. The randomized Faure appears to eliminate the regularities observed in the classical Faure sequence. The randomized points are more uniformly dispersed throughout the unit square and do not follow any specific structure. Similar phenomenon is observed if the randomized algorithm A were used.

4 Numerical Example

As we mentioned earlier, the current price of a European derivative security can be written as its discounted expectation under the equivalent martingale (or Q measure) as

$$\text{Price} = e^{-rT} E_Q[g(\boldsymbol{S})] \tag{3}$$

where r is the riskless rate and $g(\cdot)$ is the payoff function that depends on the underlying state variable(s) $\boldsymbol{S}$.

In the simplest case, when g depends only on the terminal price of a single underlying asset (stock); i.e

$$g = \max[0, S(T) - K]$$

where $S(T)$ is the asset price at maturity T and K is the strike price. If the asset price follows a geometric Brownian motion with constant drift and volatility, equation (3) can be evaluated exactly and reduces to the celebrated Black-Scholes formula.

For more complicated structures of the payoff function, the expectation in (3) generally does not have a simple analytic solution. One typical example is the arithmetic path average option where the payoff function depends on the history of the asset prices as

$$g = \max\left[0, \frac{\sum_{i=0}^{s} S(t_i)}{s+1}\right]$$

where $0 = t_0 < t_1 \cdots < t_s = T$, and the time point t_i is the reset point at which the asset price is sampled and $s+1$ is the total number of discrete sampling points. Analytic solutions for this type of security are very complicated, involving multiple integrals of oscillating functions. Monte Carlo simulation is a common technique for pricing this type of option.

Since the purpose of this paper is to compare the efficiency of the randomized LD methods to the classical LD methods, particularly in the high dimensional case, it is important to use option problems that admit simple analytic solution for any finite dimension so that the simulation results can be benchmarked.

Two possible choices are available. If the average in the path average option is taken to be geometric, rather than arithmetic, we have

$$g = \max\left[0, \left\{\prod_{i=0}^{s} S(t_i)\right\}^{\frac{1}{s+1}}\right].$$

For this choice of the payoff function, the analytic solution is known and is of the form similar to Black-Scholes since the product of the lognormal variates is also a lognormal variates. Similarly, we could have defined a payoff function which depends on the geometric average of the terminal prices on a portfolio of s assets as

$$g = \max\left[0, \left\{\prod_{i=1}^{s} S_i(T)\right\}^{\frac{1}{s}}\right],$$

where $S_i(T)$ is the terminal price for asset i on maturity T. This type of option is known as the geometric portfolio average option and also has a closed-form representation of the form similar to Black-Scholes.

To approximate the prices of these two types of options using both Monte Carlo and LD methods, the procedure can be summarized as follows: For each realization of the simulation path, the discounted payoff of the option is computed. This procedure is repeated for many simulation trials with the final option value calculated as the average over all the realized option values. As guaranteed by the strong law of numbers, the estimated value converges to the analytic value as the number of simulations tends to infinity.

In the case of the geometric path average option, it is necessary to simulate a sequence of stock prices path $\{\hat{S}(t_1), \hat{S}(t_2), \ldots, \hat{S}(t_s)\}$. Since the stock price is lognormally distributed with $\log(S(t_i)/S(t_{i-1}))$ normally distributed having mean $(r - \frac{\sigma^2}{2})\frac{T}{s}$ and variance $\frac{\sigma^2 T}{s}$, the simulated stock prices $\{\hat{S}(t_1), \hat{S}(t_2), \ldots, \hat{S}(t_s)\}$ are generated recursively using

$$\hat{S}(t_i) = \hat{S}(t_{i-1}) e^{(r-\frac{\sigma^2}{2})\frac{T}{s} + \sigma\sqrt{\frac{T}{s}}\epsilon_i}, \qquad i = 1, \ldots, s \tag{4}$$

where $\hat{S}(t_0) = S(0)$ and $\{\epsilon_1, \ldots, \epsilon_s\}$ are independent random samples from the standardized normal distribution. For Monte Carlo methods, the standardized normal variates $\{\epsilon_1, \ldots, \epsilon_s\}$ are generated using polar transformation or the Box-Muller method. For LD methods, we generate points $\{x_1, \ldots, x_s\}$ in $[0,1)^s$ and then invert to normal variates $\{\epsilon_1, \ldots, \epsilon_s\}$ using the procedure suggested by Moro (1995).

For the geometric portfolio average option, we have to simulate the terminal stock prices of the s stocks. Assuming that the s stock prices have a multivariate lognormal distribution, the terminal stock prices can be simulated jointly as

$$\left.\begin{aligned} \hat{S}_1(T) &= S_1(0)e^{(r-\frac{\sigma_1^2}{2})T+\sigma_1\sqrt{T}\xi_1} \\ \hat{S}_2(T) &= S_2(0)e^{(r-\frac{\sigma_2^2}{2})T+\sigma_2\sqrt{T}\xi_2} \\ &\vdots \\ \hat{S}_s(T) &= S_s(0)e^{(r-\frac{\sigma_s^2}{2})T+\sigma_s\sqrt{T}\xi_s} \end{aligned}\right\} \tag{5}$$

where $\{\xi_1, \ldots, \xi_s\}$ is a random sample from an s-variate normal distribution. A convenient way of obtaining $\{\xi_1, \ldots, \xi_s\}$ is to first generate s independent standardized normal variates $\{\epsilon_1, \ldots, \epsilon_s\}$. These standardized normal variates are in turn transformed to s-variate normal distribution $\{\xi_1, \ldots, \xi_s\}$ using the methods suggested by Barr and Slezak (1972) or Scheuer and Stoller (1962). As in the case of geometric path average option, the standardized normal variates $\{\epsilon_1, \ldots, \epsilon_s\}$ are inverted directly from the low disrepancy point $\{x_1, \ldots, x_s\}$ for the LD methods.

For both types of options, s denotes the dimensionality of the simulation problems. The effective dimension of the problem, on the other hand, is considerably less than s. This has an important impact on the efficiency of the LD methods. When the effective dimension is small relative to s, this implies that there is a greater emphasize on the lower dimensions. This particular case favours the use of the LD methods since it is well known that low discrepancy sequences are more uniform in lower dimensions than in higher dimensions. In other words, when two problems have the same nominal dimension, s, but different effective dimensions, we should expect that the LD method is more efficient for the problem that has lower effective dimension. For this reason, it is important to choose an option problem in which the effective dimension is as close to the nominal dimension as possible

in order to truly reflect the effectiveness of the LD methods.

We now demonstrate the effective dimension of geometric path average option is in general less than the geometric portfolio average option and hence the geometric portfolio average option will serve as our base case for the subsequent studies. Using (4), the geometric path average, $\bar{S}_{\text{path}}$ can be simulated as

$$\begin{aligned}\bar{S}_{\text{path}} &= [\hat{S}(t_0)\hat{S}(t_1)\cdots\hat{S}(t_s)]^{\frac{1}{s+1}} \\ &= S(0)\exp\{\frac{T}{2}(r-\frac{\sigma^2}{2})+\frac{\sigma\sqrt{\frac{T}{s}}}{s+1}(s\epsilon_1+(s-1)\epsilon_2+\cdots+\epsilon_s)\} \quad (6)\end{aligned}$$

while the geometric portfolio average, $\bar{S}_{\text{portfolio}}(T)$, is simulated according to

$$\begin{aligned}\bar{S}_{\text{portfolio}}(T) &= [\hat{S}_1(T)\cdot\hat{S}_2(T)\cdots\hat{S}_s(T)]^{\frac{1}{s}} \\ &= \left[\prod_{i=1}^{s}S_i(0)\right]^{\frac{1}{s}}\exp\left[\frac{T}{s}\sum_{i=1}^{s}(r-\frac{\sigma_i^2}{2})+\frac{\sqrt{T}}{s}\sum_{i=1}^{s}\sigma_i\xi_i\right]. \quad (7)\end{aligned}$$

Expression (6) reveals that in simulating $\bar{S}_{\text{path}}$, the contribution of ϵ_i decreases as i increases. Hence, the earlier ϵ_i are relatively more important. Since $\{\epsilon_1,\ldots,\epsilon_s\}$ are one-to-one mapped from $\{x_1,\ldots,x_s\}\in[0,1)^s$ in the LD methods, this implies that the coordinates of x_i in lower dimensions are relatively more important than the x_i in higher dimensions. To simulate $\bar{S}_{\text{portfolio}}$, expression (7) suggests that the contribution of ξ_i are equally important to the extent that the σ_i do not vary very much. In the case of LD methods, $\{\xi_1,\ldots,\xi_s\}$ are transformed via the low discrepancy points $\{x_1,\ldots,x_s\}$, the effect of the coordinates in lower dimension is not as dominance as in the case of simulating $\bar{S}_{\text{path}}$. For this reason, the numerical studies carried out in the next section only focus on the geometric portfolio average options.

5 Numerical Comparisons

The numerical experimentation is carried out in two phases. For the first phase, we compare the performance of the standard Monte Carlo, classical Faure and randomized Faure based on the randomization techniques A

and B by simulating the geometric portfolio average options. For Monte Carlo methods, we use the generator RAN2 from Press et. al. (1992). In each method, we compute the root-mean-squared (RMS) relative error defined by

$$\text{RMS} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{\hat{C}_i - C_i}{C_i}\right)^2}$$

where i is the i-th option problem out of the $50(= m)$ randomly generated problems, and C_i and $\hat{C}_i$ are respectively, the analytic value and the estimated value for the i-th option problem. The parameter values for the 50 option problems are obtained as follows: The strike price is fixed at 100 and correlation between stocks is also constant at 0.5. The rest of the parameter values are generated randomly such that each initial stock price is uniformly distributed between 50 and 150, the annual volatilities are uniform between 10% and 60%, the expiration date is uniform between 6 months and 2 years and the annual interest rate is uniform between 5% and 15%. If the true option value for the set of randomly generated parameter values falls below 0.5, it is discarded and is replaced by another randomly generated set until the option value is at least 0.5. Very small option values may lead to less reliable estimates of RMS relative errors.

We consider the geometric portfolio average call options having dimensions $s = 10, 50, 100$. In the case of classical Faure, $200,000$ points are generated and the same set of points are used to compute the 50 randomly generated option problems. On the other hand, the random sequence with different starting seed is used for each option problem while for the randomized Faure, different random permutation is applied to each option problem. In all cases, the RMS relative errors are computed at intervals of 5,000 points. Using different seeds for the Monte Carlo method and different random permutations for randomized Faure for each problem lead to a smoother transition of the RMS relative errors as the number of points increase.

Figure 9 plots the RMS relative errors in percentage generated from the classical Faure sequences. An immediate conclusion can be drawn from this graph is the deterioration of the rate of convergence as the dimension of the problem increases. Figures 10, 11 and 12 compare the RMS relative errors for the various methods discussed above. For $s = 10$, the classical and the randomized Faure are similar and are more efficient than the standard Monte Carlo method. When the dimensions of the options increase to 50 and 100,

the convergence rate for classical Faure deteriorates and eventually becomes worse than the Monte Carlo method. Both the randomized A and B, on the other hand, remain competitive and indicate superior rate of convergence. We also looked at the geometric portfolio average put options and the same relative rate of convergence is also observed.

In the second phase of the comparison, we investigate the effectiveness of the randomization procedure in providing the error estimates. Since the classical Faure is purely deterministic, we focus our comparison on randomized LD methods using randomization techniques A and B to the standard Monte Carlo and the Monte Carlo with antithetic variates.

The standard errors of the Monte Carlo method can be obtained as follows: suppose f_i is the result obtained from the i-th Monte Carlo simulation. The unbiased Monte Carlo estimates of the mean and variance are

$$\begin{aligned} \bar{f} &= \frac{1}{N}\sum_{i=1}^{N} f_i \\ s_f^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \bar{f})^2 \end{aligned}$$

where N is the number of simulations. Assuming that each simulated estimate is approximately normally distributed with mean equal to the true value of the integral and some finite variance. We can construct an approximate 95% confidence interval for the estimated mean as $\bar{f} \pm 1.96\sqrt{s_f/N}$.

We obtain the Monte Carlo estimates based on antithetic variates as follows: for each simulation run, two parallel estimates of the option prices are obtained. The first estimate, $\tilde{f}_i$, is computed using the s independent standardized normal variates $\{\epsilon_1, \ldots, \epsilon_s\}$ while the second estimate, $\tilde{f}_i^A$, is generated from $\{-\epsilon_1, \ldots, -\epsilon_s\}$. The overall estimate of the option price for this particular simulation trial is given by the average of these two estimates. In other words, we have

$$f_i = \frac{\tilde{f}_i + \tilde{f}_i^A}{2}.$$

as the option estimate for the i-th simulation trial. To have a fair comparison across various techniques, if N simulations are used in standard Monte Carlo method, then $N/2$ simulations are carried out for Monte Carlo with antithetic variates since in this case, each estimate essentially involving two simulation

runs. The standard errors are computed similarly as in standard Monte Carlo methods.

The error estimates of the randomized LD methods are obtained by replication. The total number of simulations is divided equally into R batches so that R replications of M simulations are performed independently using different random permutations. Let $\bar{g}_j$ be the estimated mean for batch j based on M simulations and $\bar{g}$ be the overall mean of the R replications; i.e.

$$\bar{g} = \frac{1}{R}\sum_{j=1}^{R} \bar{g}_j.$$

The unbiased estimate of the variance[2] of $\bar{g}_j$ is

$$s_g^2 = \frac{1}{R-1}\sum_{j=1}^{R}(\bar{g}_j - \bar{g})^2.$$

Similarly, R and M are chosen so that $R \times M = N$ in order to have a fair comparison between different simulation techniques. In our numerical comparisons, 10 and 30 replications are used. To construct an approximate 95% confidence interval with 30 replications, the normal assumption is appropriate and is constructed in the same way as the standard Monte Carlo methods. On the other hand, with 10 replications, the t-distribution is more appropriate and the 95% confidence interval is constructed as $\bar{g} \pm 2.26\sqrt{s_g/10}$, where the constant 2.26 comes from the t-distribution with 9 degrees of freedom.

For our numerical comparison, we consider geometric portfolio call average option with $s = 100$ and initial stock prices $S_j(0) = 100$, $\sigma_j = 0.3$ for $j = 1, \ldots, 100$, the correlation between stocks is $\rho_{ij} = 0.5$ for $i, j = 1, \ldots, 100, i \neq j$, time to maturity is 1 year and the annual interest rate is 10%. The strike price is chosen so that both call and put options have equal values.

The two panels in Figure 13 compare the upper and lower limits of the constructed 95% confidence interval for both randomization techniques A and B. In both cases, the approximate 95% confidence interval contains the true solution most of the time, regardless of whether it is replicated 10 or 30 times, indicating that the constructed confidence intervals are reasonable. Figure 14 compares the computed standard errors for the 4 approaches discussed above. The Monte Carlo method with the antithetic adjustment

is efficient in reducing the variability of the estimates, as evidenced by its smaller standard errors relative to the crude Monte Carlo method. On the other hand, the standard errors obtained from the randomized Faure are considerably smaller than both Monte Carlo methods, indicating the superiority of the randomized Faure in reducing the variability. Based on the first part of the comparison on 50 randomly generated option problems, we conclude that both randomizations A and B are equally efficient. However, when we examine the standard errors generated by both procedures, randomization A is more favourable. As the number of simulation increases, we should expect the standard error to decrease. The standard errors computed using randomized A have a smoother transition as number of simulation increases while those using randomized B fluctuate more wildly.

6 Conclusion

The classical approach to the implementation of low discrepancy sequences for the solution of problems in the finance area suffers from some drawbacks. This paper has examined ways to rectify some of the disadvantages of LD methods. These new methods try to combine the best features of both standard (crude) Monte Carlo and classical LD method. We found that our modification of scrambling procedure suggested by Owen (1995) improves the convergence rate as compared to the classical LD approach. In addition, our procedure permits the generation of confidence intervals. The availability of confidence intervals is useful because it provides a scientific method for determining the accuracy of the estimation procedure and thus providing practical termination criteria.

It is important to stress the following caveats. In this paper our investigations dealt with just one particular type of low discrepancy sequence - Faure sequences. Furthermore, all the numerical experiments were based on one particular type of derivative: the portfolio geometric average option. However, our results are consistent with those of Owen and Tavella (1996) for different problems[3]. These authors also examined the performance of scrambled nets for two low dimensional problems and documented their superiority over both standard Monte Carlo and LD methods based on Sobol' sequences. Nevertheless, it is clear that more research work needs to be done in this area. In particular we would like to see how the scrambling

approach performs across a range of complex derivatives for different types of low discrepancy sequences.

One important aspect of these comparisons relates to the computational requirement for the different methods. A technique discussing the trade-off between accuracy and computational time is available in Boyle, Broadie and Glasserman (1996). We plan to examine this issue in future work.

Acknowledgement: This work was partly supported by the Natural Sciences and Engineering Research Council of Canada and the Society of Actuaries PhD Grants.

Notes

[1]This is an unfortunate misnomer. There is nothing random about this method. Nevertheless, the name quasi-Monte Carlo is widely used to describe the evaluation of integrals using low discrepancy sequences. We prefer to call this approach the low discrepancy (LD) method.

[2]Owen (1996) proposes the following method for estimating the variance. Instead of using R independently scrambled sequences, a single scrambled sequence of N points is generated. These N points are carefully divided into R equally batches so that each of the N/R sub-points is also a (t, m, s)-net. The variance estimate using this approach is however, biased.

[3]Owen and Tavella (1996) based their analysis on problems involving only 6 dimensions. Our results are based on much higher dimensional problems.

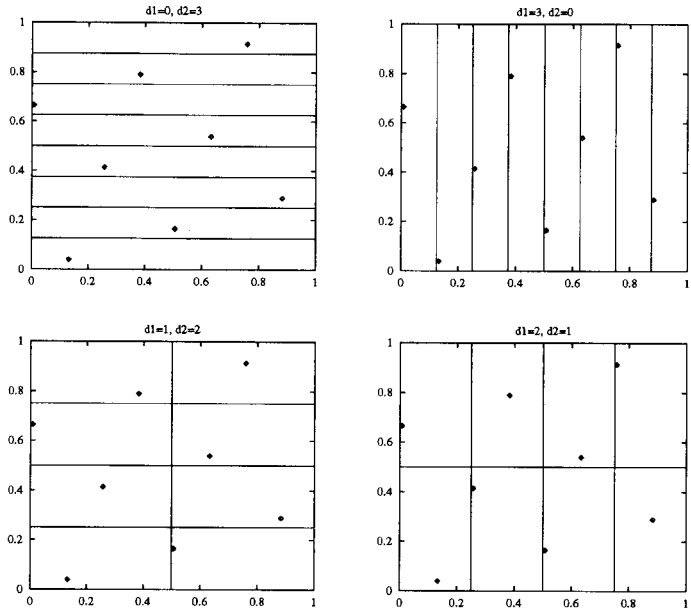

Figure 1: 4 Characterizations of Elementary Intervals for the First 8 Points

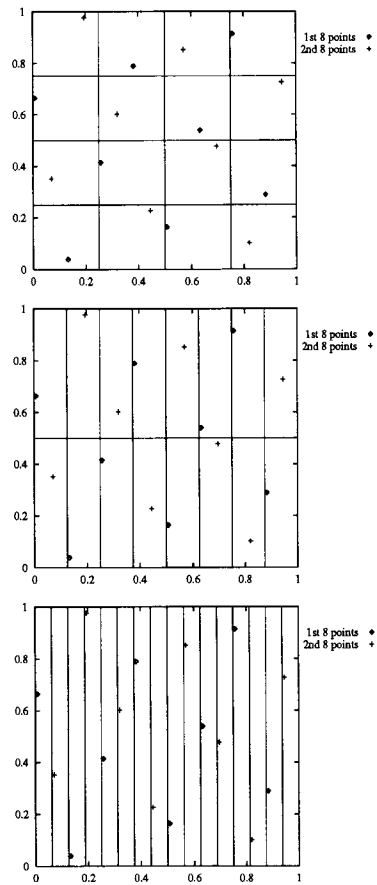

Figure 2: Different Characterization of Elementary Intervals with 8 More Points

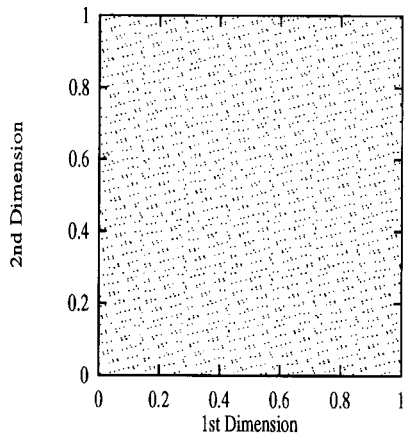

Figure 3: Orthogonal Projection of a 7-dimension Faure Sequence

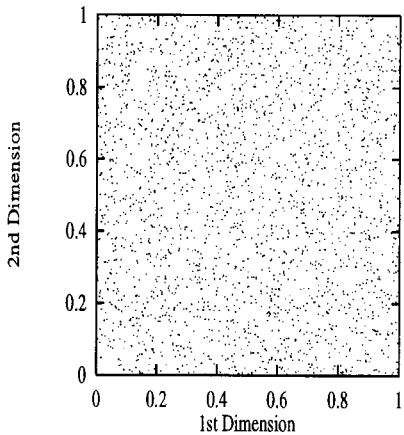

Figure 4: Orthogonal Projection of a 7-dimension Random Sequence

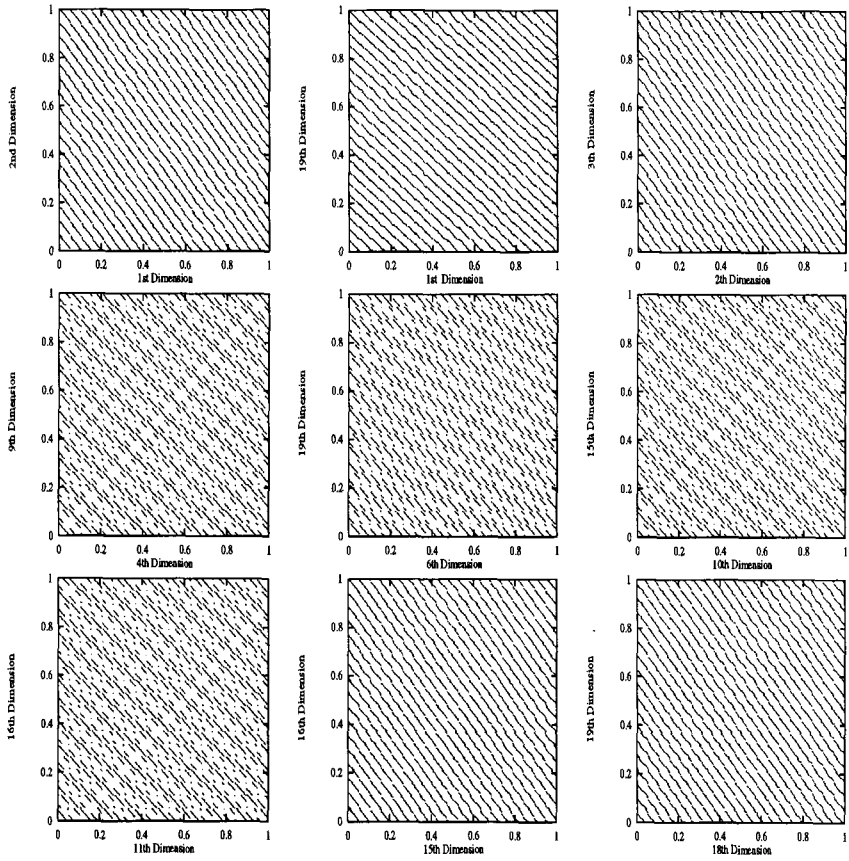

Figure 5: Orthogonal Projection of a 19-dimension Classical Faure Sequence

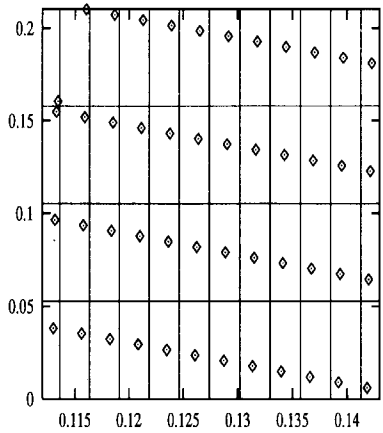

Figure 6: Blowup of Subsets of Faure Points from First Panel of Figure 5

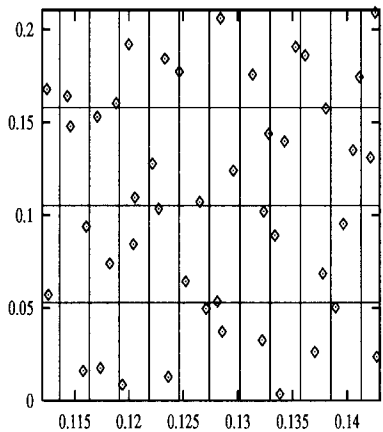

Figure 7: Blowup of Subsets of 6859 19-dimension Randomized Faure Points

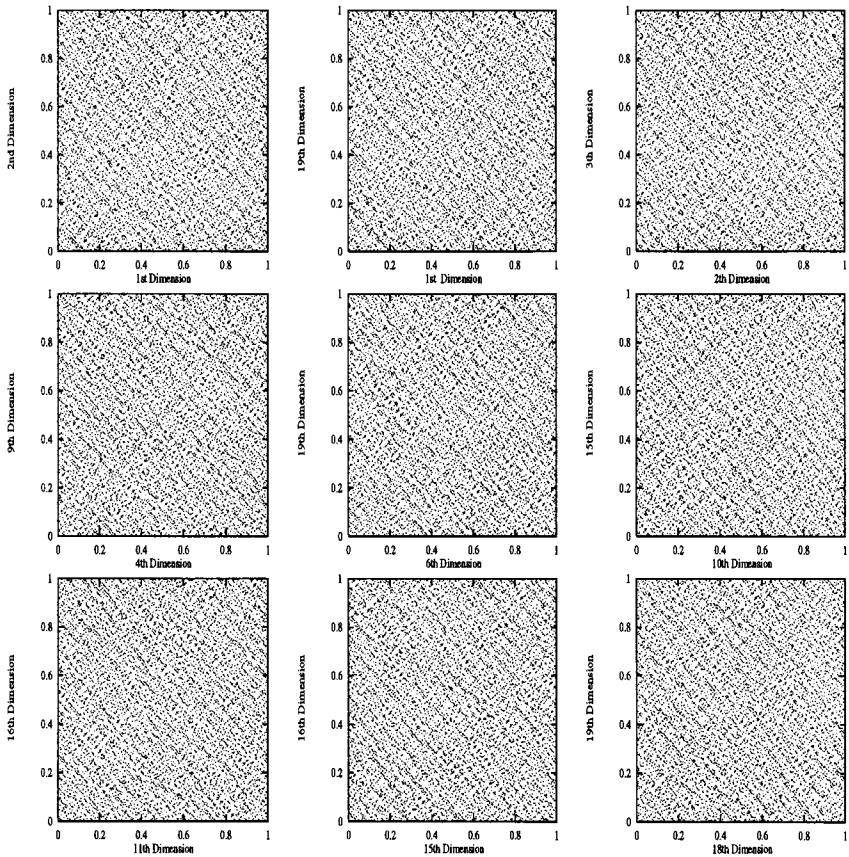

Figure 8: Orthogonal Projection of the 19-dimension Randomized Faure Sequence

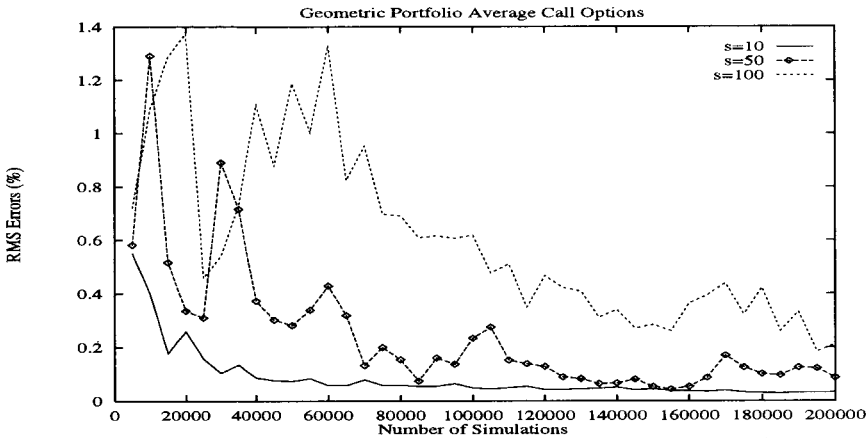

Figure 9: Efficiency of Classical Faure with Increasing Dimensions

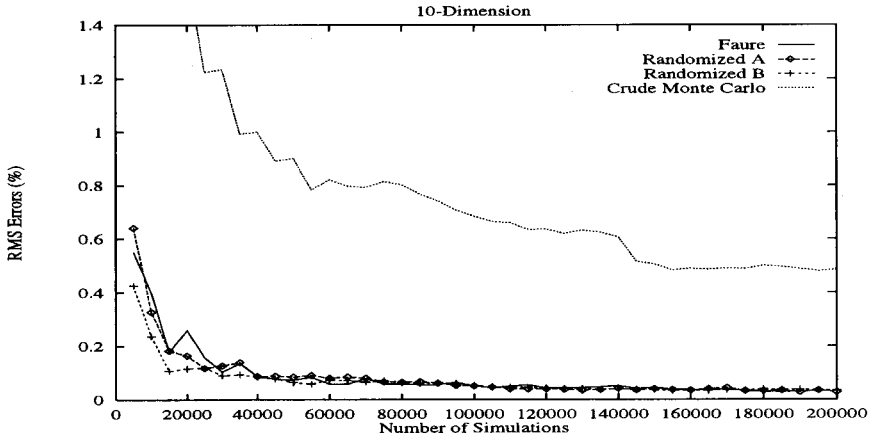

Figure 10: Efficiency of Various Techniques with $s = 10$

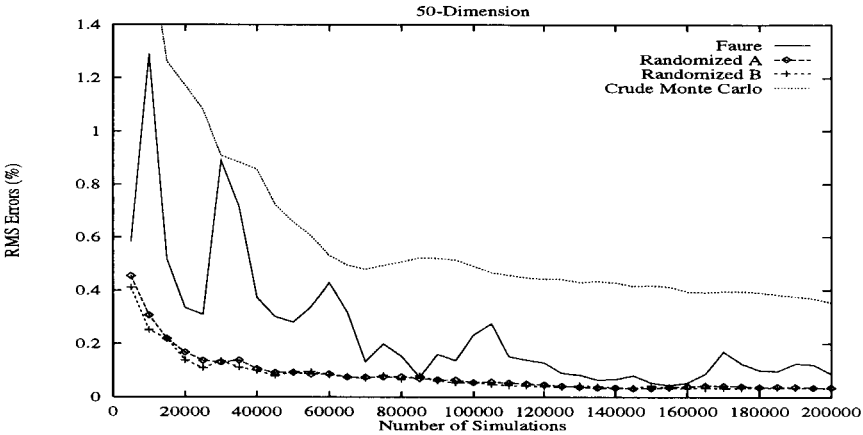

Figure 11: Efficiency of Various Techniques with $s = 50$

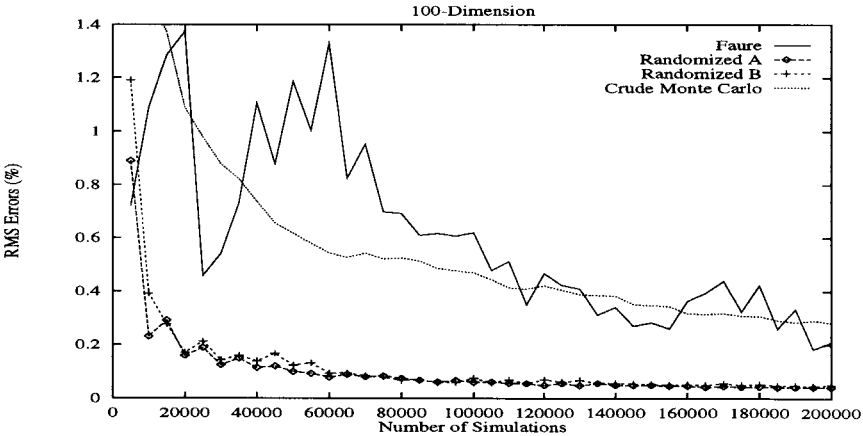

Figure 12: Efficiency of Various Techniques with $s = 100$

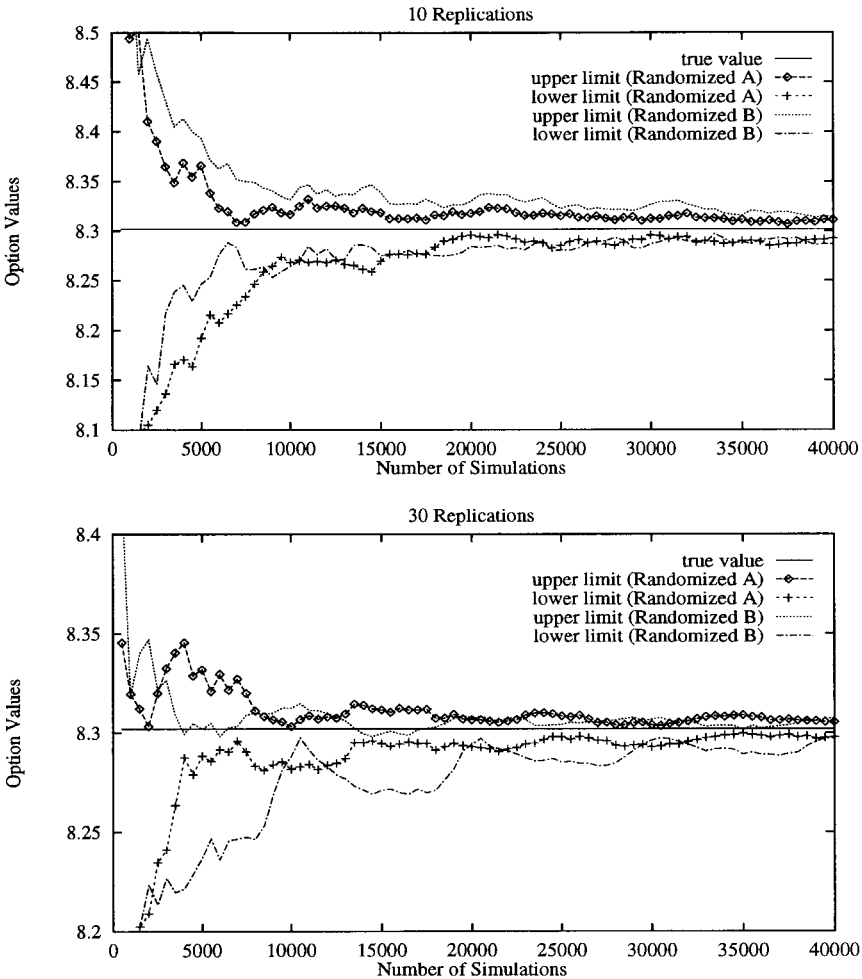

Figure 13: 95% Confidence Intervals from Randomized Faure with 10 and 30 Replications

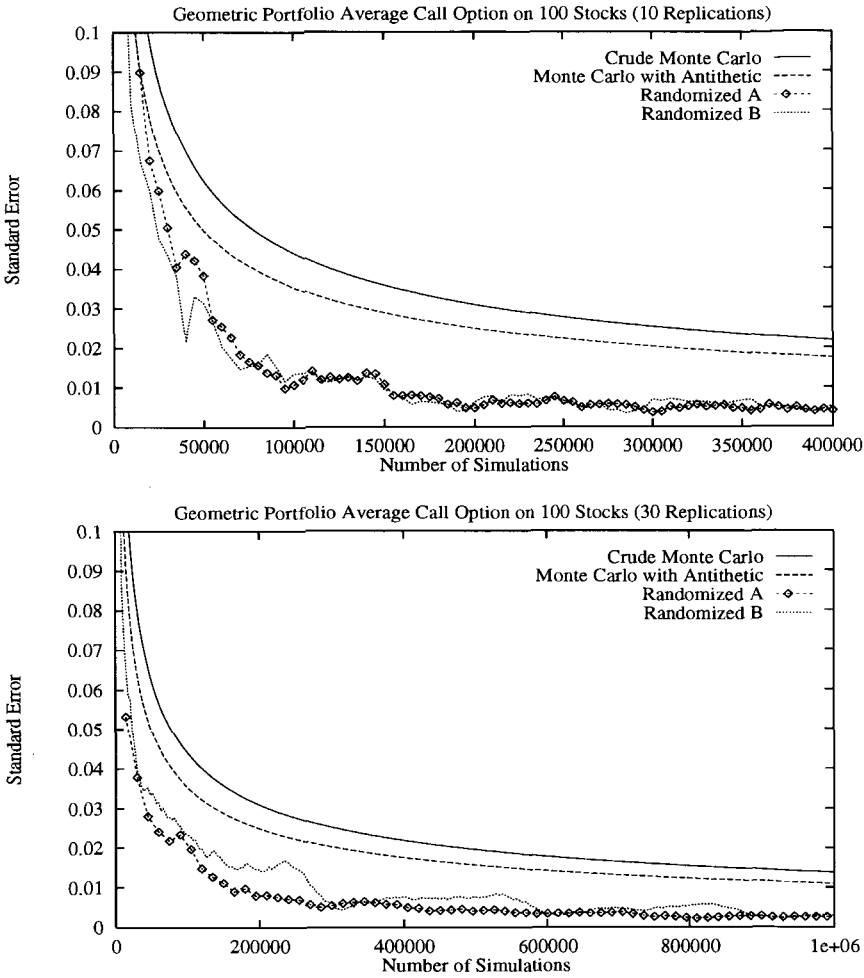

Figure 14: Comparison of Standard Errors with 10 and 30 Replications

References

[1] Barr, D.R. and N.L. Slezak "A comparison of multivariate normal generators," *Communications of the ACM,* 15, (1972) 1048-49.

[2] Boyle, P. "Options: A Monte Carlo Approach," *Journal of Financial Economics,* 4 (1977), 323-338.

[3] Boyle, P., M. Broadie and P. Glasserman "Monte Carlo Methods for Security Pricing" forthcoming, *Journal of Economics Dynamics and Control* (1996)

[4] Braaten E. and G. Weller "An improved low discrepancy sequence for multidimensional quasi-Monte Carlo integration," *Journal of Computational Physics,* 33 (1979) 249-258.

[5] Bratley, P., B.L. Fox and H. Niederreiter "Implementation and Tests of Low-discrepancy sequences", *ACM Transactions on Modeling and Computer Simulation,* 2 (1992) 195-213.

[6] Caflisch, R.E. and W.J. Morokoff "Valuation of Mortgage-Backed Securities Using the Quasi-Monte Carlo Method", International Association of Financial Engineers First Annual Computational Finance Conference, Stanford University, (1996).

[7] Cranley, R. and T.N.L. Patterson "Randomization of Number Theoretic Methods for Multiple Integration," *SIAM Journal of Numerical Analysis,* 23 (1976), 904-14.

[8] Faure, H. "Good Permutation for Extreme Discrepancy," *Journal of Number Theory,* 42, (1992) 47-56.

[9] Faure, H. "Discrépance de suites associées à un système de numération (en dimension s)," *Acta Arithmetica,* XLI (1982), 337-351.

[10] Fox, P. "ALGORITHM 647: Implementation and Relative Efficiency of Quasi-random Sequence Generators," *ACM Trans. Math Software,* 12 (1986), 362-376.

[11] Halton, J.H. "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numer. Math.* 2, (1960) 84-90.

[12] Hickernell, F.J. "The Mean Square Discrepancy of Randomized Nets" Technical Report MATH-112, Department of Mathematics, Hong Kong Baptist University, 1996.

[13] Joe S. "Randomization of Lattice Rules for Numerical Multiple Integration", *Journal of Computational and Applied Mathematics*, 31 (1990) 299-304.

[14] Joy C., P. Boyle and K.S. Tan "Quasi-Monte Carlo Methods in Numerical Finance," *Management Science*, 42, No. 6, (1996) 926-38.

[15] Moro, B. "The Full Monte", *Risk* 8 Feb 1995 57-8.

[16] Morokoff, W.J. and R.E. Caflisch "Quasi-Monte Carlo Integration", *Journal of Computational Physics*, 122 (1995) 128-30.

[17] Morokoff, W.J. and R.E. Caflisch "Quasi-Random Sequences and their Discrepancies," *SIAM Journal Scientific Computing*, 15, (1994) 1251-79.

[18] Ninomiya, S. and S. Tezuka "Toward Real-Time Pricing of Complex Financial Derivatives," *Applied Mathematical Finance* , 3, (1996) 1-20.

[19] Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods.* Philadelphia, PA: SIAM, (1992).

[20] Niederreiter, H. "Point sets and sequences with small discrepancy," *Monatsh. Math.*, 104 (1987), 273-337.

[21] Owen, A.B. "Randomly permuted (t, m, s)-Nets and (t, s)-sequences" in *"Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing"*, H. Niederreiter and J. Shiue Editors, Springer, New York (1995).

[22] Owen, A.B. and D.A. Tavella "Scrambled Nets for Value at Risk Calculations", International Association of Financial Engineers First Annual Computational Finance Conference, Stanford University, (1996)

[23] Paskov, S.H. and J.F. Traub "Faster Valuation of Financial Derivatives", *Journal of Portfolio Management*, (1995) 113-120.

[24] Press, W., Teukolsky, S. et al. *Numerical Recipes in C, 2nd ed.* New York, NY: Cambridge University Press, 1992, 309-315.

[25] Scheuer E.M. and D.S. Stoller "On the generation of normal normal vectors" *Technometrics,* 4 (1962) 434-58.

[26] Sobol', I.M. "The Distribution of Points in a Cube and the Approximate Evaluation of Integrals," *U.S.S.R. Computational Math. and Math. Phys.,* 7, no. 4, (1967) 86-112.