Comparison of Quasi-Monte Carlo-Based Methods for the Simulation of Markov Chains

Christian Lécot
Laboratoire de Mathématiques
Université de Savoie
73376 Le Bourget-du-Lac Cedex
France
E-mail: Christian.Lecot@univ-savoie.fr

Bruno Tuffin*
IRISA-INRIA
Campus universitaire de Beaulieu
35042 Rennes Cedex
France
E-mail: btuffin@irisa.fr

Abstract

Monte Carlo (MC) method is probably the most widespread simulation technique due to its ease of use. Quasi-Monte Carlo (QMC) methods have been designed in order to speed up the convergence rate of MC but their implementation requires more stringent assumptions. For instance, the direct QMC simulation of Markov chains is inefficient due to the correlation of the points used. We propose here to survey the QMC-based methods that have been developed to tackle the QMC simulation of Markov chains. Most of those methods were hybrid MC/QMC methods. We compare them with a recently developped pure QMC method and illustrate the better convergence speed of the latter.

Key words: Markov chains, Monte Carlo, Quasi-Monte Carlo, Simulation.

1 Introduction

Monte Carlo (MC) simulation technique [5] has been widely used from the first days of computer science (and even before), in all scientific fields, for computing integrals or solving differential equations for instance. It is based on the law of large numbers

*The work of this author has been partially supported by the Sure-Paths ACI

which states that by sampling and considering the mean over a sample of the set of possibilities we get a good approximation of the quantity of interest that almost surely converges when the sample size increases. Its simplicity of use and the very few required assumptions are probably the main reasons of its success. Especially when compared with traditional numerical analysis techniques, MC methods are argued to be insensitive to the problem dimensionality.

Quasi-Monte Carlo (QMC) [12] methods can be defined by analogy with MC, by replacing the random sample by a sequence of "well distributed" points (called a *low discrepancy sequence*) that is expected to produce a faster convergence to the true value. Whereas very promising theoretically, QMC methods suffer from several drawbacks. The two main problems are the following. First, except if applied non-directly and using specific and complicated techniques, the dimensionality of the problem has to be known (and relatively small in practice). Second, the error estimation, while possible in theory, is intractable in practice. For these reasons, QMC has not been as applied as MC.

The goal of this paper is to review the QMC-based methods that have been designed in order to circumvent the first of these two drawbacks. We survey how hybrid techniques help in extending the range of application of QMC, for estimating improper integrals and measures over stochastic processes such as Markov chains. The contribution of the paper is a comparison of the hybrid methods of the literature [14, 15, 18, 19] with a new *deterministic* QMC method [8] that can be seen as a deterministic scrambling version of the scrambled algorithm of [19].

The layout of the paper is as follows. In Section 2 we recall the basic definitions of MC and QMC and highlight the main drawbacks of QMC. In Section 3 we present how, with the help of pseudo-random number, the range of applications of QMC can be extended to the estimation of measures over stochastic processes. We compare in Section 4 the convergence rate of hybrid methods with a new deterministic technique that shows its superiority. Section 5 is devoted to the conclusions and the perspectives of research.

2 Monte Carlo and quasi-Monte Carlo methods

2.1 Monte Carlo

Assume that we wish to compute the integral

$$\mathcal{I} = \int_{[0,1)^s} f(x)dx.$$

Let $(X^{(n)})_{0\leq n<N}$ be a finite sequence of N random and independent vectors uniformly distributed over $[0,1)^s$. By the law of large numbers, we know that an unbiased estimator of $\mathcal{I}$ is

$$\bar{f}_N = \frac{1}{N}\sum_{n=0}^{N-1} f(X^{(n)}).$$

The variance of $\bar{f}_N$ is then σ^2/N; σ^2 being the variance of the random variable $f(X)$ where X is uniformly distributed over $[0,1)^s$. From the central limit theorem we know that the error

$$\sqrt{N}\frac{\bar{f}_N - \mathcal{I}}{\sigma}$$

converges to a Gaussian law with mean 0 and standard deviation 1. This allows us to compute a confidence interval for $\mathcal{I}$:

$$\mathcal{I} \in \left[\bar{f}_N - \frac{c_\alpha \sigma}{\sqrt{N}},\ \bar{f}_N + \frac{c_\alpha \sigma}{\sqrt{N}}\right]$$

at confidence level α, where $c_\alpha = \Phi^{-1}(\frac{1+\alpha}{2})$ and Φ is the distribution function of the Gaussian law with mean 0 and standard deviation 1. The convergence speed of this method is then, on average, $O(N^{-1/2})$, independent of the dimension of the problem s. In a practical implementation, the estimator $\mathcal{I}$ is computed by generating the uniformly distributed variables $X^{(i)}$ using pseudo-random numbers [5].

2.2 Quasi-Monte Carlo

In quasi-Monte Carlo methods [2, 12], the pseudo-random sequence is replaced by a deterministic equi-distributed one $\mathcal{P} = (\xi^{(n)})_n$ so that the estimator is

$$\frac{1}{N}\sum_{n=0}^{N-1} f(\xi^{(n)}). \tag{1}$$

A measure of equi-distribution is the following. Let B be a sub-interval of $[0,1)^s$ and $A_N(B,\mathcal{P})$ be the number of points in B among the N firsts of the sequence $\mathcal{P} = (\xi^{(n)})_{n\in\mathbb{N}}$, i.e.

$$A_N(B,\mathcal{P}) = \sum_{n=0}^{N-1} 1_B(\xi^{(n)}).$$

To measure the quality of the repartition, the *discrepancy* of the N first elements of $\mathcal{P}$ is defined by

$$D_N^*(\mathcal{P}) = \sup_{x\in[0,1]^s} \left| \frac{A_N(\prod_{i=1}^s [0,x_i),\mathcal{P})}{N} - \prod_{i=1}^s x_i \right|.$$

The sequence $\mathcal{P} = (\xi^{(n)})_{n\in\mathbb{N}}$ is then said to be equi-distributed if and only if $\lim_{N\to+\infty} D_N^*(\mathcal{P}) = 0$. Error bounds for the approximation (1) of $\mathcal{I}$ are obtained in terms of the discrepancy. Let P be a partition of $[0,1]^s$ in subintervals and $\Delta(f,J)$ be the alterned sum of f values at the edges of sub-interval J. The variation in sense of Vitali is defined by

$$V_{\mathrm{Vit}}(f) = \sup_P \sum_{J\in P} |\Delta(f,J)|.$$

From $V_{\rm Vit}(f)$, we define $V(f)$, the variation of f in sense of Hardy and Krause by

$$V(f) = \sum_{k=1}^{s} \sum_{1 \le i_1 < \cdots < i_k \le s} V_{\rm Vit}^{(k)}(f\ ; i_1, \ldots, i_k)$$

where $V_{\rm Vit}^{(k)}(f\ ; i_1, \ldots, i_k)$ is the variation in sense of Vitali applied to the restriction of f to the space of dimension k $\{(u_1, \ldots, u_s) \in [0,1]^s\ : u_j = 1 \text{ for } j \neq i_1, \ldots, i_k\}$. We then have the Koksma-Hlawka inequality:

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\xi^{(n)}) - \int_{[0,1)^s} f(x)dx \right| \le V(f) D_N^*(\mathcal{P}). \tag{2}$$

A sequence $\mathcal{P} = (\xi^{(n)})_{n \in \mathbb{N}}$ is said to be a *low discrepancy sequence* if $D_N^*(\mathcal{P}) = O(N^{-1}(\ln N)^s)$. It has been proved that, for a finite sequence, we cannot get better than $O(N^{-1}(\ln N)^{\alpha(s)})$ where $\alpha(s) = s-1$ for $s = 1, 2$ and $\alpha(s) = (s-1)/2$ otherwise [12]. There exist many low discrepancy sequences; we can quote for instance Halton sequences [6], Kronecker sequences [2], Sobol' sequences [1, 16, 17], Niederreiter sequences[10, 11], or Faure sequences [3, 4].

The Sobol, Niederreiter and Faure sequences are in the class of so-called (u, s)-sequences that we describe now in more details since they will be used later on. For an integer $b \ge 2$, an *elementary interval in base* b is an interval of the form

$$\prod_{r=1}^{s} \left[\frac{a_r}{b^{q_r}}, \frac{a_r + 1}{b^{q_r}} \right),$$

with integers $q_r \ge 0$ and $0 \le a_r < b^{q_r}$ for $1 \le r \le s$. If $0 \le u \le q$ are integers, a *(u, q, s)-net in base b* is a point set X consisting of b^q points in $[0,1)^s$ such that $D(Q; X) = 0$ for every elementary interval Q in base b with measure b^{u-q}. The analogous concept for an infinite sequence is defined as follows. If $b \ge 2$ and $u \ge 0$ are integers, a sequence $\xi^{(0)}, \xi^{(1)}, \ldots$ of points in $[0,1)^s$ is a *(u, s)-sequence in base b* if, for all integers $n \ge 0$ and $q > u$, the points $\xi^{(p)}$ with $nb^q \le p < (n+1)b^q$ form a (u, q, s)-net in base b.

2.3 A difficulty with quasi-Monte Carlo

One of the main advantages of Monte Carlo is that it can be easily applied to other types of problems such as the simulation of stochastic processes, and that the models under study require very few assumptions. On the other hand, quasi-Monte Carlo methods, due to the correlation structure of low discrepancy sequences, cannot be applied directly to such problems, by just replacing pseudo-random numbers by quasi-random ones, and require further theoretical analysis [7, 9, 13]. Indeed, consider for instance the simulation of a Markov chain defined over state space $\{0, 1\}$ with initial state 0 and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Using Van der Corput sequence (one-dimensional Halton sequence) in base $b = 2$, such that $\xi^{(n)} = \sum_{k\geq 0} a_k(n)2^{-(k+1)}$ when $n = \sum_{k\geq 0} a_k(n)2^k$ is the development of n in base 2, the simulation produces a path such that the process will never stay in the same state two consecutive instants t and $t+1$ (which is an undesirable effect). This is due to the fact that $\xi^{(n)} \geq 1/2$ for n odd and $\xi^{(n)} < 1/2$ if n is even.

In the next sections, we see how using hybrid QMC methods would circumvent this drawback.

3 Randomized quasi-Monte Carlo methods for the simulation of Markov chains

Assume that we wish to estimate a measure over a Markov chain. As described in Section 2, a direct application of QMC would not work. We review in 3.1 the randomization methods in the literature, and introduce a new deterministic scrambling technique in 3.2.

3.1 Hybrid quasi-Monte Carlo

Consider the simulation of a Markov chain $(X_t)_{t\in\mathbb{N}}$ with probability matrix P and initial distribution μ, up to a fixed time T defined over a state space E. Here E is finite ($E = \{1, \ldots, M\}$) or countable ($E = \mathbb{N}$ or $\mathbb{Z}$).

Theoretically, the Markov process can be simulated using QMC methods by considering a low discrepancy sequence $(\xi^{(n)})_{n\in\mathbb{N}}$ in dimension $T+1$ such that the point $\xi^{(n)}$ is used to sample the n-th path of the estimation:

- the initial state of the n-th realization is sampled from probability measure μ, using $\xi_1^{(n)}$ (like would be done using a pseudo-random number).
- the t-th step of the n-th realization, from $X_{t-1}^{(n)}$ to $X_t^{(n)}$, is sampled from probability measure $P_{x,\cdot}$ if $X_{t-1}^{(n)} = x$, using $\xi_{t+1}^{(n)}$.

Nevertheless, for quite large values of T, QMC is known to be inefficient. For this reason, the following hybrid algorithms have been developed.

3.1.1 Mixed strategy

The mixed strategy has been developed in [14, 18, 19]. It consists in using an s-dimensional low discrepancy sequence so that the s first steps of the Markov chains are sampled using the s coordinates of the low discrepancy sequence, and the $T+1-s$ remaining steps using pseudo-random numbers.

Formally, the t-th step of the n-th path is sampled using $\xi_{t+1}^{(n)}$ if $t < s$ and $U_t^{(n)}$ otherwise, where the $U_t^{(n)}$ ($s \leq t \leq T$) are independent and uniform random variables over $[0, 1)$.

It is expected that the good repartition of low discrepancy sequences will improve the convergence rate.

3.1.2 Latin Hypercube Sampling (LHS)

The method consists in using a Latin hypercube sample for "padding" the remaining coordinates instead of just random numbers in the above mixed strategy [15]. LHS is a form of simultaneous stratification on all the remaining dimensions. Here it is applied in the following way: the t-th step of the n-th path is (still) sampled using $\xi_{t+1}^{(n)}$ if $t < s$, and using

$$V_t^{(n)} = \frac{\pi_t(n) - U_t^{(n)}}{N}$$

if $s \leq t \leq T$, where π_t are independent uniform random permutations of the integers $0, \ldots, N-1$ and the $U_t^{(n)}$ $(s \leq t \leq T)$ are independent and uniform random variables over $[0, 1)$.

3.1.3 Scrambled strategy

This strategy scrambles the "direct" QMC method that was proved to be inefficient. Consider a one-dimensional low discrepancy sequence $(\xi^{(n)})_{n \in \mathbb{N}}$. For $0 \leq t \leq T$ let π_t be independent random permutations of integers $0, \ldots, N-1$. The N sampled paths are simulated in parallel.

- The initial state of the n-th realization $(0 \leq n < N)$ is sampled from probability measure μ, using $\xi^{(\pi_0(n))}$.
- The t-th step of the n-th realization is sampled from probability measure $P_{x,\cdot}$ if $X_{t-1}^{(n)} = x$, using $\xi^{(t-1)N+\pi_t(n)}$.

To reduce the complexity of the method, π_0 is often taken as the identity.

To the best of our knowledge, no proof of convergence has been provided for this strategy, but numerical results illustrate that it actually converges.

3.2 A deterministic scrambling

This method, developed in [8], can be seen as a deterministic version of the previous scrambled strategy. Consider a $(u, 2)$-sequence in base b $(\xi^{(n)})_{n \in \mathbb{N}}$ (then a sequence in dimension 2) and let $\Xi^{(n)}$ be the point set $\{\xi^{(p)} = (\xi_1^{(p)}, \xi_2^{(p)}) \ : \ nN \leq p < (n+1)N\}$ where $N = b^q$, with $q > u$. If proj_1 and proj_2 denote the projections defined by $\mathrm{proj}_i(x_1, x_2) = x_i$, for $i = 1, 2$, we assume that $\forall n \geq 0$

$$\mathrm{proj}_1 \Xi^{(n)} \text{ is a } (0, q, 1)\text{-net in base } b, \tag{3}$$

and that, if $E = \mathbb{Z}$,

$$0 \notin \mathrm{proj}_2 \Xi^{(n)}. \tag{4}$$

Consider N distinct chains $X^{(\ell)} = \{X_n^{(\ell)}\}_{0 \leq n \leq T}$, $0 \leq \ell < N$.

1. Using the point set $\Xi^{(0)}$, sample the initial state (i.e., at time $t = 0$) of each chain according to μ: the initial state $i_0^{\lfloor N\xi_1^{(\ell)} \rfloor}$ of the chain $\lfloor N\xi_1^{(\ell)} \rfloor$ is sampled using $\xi_2^{(\ell)}$ (as would be done using a pseudo-random number). Note that from (3), each chain will be sampled exactly once.
2. Re-order the chains according to their states: $i_n^0 \leq i_n^1 \leq \cdots \leq i_n^{N-1}$.
3. $n = n + 1$. Consider the point set $\Xi^{(n)}$. Sample the new state for each chain according to matrix P: the new state of chain $\lfloor N\xi_1^{(\ell+nN)} \rfloor$ is chosen using $\xi_2^{(\ell+nN)}$ from the distribution $P_{i_n^{\lfloor N\xi_1^{(\ell+nN)} \rfloor},\cdot}$.
4. if $n = T$ stop else goto 2).

It is proven in [8] that the method converges if the transition matrix P satisfies

$$\forall k \in E \quad \sum_{i \in E'} \Big| \sum_{j<k} p(i+1,j) - \sum_{j<k} p(i,j) \Big| \leq 1,$$

where

$$E' := \{ i \in E : i+1 \in E \}.$$

4 Numerical comparisons

To compare the efficiency of the above methods, consider a small example, consisting in a discrete-time queue such that a new customer arrives at each time t with probability 0.6 and there is no arrival with probability 0.4. Each customer in the queue is served at time t with probability 0.5. This is a so-called Geo/Geo queue. We wish to compute the mean number of customers at time $t = 10$ given that the queue is empty at time $t = 0$. For the LHS (LHS5) and mixed (M5) methods, we have arbitrarily considered a low discrepancy sequence in dimension $s = 5$, so that the last 5 transition steps are sampled using pseudo-random numbers. The results are displayed in Figure 1. Regression analysis yields the following convergence speeds:

$$\begin{aligned} \mathrm{Err}_{\mathrm{MC}} &= 0.18 N^{-0.40} \\ \mathrm{Err}_{\mathrm{M5}} &= 0.43 N^{-0.49} \\ \mathrm{Err}_{\mathrm{LHS5}} &= 0.20 N^{-0.50} \\ \mathrm{Err}_{\mathrm{Scr}} &= 0.25 N^{-0.53} \\ \mathrm{Err}_{\mathrm{QMC}} &= 0.66 N^{-0.95}. \end{aligned}$$

It illustrates that the deterministic scrambling outperforms the other methods.

Consider now a larger example, consisting in a (continuous-time) M/M/1/40 queue with arrival rate 0.97 and service rate 1.0 : we want to compute the loss probability at time $T = 150$. The model is discretized using the uniformization technique, resulting in a discrete-time process in dimension 411 (i.e., 411 transition

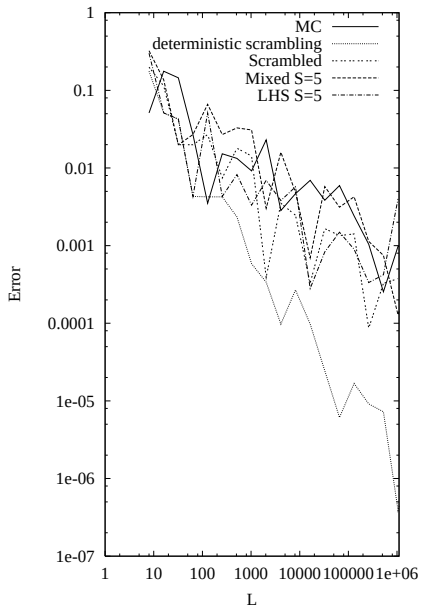

Figure 1: Discrete queue

steps, at a given pre-defined error of $\varepsilon = 10^{-10}$). For more details on the model and the uniformization technique, we refer to [8]. The results are displayed in Figure 2 (using a low discrepancy sequence in dimension $s = 30$ for the mixed and LHS methods). Regression analysis yields the following convergence speeds:

$$\begin{aligned} \mathrm{Err}_{\mathrm{MC}} &= 0.005 N^{-0.39} \\ \mathrm{Err}_{\mathrm{M30}} &= 0.062 N^{-0.64} \\ \mathrm{Err}_{\mathrm{Scr}} &= 0.075 N^{-0.68} \\ \mathrm{Err}_{\mathrm{QMC}} &= 0.012 N^{-0.84}. \end{aligned}$$

Again, the deterministic scrambling provides much better results than the other techniques.

5 Conclusions and Perspectives

We have shown that quasi-Monte Carlo methods, when applied directly, may be inefficient when simulating Markov chains. Hybrid methods have been designed in the literature to cope with this problem. We have compared those methods with a pure QMC method recently developped by the authors and have illustrated the superiority of this last one. Nevetheless, several issues remains to be investigated:

- Could this superiority be mathematically proved?

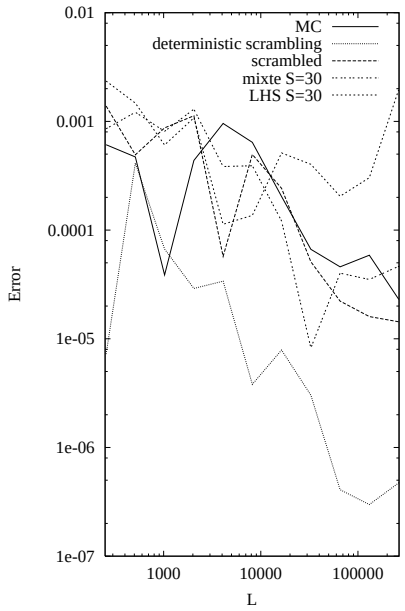

Figure 2: M/M/1/40 queue

- Could the pure QMC method be extended to the case of continuous time Markov chains, to regenerative simulation, or to multi-dimensional state spaces?

References

[1] I.A.. Antonov and V.M. Saleev. An economic method of computing LP_τ-sequences. *USSR Comput. Math. Math. Phys.*, 19:252–256, 1979.

[2] M. Drmota and R.F. Tichy. *Sequences, Discrepancies and Applications.* Springer-Verlag, Berlin, 1997.

[3] H. Faure. Discrépances de suites associées à un système de numération (en dimension un). *Bull. Soc. Math. Fr.*, 109:143–182, 1981.

[4] H. Faure. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arith.*, 41:337–351, 1982.

[5] G.S. Fishman. *Monte Carlo: Concepts, Algorithms and Applications.* Springer-Verlag, New York, 1996.

[6] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2:84–90, 1960; Berichtigung, ibid. 196.

[7] C. Lécot and A. Koudiraty. Quasi-random simulation of linear kinetic equations. *J. Complexity*, 17:795–814, 2001.

[8] C. Lécot and B. Tuffin. Quasi-Monte Carlo methods for estimating transient measures of discrete time Markov chains. In H. Niederreiter, ed., *Monte Carlo and Quasi-Monte Carlo Methods 2002*, to appear.

[9] W. J. Morokoff and R. E. Caflisch. A quasi-Monte Carlo approach to particle simulation of the heat equation. *SIAM J. Numer. Anal.*, 30:1558–1573, 1993.

[10] H. Niederreiter. Point sets and sequences with small discrepancy. *Monatsh. Math.*, 104:273–337, 1987.

[11] H. Niederreiter. Low discrepancy and low-dispersion sequences. *J. Number Theory*, 30:51–70, 1988.

[12] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.

[13] S. Ogawa and C. Lécot. A quasi-random walk method for one-dimensional reaction-diffusion equations. *Math. Comput. Simul.*, 62:487–494, 2003.

[14] G. Ökten. A probabilistic result on the discrepancy of a hybrid-monte carlo sequence and applications. *Monte Carlo Methods Appl.*, 2:255–270, 1996.

[15] A.B. Owen. Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo. In H. Niederreiter and J. Spanier, eds., *Monte Carlo and Quasi-Monte Carlo Methods 1998*, pages 86–97. Springer-Verlag, Berlin, 2000.

[16] I.M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.*, 7:86–112, 1967.

[17] I.M. Sobol'. Uniformly distributed sequences with an additional uniform property. *USSR Comput. Math. Math. Phys.*, 16:236–242, 1976.

[18] J. Spanier. Quasi-Monte Carlo methods for particle transport problems. In H. Niederreiter and P. J.-S. Shiue, eds., *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 121–148. Springer-Verlag, New York, 1995.

[19] J. Spanier and L. Li. Quasi-Monte Carlo methods for integral equations. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, eds., *Monte Carlo and Quasi-Monte Carlo Methods 1996*, pages 398–414. Springer-Verlag, New York, 1998.