A Practical View of Randomized Quasi-Monte Carlo

[Invited Presentation, Extended Abstract]

Pierre L'Ecuyer
DIRO, Université de Montréal
C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7, Canada
lecuyer@iro.umontreal.ca

ABSTRACT

In this talk we will summarize the main ideas and results on randomized quasi-Monte Carlo (RQMC) methods, discuss their practical aspects, and give several examples. RQMC methods provide unbiased estimators of a mathematical expectation whose variance sometimes converge at a faster rate than with standard Monte Carlo, as a function of the number of simulation runs. We will also discuss an RQMC variant specially designed for the simulation of Markov chains.

Categories and Subject Descriptors

I.6 [**Computing Methodologies**]: Simulation and Modeling

General Terms

Algorithms, Performance

Monte Carlo

We are interested in estimating the mathematical expectation μ of a complicated random variable X by simulation. This is a central problem for performance evaluation and optimization in a large variety of areas. Simulation methods typically obtain a realization of X by generating a sequence of independent uniform random variables over the interval $(0,1)$, say $U_1,\ldots,U_s \sim U(0,1)$, and transforming them into X by some complicated algorithm, say $X = f(U_1,\ldots,U_s)$ where $f:(0,1)^s \to \mathbb{R}$. That is, we can write

$$\mu = \mathbb{E}[X] = \mathbb{E}[f(U_1,\ldots,U_s)] = \int_{(0,1)^s} f(\mathbf{u})\,d\mathbf{u}. \qquad (1)$$

For example, f may represent transformations of the uniforms into exponential or normal random variables by inversion, followed by further transformations to eventually compute a performance measure X. If the number of required uniforms is random and unbounded, we can take s as infinite.

In the *Monte Carlo* (MC) method, this is repeated n times independently and the estimator of μ is

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{U}_i), \qquad (2)$$

where $\mathbf{U}_0,\ldots,\mathbf{U}_{n-1}$ are independent random vectors uniformly distributed over $(0,1)^s$. We have $\mu = \mathbb{E}[\hat{\mu}_n]$ and the variance is $\mathrm{Var}[\hat{\mu}_n] = \sigma^2/n$, where $\sigma^2 = \int_{(0,1)^s} f^2(\mathbf{u})\,d\mathbf{u} - \mu^2$. If $\sigma^2 < \infty$, then $\hat{\mu}_n$ obeys a central-limit theorem, which can be exploited to compute a confidence interval on μ, whose size converges as $O(n^{-1/2})$.

Quasi-Monte Carlo

The idea of *quasi-Monte Carlo* (QMC), in its plain vanilla flavor [6,25,30], is to replace the independent random points $\mathbf{U}_i$ by a set $P_n = \{\mathbf{u}_0,\ldots,\mathbf{u}_{n-1}\}$ of n carefully-selected deterministic points that cover the unit hypercube much more evenly than typical random points. The MC estimator is then replaced by the (deterministic) approximation

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{u}_i). \qquad (3)$$

Often, an infinite sequence of points $\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2,\ldots$ is constructed and one can take the first n points for a value of n deemed convenient. Construction methods for these point sets and sequences include lattice rules, digital nets (or which the sequences of Sobol', Faure, Niederreiter, and Niederreiter-Xing are special cases), and variants of the Halton sequence, for example [20, 25, 30].

As a simple illustration, a widely-used type of QMC method is a *rank-1 lattice rule*, defined as follows, for an arbitrary integer n and dimension s [17, 25, 30]. Select a vector $\mathbf{a}_1 = (a_1,\ldots,a_s)$ whose coordinates a_j all belong to $\{0,\ldots,n-1\}$. Usually we also take them relatively prime with n. Then let $\mathbf{v}_1 = \mathbf{a}_1/n$ and define $P_n = \{\mathbf{v} = i\mathbf{v}_1 \bmod 1,\ i = 0,1,\ldots n-1\}$, where the division and the modulus are coordinate-wise. This point set is the intersection of a lattice with the unit hypercube in s dimensions. Figure 1 provides a simple illustration with $n = 101$ and $\mathbf{a}_1 = (1, 12)$. The 101 points $\mathbf{u}_i = (u_{i,1}, u_{i,2})$ cover the unit square quite evenly.

The aim is to construct the points so that the absolute integration error $|\bar{\mu} - \mu|$ converges faster than the (probabilistic) MC rate of $O(n^{-1/2})$, at least for certain classes of integrands f. This error depends on both the point set P_n and the function f. To study its behavior, the usual approach is

VALUETOOLS 2009, October 20-22, Pisa, Italy

DOI 10.4108/ICST.VALUETOOLS2009.7914

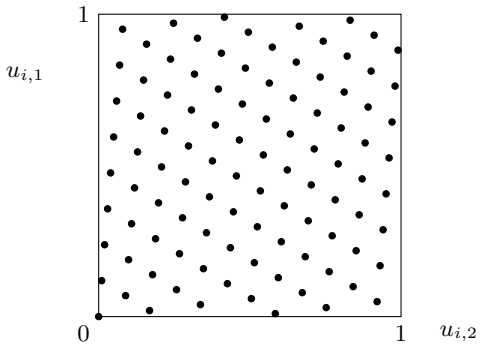

Figure 1: The point set P_n for a simple lattice rule in two dimensions with $n = 101$ points.

to consider a specific class $\mathcal{H}$ of functions f, often a reproducing kernel Hilbert space, and derive a worst-case error bound which is the product of two terms: one that depends only on P_n and measures its departure (or discrepancy) from the uniform distribution, and another that depends only on f and measures its variability. This can be written as

$$|\bar{\mu}_n - \mu| \le D(P_n)V(f) \tag{4}$$

for all $f \in \mathcal{H}$, where $V(f) = \|f - \mu\|_{\mathcal{H}}$ is the *variation* of f in $\mathcal{H}$, and $D(P_n)$ is the *discrepancy* of P_n [5,7,8]. Note that the definitions of $D(P_n)$ and $V(f)$ depend on each other. Generally speaking, a choice of definition that makes $V(f)$ smaller will make $D(P_n)$ larger, and vice-versa. Then, whenever $V(f) < \infty$, the error bound converges at the same rate as the discrepancy $D(P_n)$ as a function of n.

It turns out that for any real number $\alpha > 0$, there are Hilbert spaces $\mathcal{H}$ and corresponding variations $V(f)$ and discrepancies $D(P_n)$ for which it is possible to construct point sets P_n whose discrepancy $D(P_n)$ converges as $O(n^{-\alpha+\delta})$ for any $\delta > 0$. That is, we can beat MC by as much as we want if we are allowed to restrict ourselves to certain classes of functions. The price to pay is that for a larger α, more restrictive conditions must be made on f. For example, let α be a positive integer and assume that $\mathcal{H}$ is the (Sobolev) class of functions f that are periodic with respect to each coordinate (f takes the same value when this coordinate is 0 or 1 while the other coordinates are fixed), and whose partial derivative of order α with respect to any subset of coordinates is square integrable. The variation is defined as a weighted sum of these integrals. Then it is known that there exist lattice rules for which the corresponding discrepancy $D(P_n)$ converges as $O(n^{-\alpha+\delta})$ for any $\delta > 0$, and such rules are not difficult to construct by a computerized search. See [4, 25, 30, 31] for the details.

The best known special case of (4) is certainly the classical *Koksma-Hlawka inequality* [25], in which $V(f) = V_{\mathrm{hk}}(f)$ is the variation of f in the sense of Hardy and Krause and $D(P_n)$ is the *star discrepancy* $D^*(P_n)$, defined as follows: For each $\mathbf{u} \in (0,1)^s$, take the absolute difference between the volume of the box with corners at $\mathbf{0}$ and at $\mathbf{u}$, and the fraction of P_n that fall in that box, and define $D^*(P_n)$ be the supremum of this quantity over all $\mathbf{u} \in (0,1)^s$. Several explicit sequences $\mathbf{u}_0, \mathbf{u}_1, \ldots$, in any dimension s, are known for which $D^*(P_n) = O(n^{-1}(\ln n)^s)$ where $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$.

These error bounds are powerful in the asymptotic sense (for $n \to \infty$), but they are unfortunately unpractical in the sense that as soon as the dimension exceeds a few units, it takes an excessively large n_0 to have $n^{-1}(\ln n)^s < n^{-1/2}$ for all $n \ge n_0$ (i.e., to beat the MC error estimate). Moreover, $V_{\mathrm{hk}}(f)$ and $D^*(P_n)$ are typically much too hard to compute. Despite this, for several classes of integrands, sometimes in hundreds of dimensions, QMC approximations turn out to give much more accurate estimates than MC (empirically) [2, 13, 29]. Some explanation of this success will be given below.

Randomized Quasi-Monte Carlo

One practical difficulty with (deterministic) QMC is that reliable error estimates are hard to obtain. Another limitation is that the worst case error bound is often infinite, for example when f is unbounded (which happens frequently along the boundaries). *Randomized QMC* (RQMC) addresses these problems by randomizing P_n in a way that for its randomized version (called an RQMC point set):

(a) the entire point set retains its high uniformity;
(b) each individual point is a random vector having the uniform distribution over $(0,1)^s$.

The RQMC estimator is $\hat{\mu}_{n,\mathrm{rqmc}}$ as in (2), but $\mathbf{U}_0, \ldots, \mathbf{U}_{n-1}$ are now the n RQMC points. We have $\mathbb{E}[\hat{\mu}_{n,\mathrm{rqmc}}] = \mu$ (unbiasedness) and we hope that $\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] < \mathrm{Var}[\hat{\mu}_n]$. This can be proved in some settings. In fact, the worst-case variance over the functions $f \in \mathcal{H}$ with a given variability, say $V(f) \le 1$, does not exceed the mean square worst-case error:

$$\begin{aligned}\sup_{V(f)\le 1} \mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] &= \sup_{V(f)\le 1} \mathbb{E}[(\hat{\mu}_{n,\mathrm{rqmc}} - \mu)^2] \\ &\le \mathbb{E}\left[\sup_{V(f)\le 1} (\hat{\mu}_{n,\mathrm{rqmc}} - \mu)^2\right]\end{aligned}$$

where the expectation is with respect to the randomization. The latter often has the same convergence rate as the squared worst-case error with the non-randomized points. In some cases the variance is also strictly smaller than the mean square worst-case error. For example, for the space $\mathcal{H}$ of functions with square integrable mixed partial derivatives up to order α, it is known that there exist RQMC rules with $O(n^{-2\alpha-1+\delta})$ variance (although we do not know how to construct these rules explicitly except for $\alpha = 1$ where scrambled nets are known to achieve this rate [26,27]), whereas only $O(n^{-2\alpha+\delta})$ can be achieved for the worst-case error [9].

One very simple example of RQMC point set is as follows. Take an arbitrary QMC set P_n and add a *random shift modulo 1* to all the points simultaneously. That is, generate a single point $\mathbf{U}$ uniformly over $(0,1)^s$ and add it to each point of P_n, modulo 1, coordinate-wise [3,17,30]. If P_n is a lattice, then the lattice structure is preserved. Another example is a *random digital shift in base b* [16, 22]: generate again $\mathbf{U}$ uniformly over $(0,1)^s$, expand its coordinates in base b, and add the digits to the corresponding digits of each point of

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7914
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7914

P_n, modulo b. In the case where P_n is a digital net in base b, this preserves the digital net structure.

To estimate the variance and compute a confidence interval on μ, one may generate m independent realizations of the RQMC estimator $\hat{\mu}_{n,\mathrm{rqmc}}$, say $X_1, \ldots, X_m$, based on m independent randomizations of P_n, and compute their sample mean $\bar{X}_m$ and their sample variance and $S^2_{x,m}$. One has $\mathbb{E}[\bar{X}_m] = \mu$ and $\mathbb{E}[S^2_{x,m}] = m\mathrm{Var}[\bar{X}_m]$ [17]. If we assume that $\bar{X}_m$ is approximately normally distributed, then we can compute a confidence interval on μ in a standard way. However, this normality assumption could be far from the truth in general, especially if m is small, because the distribution of $\hat{\mu}_{n,\mathrm{rqmc}}$ does not converge to the normal distribution when $n \to \infty$ [19]. As a simple illustration, if $s = 1$ (one dimension) and we use a randomly-shifted lattice rule, the distribution of $\hat{\mu}_{n,\mathrm{rqmc}}$ (properly standardized) converges to a uniform distribution [19].

ANOVA Decomposition

Devising an method that effectively integrates all reasonable 200-dimensional functions, say, is clearly hopeless, because for example such a function could take arbitrary values in each of the 2^{200} corners (or multivariate quadrants) of the unit hypercube $(0,1)^{200}$. We would need at least 2^{200} points to have at least one in each corner, and this is not practically feasible. So how can we explain that certain classes of high-dimensional functions are integrated quite accurately (much better than with MC) by RQMC methods?

The best explanation seems to be that these functions for which RQMC performs well turn out to be well approximated by a sum of low-dimensional functions, which in turn can be integrated accurately by RQMC methods. A function that is almost linear, for example, can be well approximated by a sum of one-dimensional functions, and RQMC methods can perform extremely well on these one-dimensional functions. Of course, this is an extreme case.

This idea can be formalized via the functional ANOVA decomposition, defined as follows [21,28,32]. Suppose the variance σ^2 is finite. Then f has a unique decomposition of the form

$$f(\mathbf{u}) = \mu + \sum_{\mathfrak{u} \subseteq \mathcal{S},\, \mathfrak{u} \neq \phi} f_{\mathfrak{u}}(\mathbf{u}) \tag{5}$$

where each $f_{\mathfrak{u}} : (0,1)^s \to \mathbb{R}$ depends only on $\{u_i,\, i \in \mathfrak{u}\}$, the $f_{\mathfrak{u}}$'s integrate to zero and are orthogonal, and the variance decomposes as $\sigma^2 = \sum_{\mathfrak{u} \subseteq \mathcal{S}} \sigma^2_{\mathfrak{u}}$ where $\sigma^2_{\mathfrak{u}} = \mathrm{Var}[f_{\mathfrak{u}}(\mathbf{U})]$ where $\mathbf{U}$ is uniformly distributed over $(0,1)^s$.

Let $P_n(\mathfrak{u})$ denotes the projection of P_n over the subspace determined by the subset of coordinates $\mathfrak{u}$. For a given function f, if there is a class $\mathcal{J}$ of small subsets of $\mathcal{S}$ for which

$$\sum_{\mathfrak{u} \in \mathcal{J}} \sigma^2_{\mathfrak{u}} \geq \rho \sigma^2 \tag{6}$$

for some ρ close to 1, then we can focus on constructing the point set P_n so that its projections $P_n(\mathfrak{u})$ are as uniform as possible for all $\mathfrak{u} \in \mathcal{J}$ (after the randomization), and neglect the other projections. If we can achieve that, then the variance terms on the left side of (6) can be reduced significantly, which would yield an RQMC method having much less variance than MC. When (6) holds for a class of low-cardinality subsets $\mathcal{J}$ and ρ close to 1, we say that f has *low effective dimension*. There are special cases of this definition, for specific ways of choosing $\mathcal{J}$ [2, 17, 28]. For example, $\mathcal{J}$ may contain only the subsets of the first d coordinates, for a small d, or only the subsets of cardinality d or less (e.g., with $d = 2$ or 3).

Low effective dimension can sometimes be achieved by transforming the function f via a change of variables that preserves the mean μ, for example in a way that the first few uniforms account for most of the variance [1,2,10,13,23,33]. That is, we change the way the uniforms are used by the simulation. For example, this can be achieved by using a bridge sampling technique or principal component decomposition to generate the sample path of a Brownian motion or a Lévy process, and sometimes by replacing some random variables by their conditional expectations [1,6,10,13,17,24].

Specific Methods, Examples, and More

In the talk, we will examine specific RQMC techniques, and provide examples and numerical illustrations of their performance. Some of these examples will be taken from [17] and [13].

For more on RQMC methods and their practical aspects, we refer the reader to [11, 13, 18, 20]. Software implementations and programming examples can be found in the Java library SSJ [12].

Array-RQMC for Markov Chains

Simulating a Markov chain usually requires a very long sequence of uniform random numbers. That is, we face a high-dimensional integration problem for which RQMC is typically ill-suited, and it is generally difficult to reduce the effective dimension to a small value. However, a special RQMC technique called *array-RQMC*, designed for that situation, has been developed recently [14, 15]. The algorithm simulates n copies of the chain in parallel, advances all copies by one step using an RQMC point set at each iteration, and induces negative dependence between these copies in a way that the empirical distribution of the n states at any given step provides a more accurate estimate of the exact distribution than if the n copies were simulated independently. As a result, it provides lower variance for the average of a state-dependent cost function at any step. In some examples, the variance can be reduced by factors of more than 1000, even for Markov chains that evolve over a few hundred steps.

Estimating Something else than a Mean

RQMC can be used as well for other purposes than estimating an expectation. For example, it can be effective for estimating a quantile, or a function of several expectations, or the gradient of an expectation with respect to a vector of parameters, or to obtain an approximation of a function f over a given domain, or to estimate the solution of an optimization problem in which the objective function or the constraints (or both) involve mathematical expectations. It can also replace MC in algorithms that combine MC with approximate dynamic programming, such as least-squares Monte Carlo [6].

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7914
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7914

Acknowledgment

This work has been supported by an NSERC-Canada Discovery Grant and a Canada Research Chair to the author.

1. REFERENCES

[1] T. Avramidis and P. L'Ecuyer. Efficient Monte Carlo and quasi-Monte Carlo option pricing under the variance-gamma model. *Management Science*, 52(12):1930–1944, 2006.

[2] R. E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997.

[3] R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976.

[4] J. Dick, I. H. Sloan, X. Wang, and H. Wozniakowski. Liberating the weights. *Journal of Complexity*, 20(5):593–623, 2004.

[5] J. Dick, I. H. Sloan, X. Wang, and H. Wozniakowski. Good lattice rules in weighted Korobov spaces with general weights. *Numerische Mathematik*, 103:63–97, 2006.

[6] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[7] F. J. Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67:299–322, 1998.

[8] F. J. Hickernell. What affects the accuracy of quasi-Monte Carlo quadrature? In H. Niederreiter and J. Spanier, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1998*, pages 16–55, Berlin, 2000. Springer-Verlag.

[9] F. J. Hickernell. My dream quadrature rule. *Journal of Complexity*, 19:420–427, 2003.

[10] J. Imai and K. S. Tan. A general dimension reduction technique for derivative pricing. *Journal of Computational Finance*, 10(2):129–155, 2006.

[11] A. Keller, S. Heinrich, and H. Niederreiter, editors. *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Springer-Verlag, Berlin, 2008.

[12] P. L'Ecuyer. *SSJ: A Java Library for Stochastic Simulation*, 2008. Software user's guide, available at `http://www.iro.umontreal.ca/~lecuyer`.

[13] P. L'Ecuyer. Quasi-Monte Carlo methods with applications in finance. *Finance and Stochastics*, 13(3):307–349, 2009.

[14] P. L'Ecuyer, C. Lécot, and A. L'Archevêque-Gaudet. On array-RQMC for Markov chains: Mapping alternatives and convergence rates. In P. L'Ecuyer and A. B. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, Berlin, 2009. Springer-Verlag. To appear.

[15] P. L'Ecuyer, C. Lécot, and B. Tuffin. A randomized quasi-Monte Carlo simulation method for Markov chains. *Operations Research*, 56(4):958–975, 2008.

[16] P. L'Ecuyer and C. Lemieux. Quasi-Monte Carlo via linear shift-register sequences. In *Proceedings of the 1999 Winter Simulation Conference*, pages 632–639. IEEE Press, 1999.

[17] P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235, 2000.

[18] P. L'Ecuyer and A. B. Owen, editors. *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer-Verlag, Berlin, 2010. To appear.

[19] P. L'Ecuyer and B. Tuffin. On the error distribution for randomly-shifted lattice rules. In *Proceedings of the 2009 Winter Simulation Conference*. IEEE Press, 2009. To appear.

[20] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer-Verlag, New York, NY, 2009.

[21] R. Liu and A. B. Owen. Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721, 2006.

[22] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide*. Springer-Verlag, Berlin, 1999.

[23] W. J. Morokoff. Generating quasi-random paths for stochastic processes. *SIAM Review*, 40(4):765–788, 1998.

[24] B. Moskowitz and R. E. Caflisch. Smoothness and dimension reduction in quasi-Monte Carlo methods. *Journal of Mathematical and Computer Modeling*, 23:37–54, 1996.

[25] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA, 1992.

[26] A. B. Owen. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997.

[27] A. B. Owen. Scrambled net variance for integrals of smooth functions. *Annals of Statistics*, 25(4):1541–1562, 1997.

[28] A. B. Owen. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions on Modeling and Computer Simulation*, 8(1):71–102, 1998.

[29] S. Paskov and J. Traub. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22:113–120, 1995.

[30] I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration*. Clarendon Press, Oxford, 1994.

[31] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals. *Journal of Complexity*, 14:1–33, 1998.

[32] X. Wang and I. H. Sloan. Why are high-dimensional finance problems often of low effective dimension? *SIAM Journal on Scientific Computing*, 27(1):159–183, 2005.

[33] X. Wang and I. H. Sloan. Brownian bridge and principal component analysis: Toward removing the curse of dimensionality. *IMA Journal of Numerical Analysis*, 27:631–654, 2007.

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7914
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7914