On the Use of Quasi-Monte Carlo Methods in Computational Finance

Christiane Lemieux\(^1\) and Pierre L’Ecuyer\(^2\)

\(^1\) Department of Mathematics and Statistics, University of Calgary, 2500 University Drive N.W., Calgary, AB, T2N 1N4, Canada
leminieux@math.ucalgary.ca

\(^2\) Département IRO, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, H3C 3J7, Canada
lecuyer@iro.umontreal.ca

Abstract. We give the background and required tools for applying quasi-Monte Carlo methods efficiently to problems in computational finance, and survey recent developments in this field. We describe methods for pricing European path-dependent options, and also discuss problems involving the estimation of gradients and the simulation of stochastic volatility models.

1 Introduction

The Monte Carlo (MC) method has been introduced in finance in 1977, in the pioneering work of Boyle [5]. In 1995, Paskov and Traub published a paper [42] in which they used quasi-Monte Carlo (QMC) methods to estimate the price of a collateralized mortgage obligation. The problem they considered was in high dimensions (360) but nevertheless, they obtained more accurate approximations with QMC methods than with the standard MC method. Since then, many people have been looking at QMC methods has a promising alternative for pricing financial products [20, 37, 53, 1, 10, 3, 49]. Researchers studying QMC methods have also been very interested by these advances in computational finance because they provided convincing numerical results suggesting that QMC methods could do better than MC even in high dimensions, a task that was generally believed to be out of reach.

The aim of this paper is to provide the required background and tools for applying QMC methods to computational finance problems. We first review the idea of QMC methods and recall general results about their performance in comparison with the MC method. We give pseudocode for implementing Korobov rules [22], which constitute one type of QMC method, and provide references to papers and websites where other constructions (and code) can be found. Different randomizations are also discussed. In Section 3, we describe how randomized QMC methods can be applied for pricing European path-dependent options under the Black-Scholes model. Various methods that can be used in combination with QMC methods to enhance their performance are discussed in Section 4.

We conclude in Section 5 by discussing more complex applications such as the simulation of stochastic volatility models.
\section{Quasi-Monte Carlo Methods}

The general problem for which QMC methods have been proposed as an alternative to the MC method is multidimensional numerical integration. Hence for the remainder of this section, we assume the problem under consideration is to evaluate

\[ \mu = \int_{[0,1]^t} f(u)du, \]

where \( f \) is a square-integrable function. Many problems in finance amount to evaluate such integrals, as we discuss in Section 3. To approximate \( \mu \), both MC and QMC proceed by choosing a point set \( P_n = \{u_0, \ldots, u_{n-1}\} \subset [0,1)^t \), and then the average value of \( f \) over \( P_n \) is computed, i.e., we get

\[ Q_n = \frac{1}{n} \sum_{i=0}^{n-1} f(u_i). \]  

(1)

In the MC method, the points \( u_0, \ldots, u_{n-1} \) are independent and uniformly distributed over \([0,1)^t\). In practice, one uses a pseudorandom number generator to choose these points. The idea of QMC methods is to use a more regularly distributed point set, so that a better sampling of the function can be achieved. An important difference with MC is that the set \( P_n \) is typically deterministic when a QMC method is applied. Niederreiter presents these methods in detail in his book [35], and describes different ways of measuring the quality of the point sets \( P_n \) on which QMC methods rely. More specifically, the goal is to measure how far is the empirical distribution induced by \( P_n \) from the uniform distribution over \([0,1)^t\).

Such measures can be useful for providing upper bounds on the deterministic integration error \( |Q_n - \mu| \). For example, the rectangular-star discrepancy \( D^*(P_n) \) looks at the difference (in absolute value) between the volume of a rectangular “box” aligned with the axes of \([0,1)^t\) and having a corner at the origin, and the number of points from \( P_n \) contained in the box, and then take the maximum difference over all such boxes. Typically, a point set \( P_n \) is called a low-discrepancy point set if \( D^*(P_n) = O(n^{-1}\log^t n) \). For a function of bounded variation in the sense of Hardy and Krause, the integration error \( |Q_n - \mu| \) is in \( O(n^{-1}\log^t n) \) when \( P_n \) is a low-discrepancy point set (see [35, 30, 31] and the references therein for the details).

This type of upper bound suggests that the advantage of QMC methods over MC, which has a probabilistic error in \( O(n^{-1/2}) \), will eventually be lost as the dimension \( t \) increases, or more precisely, it suggests that QMC methods will require a sample size \( n \) too large, for practical purposes, to improve upon MC when \( t \) is large. In this context, numerical results showing an improvement of QMC over MC in high dimensions and using a relatively small sample size \( n \) [42, 10, 32, 26] seem hard to explain. To reconcile this apparent contradiction, two main approaches have been used. First, the study of \textit{randomized} QMC methods [12, 38, 51, 52, 41, 26] has provided new tools to understand the advantages of QMC over MC. Second, the notion of \textit{effective dimension}, introduced by Paskov
[43] and redefined in [10, 18], has been very useful to understand how QMC methods could improve upon MC even in large dimensions, as we now explain.

2.1 Effective Dimension

The effective dimension of a function is linked to its ANOVA decomposition [19, 14, 41], which rewrites any square-integrable function \( f : [0, 1]^t \rightarrow \mathbb{R} \) as a sum of \( 2^t \) components; there is one component \( f_I \) per subset \( I \) of \( \{1, \ldots, t\} \), i.e.,

\[
 f(u) = \sum_{I \subseteq \{1, \ldots, t\}} f_I(u),
\]

and the \( f_I \)'s are such that \( \int_{[0,1]^t} f_I(u) du = 0 \) for any nonempty \( I \), and \( \int_{[0,1]^t} f_I(u) du = 0 \) for any \( I \neq J \). Hence this decomposition is orthogonal and we get

\[
 \sigma^2 = \text{Var}(f) = \sum_{I \subseteq \{1, \ldots, t\}} \sigma_I^2,
\]

where \( \sigma_I^2 = \text{Var}(f_I) \). Therefore, the quantity \( \sigma_I^2 / \sigma^2 \) can be used as a measure of the relative importance of the component \( f_I \) for explaining the variance of \( f \). If the \( l \)-dimensional components with \( l \leq s \) contribute to more than 100\% of the variance (i.e., if \( \sum_{|I| = s} \sigma_I^2 \geq \alpha \sigma^2 \)), then \( f \) is said to have an effective dimension of at most \( s \) in the superposition sense [10, 18] in proportion \( \alpha \). Similar definitions can be given for the effective dimension in the truncation sense (if \( \sum_{|I| \leq s} \sigma_I^2 \geq \alpha \sigma^2 \)) [10, 18], or in the successive dimensions sense (if \( \sum_{|I| \leq s} \sigma_I^2 \geq \alpha \sigma^2 \)) [26].

It is often the case in computational finance that the functions to be integrated have a low effective dimension in some sense. When this happens, it means that even if the function is \( t \)-dimensional with \( t \) large, a QMC method based on a point set \( P_n \) that has good low-dimensional projections (i.e., such that when \( |I| \) is small, the projection \( P_n(I) \) of \( P_n \) over the subspace of \( [0,1]^t \) indexed by the coordinates in \( I \) is well distributed) can provide an accurate approximation for \( \mu \). Hence the success of QMC methods rely on a combination of “tractable” problems (i.e., problems involving functions with a low effective dimension), and point sets having good low-dimensional projections.

Note that in the study of the effective dimension, the variability of \( f \) is measured by its variance rather than, e.g., the bounded variation used in the upper bounds discussed earlier. In this context, it seems natural to measure the quality of an estimator for \( \mu \) by also looking at its variance. This can be achieved for QMC methods if we randomize their associated point set. By doing so, the integration error can be estimated easily. Also, the analysis of the variance has the advantage of requiring much weaker conditions on the function \( f \) than when the deterministic error is studied.
2.2 Constructions

Two main families of QMC methods are the *lattice rules* and the *digital nets* [35, 47]. Korobov rules are a special case of lattice rules that are easy to implement, as we now describe. For a given sample size \( n \), the only parameter required to generate a point set \( P_n \) in \( t \) dimensions is an integer \( a \) relatively prime to \( n \). We then get

\[
P_n = \left\{ \frac{i}{n}(1, a, a^2, \ldots, a^{t-1}) \mod 1, \ i = 0, \ldots, n - 1 \right\}, \tag{2}
\]

where the modulo 1 is applied component-wise. The choice of the generator \( a \) is important and tables of values of \( a \) leading to point sets \( P_n \) that are “good” for many values of \( t \) are given in [26]. What do we mean by good? The criterion used to measure the quality of \( P_n \) in [26] looks at many low-dimensional projections of the point set \( P_n \) and makes sure they are well distributed (with respect to the *spectral test*), in agreement with the requirements mentioned in the previous subsection. The points in \( P_n \) can be generated very easily as follows:

<table>
<thead>
<tr>
<th>Input: ( a, n, t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_0 = 1 )</td>
</tr>
<tr>
<td>for ( j = 1 ) to ( t - 1 ) do ( g_j = (g_{j-1} \times a) \mod n )</td>
</tr>
<tr>
<td>( u = (0, \ldots, 0) )</td>
</tr>
<tr>
<td>for ( i = 1 ) to ( n - 1 ) do ( u = (u + (g_0/n, \ldots, g_{t-1}/n)) \mod 1 )</td>
</tr>
</tbody>
</table>

The generation of the points in (2) can be done in an even simpler and faster way than that illustrated above when \( n \) is prime and \( a \) is a primitive element modulo \( n \); see [26] for more details. In any case, generating the point set \( P_n \) is faster than when MC is used, and this holds for most QMC methods.

Two nice properties of the point set (2) are that it is *dimension-stationary* and *fully projection-regular* [26, 47]. The first property means that if two subsets \( I = \{i_1, \ldots, i_s\}, J = \{j_1, \ldots, j_s\} \) of equal cardinality are such that \( j_l - i_l \) is constant for \( l = 1, \ldots, s \), then \( P_n(I) = P_n(J) \), i.e., the projection of \( P_n \) over the subspaces of \([0, 1]^t\) indexed by the the coordinates in \( I \) and \( J \) is the same. For example, it means that all the two-dimensional projections of the form \( P_n(\{j, j+1\}) \), for \( j = 1, \ldots, t - 1 \), are the same. Not all QMC methods have this property. The second property simply means that all projections of \( P_n \) have \( n \) distinct points, which is certainly desirable.

Another construction that shares many similarities with lattice rules are the *polynomial lattice rules* [25, 29]. As explained in [33, 34, 26], special cases of both methods can be constructed by using all overlapping \( t \)-tuples output by a *linear congruential generator* and a *Tausworthe generator*, respectively, from all possible initial seeds.

As for digital nets, details on their construction can be found in [35] and the references therein. Improved constructions are presented in, e.g., [50, 45, 36, 46, 11, 44]. Details on the implementation of Sobol’s sequence [48], and Faure’s
sequence [15], which were the first constructions proposed in the family of digital nets, are given in [7] and [16], respectively. The code that goes with these two papers can be found at www.acm.org/calgo/. More recent software for these methods and other ones can be found at www.mathdirect.com/products/qrn/ and www.cs.columbia.edu/~ap/html/finder.html, which is the link to the FinDer software [42]. The MC and QMC methods’ website www.mcqmc.org contains other relevant links.

2.3 Randomizations

As mentioned earlier, it is often useful for the purpose of error estimation to randomize QMC point sets. Two desirable properties that a given randomization should have are: (1) each point in the randomized point set should have a uniform distribution on 

\[ [0, 1]^2 \]; (2) the regularity of the point set should be preserved. The three randomizations discussed below have these properties.

For lattice rules, Cranley and Patterson [12] suggested to randomly generate a vector \( \Delta \) in \([0, 1)^t\), and then add it to each point of \( P_n \), modulo 1. This means that in the pseudocode given above, before the loop over \( i \), one simply needs to call a pseudorandom generator \( t \) times to generate the vector \( \Delta = (\Delta_1, \ldots, \Delta_t) \), and then output \( (u + \Delta) \mod 1 \) instead of \( u \). The variance of the estimator

\[
\frac{1}{n} \sum_{i=0}^{n-1} f((u_i + \Delta) \mod 1)
\]

based on a randomly shifted lattice rule is studied in [26, 52]; in [26], the only condition required on \( f \) is that it must be square-integrable. This randomization can be applied to other types of QMC point sets, as suggested in [51]. However, for digital nets and polynomial lattice rules, using a “XOR-shift” as proposed by Raymond Couture [25, 29] is more natural because it preserves the equidistribution properties of this type of point sets. The idea is to generate a random vector \( \Delta \) in \([0, 1)^t\), but instead of adding it to each point \( u_i = (u_{i1}, \ldots, u_{it}) \) of \( P_n \) modulo 1, an exclusive-or operation between the binary representation of \( \Delta_j \) and \( u_{ij} \) is performed, for each dimension \( j = 1, \ldots, t \). The variance of the estimator based on a polynomial lattice rule that has been XOR-shifted is studied in [29]. Another randomization that can be used for those point sets is the scrambling of Owen [38], which leads to tighter bounds on the variance of the associated estimators [38–40], but it requires more computation time than the XOR-shift.

3 Pricing under the Black-Scholes Model

In this section, we describe how to use QMC methods for estimating the value of a financial contract, such as an option, whose underlying assets follow the Black-Scholes model [4]. More precisely, we assume the goal is to estimate

\[
\mu = E^*(g_p(S_1, \ldots, S_t)),
\]
where $S_1, \ldots, S_t$ are prices (e.g., from one asset at $t$ different times, or from $t$
assets at the expiration date of the contract) that have a lognormal distribution. The function $g_p$ is assumed to be square-integrable and it represents the discounted payoff of the contract, and $p$ is a vector of parameters (e.g., containing the risk-free rate $r$, the volatility $\sigma$, the strike price $K$, the expiration time $T$, etc.). The expectation is taken under the risk-neutral measure [13]. Written similarly as in (1), the MC estimator for $\mu$ is given by

$$
\frac{1}{n} \sum_{i=0}^{n-1} f_p(u_i),
$$

where the $u_i$’s are independent and uniformly distributed over $[0, 1]^t$, and $f_p : [0, 1]^t \rightarrow \mathbb{R}$ is a function that takes as an input a sequence of $t$ numbers $u_1, \ldots, u_t$ between 0 and 1, transforms them into observations of $S_1, \ldots, S_t$, and then evaluates $g_p(S_1, \ldots, S_t)$. Also, $f_p$ is such that $E(f_p(u_i)) = \int_{[0,1]} f_p(u) \, du = \mu$. For example, if $f$ represents the discounted payoff of a path-dependent option on one asset, then $S_1, \ldots, S_t$ would represent observed prices on one path of this asset. To generate these prices, start with $u_1$, transform it into an observation $x$ from the standard normal distribution (using inversion, see, e.g., [23]), and then generate the first price by letting $S_1 = S_0 e^{r t_1 + \sigma \sqrt{t_1} x}$, where $t_1$ is the time at which $S_1$ is observed, and $S_0$ is the price of the underlying asset at time 0. In a similar way, $u_2$ can be used to generate the second price $S_2$, and so on. The precise definition of $f_p$ for an Asian option pricing problem is given in [26]. In the case where one has to generate prices of correlated assets, procedures requiring one uniform number per observed price can be found in, e.g., [1].

The QMC estimator for $\mu$ can be built in the exact same way as for MC if we use a randomized QMC point set: just take the estimator (3) but with the $u_i$’s coming from a randomized QMC point set. With an appropriate randomization, each point $u_i$ has a uniform distribution over $[0, 1]^t$ and thus the observation $f_p(u_i)$ has the same distribution as in the MC setting. Hence (3) is an unbiased estimator of $\mu$ in both cases. The only difference with MC is that with QMC, the observations $f_p(u_i), \ldots, f_p(u_{n-1})$ are correlated instead of being independent. With a carefully chosen QMC point set, the induced correlation should be such that the estimator has a smaller variance than the MC estimator. The variance of the randomized QMC estimator can be estimated by constructing $M$ i.i.d. copies of the estimator (3) (e.g., with $M$ i.i.d. random shifts), and then computing the sample variance.

4 Reducing the Variance and/or the Dimension

To increase the efficiency of MC simulations, many variance reduction techniques are available [24, 23] and can be applied for financial simulations; see the survey [6] for an overview. A good example of such technique is for pricing an Asian option on the arithmetic average; one can then use as a control variable the price of the same option but taken on the geometric average [21]. This can reduce the
variance by very large factors. These techniques can usually be combined with (randomized) QMC methods in a straightforward way; see, e.g., [10, 53, 26]. The combination almost always improves the naïve QMC estimators, but usually the advantage of QMC over MC is decreased by applying variance reduction techniques. Intuitively, this can be explained by the fact that these techniques might reduce the variance of the function in a way that concentrates the remaining variance in very small regions, and this makes it harder for QMC to improve upon MC.

Techniques for reducing the effective dimension are useless to improve MC simulations, but they can greatly enhance the performance of QMC methods [9, 32, 1, 10, 27, 2]. We discuss two of them: the Brownian bridge (BB) and the principal components (PC) techniques. The idea of BB was first introduced in [9] and can be used to generate a Brownian motion at \( T \) different times \( B(t_1), \ldots, B(t_T) \) by using \( T \) uniform numbers \( u_1, \ldots, u_T \). Instead of generating these observations sequentially (as outlined in the previous section), \( u_1 \) is used to generate \( B(t_1) \), \( u_2 \) is used to generate \( B(t_{1T/2}) \), \( u_3 \) for \( B(t_{T/4}) \), \( u_4 \) for \( B(t_{3T/4}) \), etc. This can be done easily since for \( u < v < w \), the distribution of \( B(v) \) given \( B(u) \) and \( B(w) \) is Gaussian with parameters depending only on \( u, v, w \). The reason why this can be helpful for QMC methods is that by generating the Brownian motion path in this way, more importance is given to the first few uniform numbers, and thus the effective dimension of a function depending on \( B(t_1), \ldots, B(t_T) \) should be decreased by doing that.

In the same spirit as BB, one can decompose the variance-covariance matrix of the prices to be simulated using principal components, and then generate the prices using this decomposition [1]. An advantage of this approach over BB is that it can be used to generate prices of correlated assets whereas BB can only be applied for generating prices coming from a single path. However, PC requires more computation time than BB for the generation of the prices, but see [2] for a way of speeding up PC.

5 Broadening the Range of Applications

We conclude this paper by discussing applications that go beyond the context discussed in Section 3. One such application is the estimation of derivatives (or gradients) of prices with respect to one (or more) parameter(s) (which are often called the greeks). For example, one might be interested in estimating how sensitive an option’s price is to the volatility of the underlying asset. Broadie and Glasserman [8] discuss how to do this using MC combined with some variance reduction techniques; a QMC approach based on randomly shifted lattice rules that improves upon MC is presented in [28]. These estimators could also be used for the more complex problem of American option pricing since the latter can be addressed using stochastic approximation methods that require gradient estimators, as discussed in [17].

QMC estimators can also be used for pricing contracts that depend on assets whose volatility is assumed to be stochastic. The difference with the problems
discussed in Section 3 is that one needs to discretize the price process and at each time step in the discretization, an observation from the volatility process must be generated in addition to one from the asset’s price. Hence for $T$ time steps, at least $2T$ random numbers are required to generate one path, which means the dimension of the problem is also at least $2T$. When using QMC, such simulations require a careful assignment of the uniform numbers $u_1, \ldots, u_{2T}$ to the generation of the prices and the volatilities [3], but improvement upon MC can still be achieved in this context [53, 3].

These applications are but a small sample of the possible problems for which QMC can provide more precise estimators than MC in computational finance. We believe that basically any problem that can be addressed using MC also has a QMC counterpart that can not only reduce the variance of the estimators, but that also typically requires less computation time.

References


