
Quasi-Monte Carlo Simulation of Discrete-Time Markov Chains on Multidimensional State Spaces

Rami El Haddad¹, Christian Lécot², and Pierre L'Ecuyer³

¹ Département de Mathématiques, Université Saint-Joseph,
BP 11-514, Riad El Solh Beyrouth 1107 2050, Liban
Rami.El-Haddad@univ-savoie.fr

² LAMA, Université de Savoie,
73376 Le Bourget-du-Lac Cedex, France
Christian.Lecot@univ-savoie.fr

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

Summary. We propose and analyze a quasi-Monte Carlo (QMC) method for simulating a discrete-time Markov chain on a discrete state space of dimension $s \geq 1$. Several paths of the chain are simulated in parallel and reordered at each step. We provide a convergence result when the number N of simulated paths increases toward infinity. Finally, we present the results of some numerical experiments showing that our QMC algorithm converges faster as a function of N , at least in some situations, than the corresponding Monte Carlo (MC) method.

Keywords. Markov chains, quasi-Monte Carlo, simulation.

1 Introduction

Markov chains are used in many fields such as physics, queueing theory, telecommunications, option pricing, etc. Very often the state space of the chain is so large that using the matrix equations to compute the state probabilities (or the expected cost) at each step is simply impractical, both in terms of computing time and required memory size. Then, the only viable method is MC simulation. Despite the versatility of MC methods, one drawback is their slow convergence: Based on the central limit theorem, their convergence rate is roughly of $\mathcal{O}(N^{-1/2})$ if N denotes the number of copies of the chain that are simulated.

One possible approach to accelerate the computation is to replace the random numbers by low discrepancy sequences, i.e., quasi-random numbers.

This is the general idea of QMC methods, which outperform MC methods in some cases, but also have limitations. Frequently, we want to estimate the expectation of a cost function that depends on the sample path of a Markov chain over several steps. This can be reformulated as an integration problem over the s' -dimensional unit hypercube, where s' represents the total number of $\mathcal{U}(0, 1)$ (uniform over $(0,1)$) random variates needed to run the simulation. In this case, the classical QMC method would use an s' -dimensional low-discrepancy point set of cardinality N , use each point to simulate one copy of the chain, and estimate the expected cost by the average over these N copies. But this s' is usually very large, so we end up with an integration problem in a very large number of dimensions, in which case QMC is typically not very effective.

A QMC algorithm for the simulation of Markov chains with a one-dimensional state space was studied in [LT04]. Randomized variants of this method were proposed and examined in [LLT06]. In the present paper, we generalize the method of [LT04] by proposing a QMC algorithm for Markov chains with multidimensional state spaces. This algorithm employs a low-discrepancy sequence with the property that each subsequence of length N starting at an index which is a multiple of N in the sequence has a low discrepancy in a sense to be specified. At each step, it uses one such subsequence to advance the N chains by one step, after matching the chains with the points in a clever way. This matching is done by sorting both the chains and the points according to their successive coordinates.

The remainder of the paper is organized as follows. In Section 2, we present the algorithm, which simulates the N sample paths of the chain in parallel using a low-discrepancy sequence. In Section 3 we adapt the basic concepts of QMC methods to the present study and we recall some definitions and properties related to the variation of multi-dimensional sequences. In Section 4, under a certain assumption on the transition matrix, we prove a convergence bound on the *worst-case error* for our method. This assumption could certainly be relaxed, at the expense of more complicated notation in the convergence proof. Finally, in Section 5, we present the results of numerical experiments that compare our method with standard MC. The convergence rate observed empirically for our method is much better than for MC and also much better than what is guaranteed by the worst-case bound.

2 The method

We consider a time-homogeneous discrete-time Markov chain $\{X_n, n \in \mathbb{N}\}$ with state space \mathbf{E} of the form $\mathbf{E} := \prod_{r=1}^s E_r$, where $E_r \subseteq \mathbb{Z}$. The initial state X_0 has distribution λ_0 , so $\mathbb{P}[X_0 = \mathbf{i}] = \lambda_0\{\mathbf{i}\}$ for each $\mathbf{i} = (i_1, \dots, i_s) \in \mathbf{E}$, and the transition matrix is $\mathcal{P} = (p(\mathbf{i}, \mathbf{j}) : \mathbf{i}, \mathbf{j} \in \mathbf{E})$, where $p(\mathbf{i}, \mathbf{j}) = \mathbb{P}[X_n = \mathbf{j} \mid X_{n-1} = \mathbf{i}]$. The probability that the chain is in state \mathbf{i} after n steps is

$$\lambda_n\{\mathbf{i}\} = \mathbb{P}[X_n = \mathbf{i}] = \lambda_0 \mathcal{P}^n \{\mathbf{i}\}$$

for all $\mathbf{i} \in \mathbf{E}$. Our aim is to estimate either (a) the function (or vector of probabilities) λ_n , or (b) the expected cost $\mathbb{E}[w(X_n)]$ at step n , for some function $w : \mathbf{E} \rightarrow [0, \infty)$. For the remainder of this section, w denotes a bounded function $w : \mathbf{E} \rightarrow [0, \infty)$; it is also called a *sequence with multivariate indices* (the indices are the elements of \mathbf{E}).

Let $\delta_{\mathbf{i}}$ be the row vector of unit mass at $\mathbf{i} = (i_1, \dots, i_s)$ defined, for all $\mathbf{j} = (j_1, \dots, j_s) \in \mathbf{E}$, by

$$\delta_{\mathbf{i}}\{\mathbf{j}\} = \begin{cases} 1 & \text{if } i_1 = j_1, \dots, i_s = j_s, \\ 0 & \text{otherwise.} \end{cases}$$

We denote by $\delta_{\mathbf{i}}w$ the real number $w(\mathbf{i})$. In case (a) above, we are looking for an approximation of λ_n , of the form

$$\widehat{\lambda}_n := \frac{1}{N} \sum_{0 \leq \ell < N} \delta_{\mathbf{i}_{\ell}^n}$$

for some integer N and for judiciously selected states $\mathbf{i}_0^n, \dots, \mathbf{i}_{N-1}^n \in \mathbf{E}$.

A simple approximation can be obtained by the *MC method*. For that, we assume that from each state $\mathbf{i} \in \mathbf{E}$, the interval $I = [0, 1)$ has been partitioned in subintervals $\mathcal{I}_{\mathbf{i}, \mathbf{j}} := [m_{\mathbf{i}, \mathbf{j}}, m'_{\mathbf{i}, \mathbf{j}})$ where $m'_{\mathbf{i}, \mathbf{j}} = m_{\mathbf{i}, \mathbf{j}} + p(\mathbf{i}, \mathbf{j})$, for all $\mathbf{j} \in \mathbf{E}^{\mathbf{i}} := \{\mathbf{j} \in \mathbf{E} : p(\mathbf{i}, \mathbf{j}) > 0\}$, so that $\cup_{\mathbf{j} \in \mathbf{E}^{\mathbf{i}}} \mathcal{I}_{\mathbf{i}, \mathbf{j}} = [0, 1)$. Now, for all $y \in I$, there is a unique $\mathbf{j} \in \mathbf{E}^{\mathbf{i}}$ such that $y \in \mathcal{I}_{\mathbf{i}, \mathbf{j}}$. We denote this element by $\mathbf{j}(\mathbf{i}, y)$. The MC method simulates the chain as follows. At step n , if we are in state $X_{n-1} = \mathbf{i}$, we generate a $\mathcal{U}(0, 1)$ random variate U_n and the next state is $X_n = \mathbf{j}(\mathbf{i}, U_n)$. This is repeated N times independently to obtain N i.i.d. copies of X_n .

Before defining our QMC approximation, we recall the notion of $(t, s+1)$ -sequence and (t, m, s) -net. Let $I^s := [0, 1)^s$ denotes the s -dimensional half open unit cube. For an integer $b \geq 2$, an *elementary interval in base b* is a subinterval of I^s of the form

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

for some integers $d_r \geq 0$ and $0 \leq a_r < b^{d_r}$ for all $1 \leq r \leq s$. If $0 \leq t \leq m$ are integers, a (t, m, s) -net in base b is a set Y of b^m points in I^s such that every elementary interval Q in base b with Lebesgue-measure (or volume) b^{t-m} contains exactly b^t points of Y . An infinite sequence of points $\mathbf{y}_0, \mathbf{y}_1, \dots$ in I^s is a (t, s) -sequence in base b if for all integers $n \geq 1$ and $m > t$, the set $Y_n = \{\mathbf{y}_p : (n-1)b^m \leq p < nb^m\}$ is a (t, m, s) -net in base b .

Choose a base $b \geq 2$ and non-negative integers d_1, \dots, d_s . Put $m := d_1 + \dots + d_s$ and $N := b^m$. For the QMC approximation, we assume that $Y = \{\mathbf{y}_0, \mathbf{y}_1, \dots\} \subset I^{s+1}$ is a $(t, s+1)$ -sequence in base b for some integer $t \geq 0$, and such that if $\Pi : I^{s+1} \rightarrow I^s$ denotes the projection defined by

$$(\mathbf{x}_1, \dots, \mathbf{x}_{s+1}) \xrightarrow{\Pi} (\mathbf{x}_1, \dots, \mathbf{x}_s) =: \mathbf{x}',$$

then the point set $\Pi(Y_n)$ is a $(0, m, s)$ -net in base b for each n . This implies that $b \geq s - 1$. The method is comprised of three steps:

1. Discretize the initial distribution.
2. Relabel the states before each transition.
3. Perform QMC integration for one transition of the chains.

We now explain them.

Discretizing the initial distribution. Initially, a set Ξ_0 of N states $\mathbf{i}_0^0, \dots, \mathbf{i}_{N-1}^0$ is chosen such that

$$\widehat{\lambda}_0 := \frac{1}{N} \sum_{0 \leq \ell < N} \delta_{\mathbf{i}_\ell^0} \approx \lambda_0.$$

The approximation sign here means that the point set Ξ_0 has a small star λ_0 -discrepancy (this will be defined more precisely in Section 3).

At step $n + 1$, given that we have a set Ξ_n of N states $\mathbf{i}_0^n, \dots, \mathbf{i}_{N-1}^n$ such that

$$\widehat{\lambda}_n := \frac{1}{N} \sum_{0 \leq \ell < N} \delta_{\mathbf{i}_\ell^n} \approx \lambda_n,$$

we start computing $\widehat{\lambda}_{n+1}$ by sorting the set Ξ_n as we now explain.

Relabeling the states. The states are labeled $\mathbf{i}_\mathbf{k}^n = (i_{\mathbf{k},1}^n, \dots, i_{\mathbf{k},s}^n)$, using a multi-dimensional index $\mathbf{k} = (k_1, \dots, k_s)$, with $0 \leq k_r < b^{d_r}$ for $1 \leq r \leq s$, such that:

$$\begin{aligned} & \text{if } k_1 < l_1, \text{ then } i_{\mathbf{k},1}^n \leq i_{\mathbf{l},1}^n, \\ & \text{if } k_1 = l_1, k_2 < l_2, \text{ then } i_{\mathbf{k},2}^n \leq i_{\mathbf{l},2}^n, \\ & \quad \vdots \\ & \text{if } k_1 = l_1, \dots, k_{s-1} = l_{s-1}, k_s < l_s, \text{ then } i_{\mathbf{k},s}^n \leq i_{\mathbf{l},s}^n. \end{aligned}$$

These conditions can be interpreted as follows. The N states are first sorted in b^{d_1} batches of size Nb^{-d_1} according to their first coordinates; then each batch is sorted in subgroups of b^{d_2} batches of size $Nb^{-d_1-d_2}$ by order of the second coordinates, and so on. At the last step of the sort, subgroups of size b^{d_s} are ordered according to the last coordinate of the state.

A graphical illustration is given in Figure 1, with $b = 2$, $s = 2$, and $d_1 = d_2 = 2$. Here we have $N = b^{d_1+d_2} = 16$ points in $s = 2$ dimensions. We first sort these points in four groups of four points, according to their first coordinate (which corresponds to the horizontal axis in the figure), and then sort each group according to the second coordinate.

Advancing the chains by one step via QMC integration. Let $\widetilde{\lambda}_{n+1} = \widehat{\lambda}_n \mathcal{P}$. We have

$$\widetilde{\lambda}_{n+1} w = \frac{1}{N} \sum_{\mathbf{k} \in \mathcal{K}} (\mathcal{P}w)(\mathbf{i}_\mathbf{k}^n) = \frac{1}{N} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{\mathbf{j} \in \mathbf{E}} p(\mathbf{i}_\mathbf{k}^n, \mathbf{j}) w(\mathbf{j}) \quad (1)$$

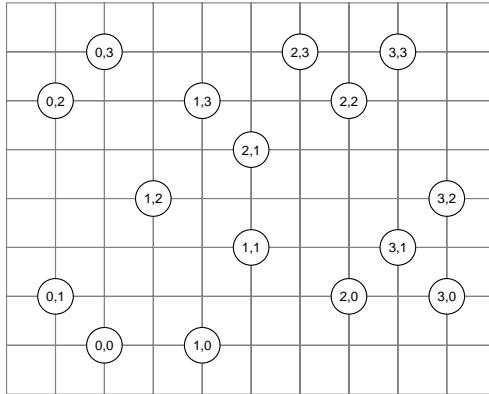


Fig. 1. Relabeling the states ($b = 2, s = 2, d_1 = d_2 = 2$)

where $\mathcal{K} = \{0, \dots, b^{d_1} - 1\} \times \dots \times \{0, \dots, b^{d_s} - 1\}$. This expression could be seen as the expected average cost at the next step, given the current set of states Ξ_n . For $\mathbf{k} = (k_1, \dots, k_s) \in \mathcal{K}$, denote by $\chi_{\mathbf{k}}$ the indicator function of the s -dimensional elementary interval

$$\mathcal{I}_{\mathbf{k}} = \prod_{r=1}^s \left[\frac{k_r}{b^{d_r}}, \frac{k_r + 1}{b^{d_r}} \right).$$

At any given step, if a chain in state \mathbf{i} is matched with a point $\mathbf{y} = (\mathbf{y}', y_{s+1})$, then this chain will move to state $\mathbf{j}(\mathbf{i}, y_{s+1})$, where the latter is defined as in the MC method.

For any given point $\mathbf{y} = (\mathbf{y}', y_{s+1}) \in I^{s+1}$, let

$$\mathcal{G}^n w(\mathbf{y}) := \sum_{\mathbf{k} \in \mathcal{K}} \chi_{\mathbf{k}}(\mathbf{y}') w(\mathbf{j}(\mathbf{i}_{\mathbf{k}}^n, y_{s+1})), \tag{2}$$

which represents the cost at the next transition if we use \mathbf{y} to move by one step the chain associated with the point of index \mathbf{k} . Integrating with respect to \mathbf{y} , we get

$$\tilde{\lambda}_{n+1} w = \int_{I^{s+1}} \mathcal{G}^n w(\mathbf{y}) d\mathbf{y}, \tag{3}$$

whereas averaging over the point set used at step $n + 1$, we obtain the QMC estimator $\hat{\lambda}_{n+1}$ of λ_{n+1} defined by

$$\hat{\lambda}_{n+1} w = \frac{1}{N} \sum_{nN \leq p < (n+1)N} \mathcal{G}^n w(\mathbf{y}_p).$$

For any $\mathbf{y}' = (y_1, \dots, y_s) \in I^s$, we define the multidimensional index

$$\mathbf{k}(\mathbf{y}') := ([b^{d_1} y_1], \dots, [b^{d_s} y_s]) \in \mathcal{K},$$

where $[x]$ denotes the largest integer $\leq x$. Because $\Pi(Y_n)$ is a $(0, m, s)$ -net in base b , the function

$$p \in \{nN, nN + 1, \dots, (n+1)N - 1\} \mapsto \mathbf{k}(\mathbf{y}'_p) \in \mathcal{K}$$

is one-to-one. Combining this with (2), we have

$$\frac{1}{N} \sum_{nN \leq p < (n+1)N} \mathcal{G}^n w(\mathbf{y}_p) = \frac{1}{N} \sum_{nN \leq p < (n+1)N} w(\mathbf{j}(\mathbf{i}_{\mathbf{k}(\mathbf{y}'_p)}^n, y_{p,s+1})). \quad (4)$$

Then, at step $n+1$, the point set $\Xi_{n+1} = \{\mathbf{i}_0^{n+1}, \dots, \mathbf{i}_{N-1}^{n+1}\} \subset \mathbf{E}$ is computed according to

$$\mathbf{i}_{p-nN}^{n+1} = \mathbf{j}(\mathbf{i}_{\mathbf{k}(\mathbf{y}'_p)}^n, y_{p,s+1}), \quad nN \leq p < (n+1)N.$$

This means that the projection \mathbf{y}'_p of each point \mathbf{y}_p of the low discrepancy sequence on the first s axes is used to select the state that is matched to this point at that step (i.e., which chain advances by one step), while the remaining component $y_{p,s+1}$ is used to determine the evolution (the next state).

3 Discrepancies and variations of sequences

The efficiency of a QMC method depends on the uniformity of the quasi-random points that are used. These points should form a low discrepancy point set. In this section, after recalling classical notions of discrepancy from [Nie92], we define and examine discrepancy measures adapted to the context of our method.

The star discrepancy. For an s -dimensional point set $Y = \{\mathbf{y}_0, \dots, \mathbf{y}_{N-1}\} \subset I^s$ and for a Lebesgue-measurable subset Q of I^s we define the *local discrepancy* by

$$D(Q, Y) := \frac{1}{N} \sum_{0 \leq p < N} \chi_Q(\mathbf{y}_p) - \int_{I^s} \chi_Q(\mathbf{x}) d\mathbf{x},$$

where χ_Q is the indicator function of Q . The *discrepancy* of the point set Y is defined by

$$D(Y) := \sup_Q |D(Q, Y)|,$$

the supremum being taken over all subintervals of I^s . The *star discrepancy* of Y is

$$D^*(Y) := \sup_{Q^*} |D(Q^*, Y)|,$$

where Q^* runs through all subintervals of I^s with one vertex at the origin.

The following result is shown in [Nie87].

Lemma 1. *Let Y be a (t, m, s) -net in base b . For any elementary interval $Q' \subset I^{s-1}$ in base b and for any $\xi \in \bar{I} := [0, 1]$, we have*

$$|D(Q' \times [0, \xi], Y)| \leq b^{t-m}.$$

The star λ -discrepancy. Let λ be a distribution on \mathbf{E} and consider a set of states $\Xi = \{\mathbf{i}_0, \dots, \mathbf{i}_{N-1}\} \subset \mathbf{E}$. For an arbitrary $F \subset \mathbf{E}$, we define the *local λ -discrepancy* of Ξ for F by

$$D(F; \Xi, \lambda) := \frac{1}{N} \sum_{0 \leq \ell < N} \chi_F(\mathbf{i}_\ell) - \sum_{\mathbf{i} \in F} \lambda\{\mathbf{i}\},$$

where χ_F denotes the function

$$\chi_F(\mathbf{i}) = \begin{cases} 1 & \text{if } \mathbf{i} \in F, \\ 0 & \text{otherwise.} \end{cases}$$

The *star λ -discrepancy* of the point set Ξ is defined by

$$D^*(\Xi, \lambda) := \sup_{\mathbf{h} \in \mathbf{E}} |D(F_{\mathbf{h}}; \Xi, \lambda)|,$$

where $F_{\mathbf{h}} = \prod_{r=1}^s ((-\infty, h_r) \cap E_r)$. If w is a nonnegative and bounded sequence, we set

$$D(w; \Xi, \lambda) := \frac{1}{N} \sum_{0 \leq \ell < N} w(\mathbf{i}_\ell) - \sum_{\mathbf{i} \in \mathbf{E}} \lambda\{\mathbf{i}\} w(\mathbf{i}),$$

so that if $F \subset \mathbf{E}$, we have $D(F; \Xi, \lambda) = D(\chi_F; \Xi, \lambda)$. Similarly, if $Y' = \{\mathbf{y}'_0, \dots, \mathbf{y}'_{N-1}\} \subset I^s$ and f is a nonnegative and bounded function defined on I^s , we put

$$D(f, Y') = \frac{1}{N} \sum_{0 \leq \ell < N} f(\mathbf{y}'_\ell) - \int_{I^s} f(\mathbf{x}') d\mathbf{x}'.$$

If $Q \subset I^s$, then $D(\chi_Q, Y') = D(Q, Y')$.

Variations of sequences. For $w : \mathbf{E} \rightarrow \mathbb{R}$ and $\mathbf{j}, \mathbf{j}' \in \mathbf{E}$, let $T_{\mathbf{j}}^r w$ and $\Delta_{\mathbf{j}, \mathbf{j}'}^r w$ be the functions (or multivariate sequences) defined by

$$T_{\mathbf{j}}^r w(\mathbf{i}) := w(i_1, \dots, i_{r-1}, j_r, i_{r+1}, \dots, i_s) \quad \text{and} \quad \Delta_{\mathbf{j}, \mathbf{j}'}^r w := T_{\mathbf{j}'}^r w - T_{\mathbf{j}}^r w.$$

If $R = \{r_1, \dots, r_q\} \subset S = \{1, \dots, s\}$, we denote

$$T_{\mathbf{j}}^R w := T_{\mathbf{j}}^{r_1} \dots T_{\mathbf{j}}^{r_q} w \quad \text{and} \quad \Delta_{\mathbf{j}, \mathbf{j}'}^R w := \Delta_{\mathbf{j}, \mathbf{j}'}^{r_1} \dots \Delta_{\mathbf{j}, \mathbf{j}'}^{r_q} w.$$

When $R = S$, we put $T_{\mathbf{j}} w := T_{\mathbf{j}}^S w$ and $\Delta_{\mathbf{j}, \mathbf{j}'} w := \Delta_{\mathbf{j}, \mathbf{j}'}^S w$. For $\mathbf{j} \in \mathbf{E}$, let $\mathbf{j}+$ be the vector $(j_1 + 1, \dots, j_s + 1)$ and $\Delta_{\mathbf{j}} w$ the sequence $\Delta_{\mathbf{j}, \mathbf{j}+} w$. Let $\mathbf{E}' = \{\mathbf{i} \in \mathbf{E} : \mathbf{i}+ \in \mathbf{E}\}$. The *variation in the sense of Vitali* of $w : \mathbf{E} \rightarrow \mathbb{R}$ is defined by

$$V^s(w) = \sum_{\mathbf{j} \in \mathbf{E}'} |\Delta_{\mathbf{j}} w|,$$

and the *variation of w in the sense of Hardy and Krause* is the sum

$$V(w) = \sum_{r=1}^s \sum_{\substack{R \subset S \\ \#R=r}} V^r(T_{\mathbf{0}}^{R^c} w)$$

where $\mathbf{0} = (0, \dots, 0)$.

Let $\mathbf{M} = (M_1, \dots, M_s)$ be the vector with coordinates $M_r = \sup E_r = \sup\{i : i \in E_r\}$. If w can be extended to \mathbf{M} , we define the *upper variation* of w as

$$V^*(w) = \sum_{r=1}^s \sum_{\substack{R \subset S \\ \#R=r}} V^r(T_{\mathbf{M}}^{R^c} w).$$

One can prove that if w is of bounded variation in the sense of Hardy and Krause, then w may be extended to \mathbf{M} and has a bounded upper variation. The next Lemma is a version of the classical Koksma-Hlawka inequality. The proof follows the general outline of the proof given by Zaremba [Zar68].

Lemma 2. *Let λ be a distribution on \mathbf{E} . If w is a sequence of bounded variation in the sense of Hardy and Krause and if $\Xi = \{\mathbf{i}_0, \dots, \mathbf{i}_{N-1}\} \subset \mathbf{E}$, then*

$$|D(w; \Xi, \lambda)| \leq V^*(w) D^*(\Xi, \lambda).$$

The following Lemma is an analogue of a result previously given in the continuous case in [Lec96] and can be proved with similar arguments.

Lemma 3. *Let $w : \mathbf{E} \rightarrow \mathbb{R}$ be a sequence of bounded variation in the sense of Hardy and Krause and p_1, \dots, p_s be integers. We consider a nested partition of \mathbf{E} of the form*

$$\begin{aligned} & \cdot g_{0,1} \leq g_{1,1} \leq \dots \leq g_{p_1,1} \in E_1 \\ & \cdot g_{k_1,0,2} \leq g_{k_1,1,2} \leq \dots \leq g_{k_1,p_2,2} \in E_2 \text{ for } 0 \leq k_1 < p_1. \\ & \cdot \dots \\ & \cdot g_{\mathbf{k}',0,s} \leq g_{\mathbf{k}',1,s} \leq \dots \leq g_{\mathbf{k}',p_s,s} \in E_s \text{ for } \mathbf{k}' = (k_1, \dots, k_{s-1}) \text{ with } 0 \leq k_1 < p_1, \dots, 0 \leq k_{s-1} < p_{s-1}. \end{aligned}$$

For each $\mathbf{k} = (k_1, \dots, k_s) \in P := \prod_{r=1}^s \{0, \dots, p_{r-1}\}$, let

$$\mathbf{i}_{\mathbf{k}}, \mathbf{j}_{\mathbf{k}} \in \{g_{k_1,1}, \dots, g_{k_1+1,1}\} \times \{g_{k_1,k_2,2}, \dots, g_{k_1,k_2+1,2}\} \times \dots \times \{g_{\mathbf{k}',k_s,s}, \dots, g_{\mathbf{k}',k_s+1,s}\}.$$

Then we have the following inequality

$$\sum_{\mathbf{k} \in P} |w(\mathbf{j}_{\mathbf{k}}) - w(\mathbf{i}_{\mathbf{k}})| \leq V^*(w) \prod_{r=1}^s p_r \sum_{r=1}^s \frac{1}{p_r}.$$

Equipped with these tools, we now return to the convergence of the QMC algorithm.

4 Convergence analysis

We provide a convergence result under the following simplifying assumption: We suppose that the transition probabilities satisfy

$$p(\mathbf{i}, \mathbf{j}) = 0 \text{ if there exists } 1 \leq q \neq r \leq s \text{ such that } i_q \neq j_q \text{ and } i_r \neq j_r. \quad (5)$$

In other words, we assume that only one coordinate of the state can be changed at any one step. (The result could certainly be generalized, but this would require more complicated notations.) Under this condition, we can take

$$m_{\mathbf{i}, \mathbf{j}} = p(\mathbf{i}, \mathbf{i}) + \sum_{\substack{\mathbf{g} \in \mathbf{E}^{\mathbf{i}^*} \\ g_1 \neq i_1}} p(\mathbf{i}, \mathbf{g}) + \cdots + \sum_{\substack{\mathbf{g} \in \mathbf{E}^{\mathbf{i}^*} \\ g_{r-1} \neq i_{r-1}}} p(\mathbf{i}, \mathbf{g}) + \sum_{\substack{\mathbf{g} \in \mathbf{E}^{\mathbf{i}^*}, g_r < j_r \\ g_\ell = i_\ell, \ell \neq r}} p(\mathbf{i}, \mathbf{g})$$

in the definition of the intervals $\mathcal{I}_{\mathbf{i}, \mathbf{j}}$. We set:

$$q(\mathbf{i}) := p(\mathbf{i}, \mathbf{i}), \quad p_r(\mathbf{i}) := \sum_{\substack{\mathbf{g} \in \mathbf{E}^{\mathbf{i}^*} \\ g_r \neq i_r}} p(\mathbf{i}, \mathbf{g}),$$

$$p_{h_r, r}(\mathbf{i}) := \sum_{\substack{\mathbf{g} \in \mathbf{E}^{\mathbf{i}^*}, g_r < h_r \\ g_\ell = i_\ell, \ell \neq r}} p(\mathbf{i}, \mathbf{g}), \quad \forall h_r \in E_r.$$

We assume that these sequences are of bounded variation in the sense of Hardy and Krause and that there exists positive constants c_0 , $c_{1,r}$ and $c_{2,r}$, for $1 \leq r \leq s$, such that

$$V^*(q) \leq c_0, \quad V^*(p_r) \leq c_{1,r}, \quad \text{and} \quad V^*(p_{h_r, r}) \leq c_{2,r}.$$

We have the following *worst-case* error bound.

Proposition 1. *If the transition matrix of the chain satisfies $V^*(\mathcal{P}\chi_{F_{\mathbf{h}}}) \leq 1$ for all $\mathbf{h} \in \mathbf{E}$ and under the previous assumptions, there exists a constant $C(s) \geq 0$ depending only on s , such that*

$$D^*(\Xi_n, \lambda_n) \leq D^*(\Xi_0, \lambda_0) + \frac{(4s+1)n}{b^{\lfloor (d_s-t)/2 \rfloor}} + C(s)n \left(\frac{1}{b^{d_1}} + \cdots + \frac{1}{b^{d_{s-1}}} + \frac{1}{b^{\lfloor (d_s-t)/2 \rfloor}} \right).$$

Proof. The proof of this proposition is quite technical; We only give a sketch of the proof in what follows. For any nonnegative and bounded sequence w , write

$$D(w; \Xi_{n+1}, \lambda_{n+1}) = D(\mathcal{P}w; \Xi_n, \lambda_n) + D(\mathcal{G}^n w, Y_n).$$

If we take $w = \chi_{F_{\mathbf{h}}}$, $\mathbf{h} \in E$, we get

$$D(F_{\mathbf{h}}; \Xi_{n+1}, \lambda_{n+1}) = D(\mathcal{P}\chi_{F_{\mathbf{h}}}; \Xi_n, \lambda_n) + D(\mathcal{G}^n \chi_{F_{\mathbf{h}}}, Y_n).$$

By Lemma 2 and using the inequality $V^*(\mathcal{P}\chi_{F_{\mathbf{h}}}) \leq 1$, we have

$$|D(\mathcal{P}\chi_{F_{\mathbf{h}}}; \Xi_n, \lambda_n)| \leq V^*(\mathcal{P}\chi_{F_{\mathbf{h}}})D^*(\Xi_n, \lambda_n) \leq D^*(\Xi_n, \lambda_n).$$

On the other hand, $\mathcal{G}^n \chi_{F_{\mathbf{h}}}$ is the indicator function of

$$Q_{\mathbf{h}}^n := \bigcup_{\mathbf{k} \in \mathcal{K}} (\mathcal{I}_{\mathbf{k}} \times \bigcup_{\substack{\mathbf{j} \in \mathbf{E}^{\mathbf{i}_{\mathbf{k}}} \\ \mathbf{j} < \mathbf{h}}} \mathcal{I}_{\mathbf{i}_{\mathbf{k}}, \mathbf{j}}^n),$$

where $\mathbf{j} < \mathbf{h}$ means that $j_1 < h_1, \dots, j_s < h_s$. Thus, $D(\mathcal{G}^n \chi_{F_{\mathbf{h}}}, Y_n) = D(Q_{\mathbf{h}}^n, Y_n)$. By decomposing $Q_{\mathbf{h}}^n$ into $s+1$ disjoint subsets and studying the local discrepancy for each one of them apart using Lemmas 1 and 3, we get the following bound

$$|D(Q_{\mathbf{h}}^n, Y_n)| \leq \frac{4s+1}{b^{\lfloor (d_s-t)/2 \rfloor}} + C(s) \left(\frac{1}{b^{d_1}} + \dots + \frac{1}{b^{d_{s-1}}} + \frac{1}{b^{\lfloor (d_s-t)/2 \rfloor}} \right),$$

where $C(s)$ is a positive constant. The desired inequality is then obtained by induction on n .

By taking

$$d_1 = \dots = d_{s-1} = \left\lfloor \frac{m-t}{s+1} \right\rfloor \quad \text{and} \quad d_s = m - (s-1) \left\lfloor \frac{m-t}{s+1} \right\rfloor,$$

the proposition shows that the error converges as $\mathcal{O}(N^{-1/(s+1)})$ in the worst case.

5 Numerical examples

In this Section, we assess the accuracy of the QMC algorithm empirically, through two (academic) examples where exact solutions are known. We show the kind of improvement that our method can bring with respect to MC. For our experiment, we use Niederreiter's sequences in base 2. For MC, the pseudo-random numbers are produced by the generator MRG32k3a of [L'Ec99]. Convergence speed is assessed by looking at the absolute difference between the empirical and theoretical means, for several values of N .

5.1 A bivariate Markovian asset valuation model

We consider a bivariate extension of Cox-Ross-Rubinstein's binomial single asset pricing model, as proposed in [HKY03]. At time n , the values of the two risky assets are denoted by S_n^1 and S_n^2 , and $S_n = (S_n^1, S_n^2)^{\mathbf{t}}$ is the price vector. The sequence $(S_n)_{n \geq 0}$ is defined in term of two independent and identically

distributed sequences $(U_n^1)_{n \geq 0}$ and $(U_n^2)_{n \geq 0}$: For all $n \in \mathbb{N}$, U_n^1 can take the two values a and b , where $-1 < a < b$, with probabilities p and $1 - p$, respectively, while U_n^2 can take the two values c and d , where $-1 < c < d$, with probabilities q and $1 - q$, respectively. We suppose that $0 < p, q < 1$. The model is a Markov chain whose state evolves as

$$\begin{pmatrix} S_n^1 \\ S_n^2 \end{pmatrix} = \begin{pmatrix} 1 + U_n^1 & \varepsilon(U_n^2 - r) \\ \delta(U_n^1 - r) & 1 + U_n^2 \end{pmatrix} \begin{pmatrix} S_{n-1}^1 \\ S_{n-1}^2 \end{pmatrix}, \quad (6)$$

where ε and δ are small numbers expressing the perturbation caused by S_n^2 on S_n^1 and by S_n^1 on S_n^2 . We suppose $\varepsilon\delta \neq 1$ and $(\varepsilon, \delta) \neq (0, 0)$. At each step, there are four possibilities for the vector (U_n^1, U_n^2) . In order to follow the geometry of the transition, we partition the unit square $[0, 1]^2$ in four rectangles

$$[0, p) \times [0, q), \quad [0, p) \times [q, 1), \quad [p, 1) \times [0, q), \quad [p, 1) \times [q, 1);$$

each one is assigned to one of the four possible outcomes.

We want to estimate $\mathbb{E}(S_n^1)$ and $\mathbb{E}(S_n^2)$ for some fixed n . These (exact) values are

$$\mathbb{E}(S_n^1) = (1 + r)^n \mathbb{E}(S_0^1) \text{ and } \mathbb{E}(S_n^2) = (1 + r)^n \mathbb{E}(S_0^2) \quad (7)$$

but for the purpose of our experiment, we pretend that they are unknown and have to be estimated. Note that here $Y \subset I^{s+2}$ and that this bidimensional Markov chain does not satisfy condition (5).

We take the following parameters: $n = 20$, $a = 0.074$, $b = 0.141$, $c = 0.086$, $d = 0.182$, $r = 0.1$, $\varepsilon = 0.30$, and $\delta = 0.20$. The initial values are $S_0^1 = 120$ and $S_0^2 = 130$. The number of states N varies from 2^4 to 2^{20} , and want to estimate the error as a function of N , say $\text{Err}_{\text{MC}}(N)$ for MC and say $\text{Err}_{\text{QMC}}(N)$ for our QMC method. Figure 2 shows the empirical values of these errors, in log-log scale, for both S_n^1 and S_n^2 . A linear regression analysis with this data gives the following empirical convergence rates:

$$\begin{aligned} \text{For } S_n^1, \text{ Err}_{\text{MC}} &= \mathcal{O}(N^{-0.41}) \text{ and } \text{Err}_{\text{QMC}} = \mathcal{O}(N^{-0.91}). \\ \text{For } S_n^2, \text{ Err}_{\text{MC}} &= \mathcal{O}(N^{-0.34}) \text{ and } \text{Err}_{\text{QMC}} = \mathcal{O}(N^{-0.89}). \end{aligned}$$

Clearly, the QMC method enjoys a much faster convergence than MC.

5.2 Pricing a European call on the maximum of two risky assets

For our second example, we consider the pricing of an European call option on the maximum of two risky assets, in a setting where the (continuous-time) evolution of the asset price vector is approximated by a (discrete-time) binomial lattice model. Again, the example is artificial and simplified, since the option price can be computed exactly in this case, but we want to use it as a benchmark to evaluate the viability of our method.

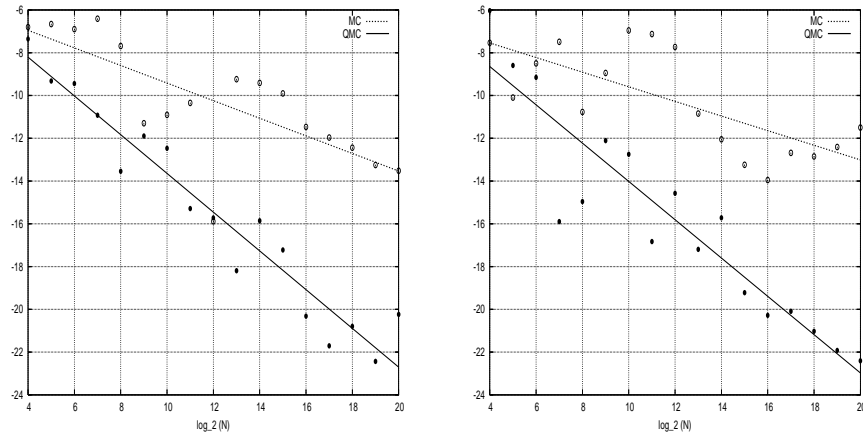


Fig. 2. Bivariate Markovian asset valuation model. Linear fits to the error as a function of N on log-log scale (in base 2), for S_1 (left) and S_2 (right), with MC (thin line) and QMC (thick line).

The original (continuous-time) model is a bivariate geometric Brownian motion (GBM) $\{S(t) = (S_1(t), S_2(t))^t, t \geq 0\}$ with *drift parameter* μ_i , *volatility parameter* σ_i , and *correlation parameter* ρ . Thus, for $i = 1, 2$,

$$S_i(t) = S_i(0) \exp [(\mu_i - \sigma_i^2/2)t + \sigma_i W_i(t)]$$

where W_i is a standard Brownian motion, and $\text{Cov}[W_1(t + \delta) - W_1(t), W_2(t + \delta) - W_2(t)] = \rho\delta$ for all $\delta > 0$. The option has discounted payoff

$$e^{-rT} \max[\max(S_1(T), S_2(T)) - K, 0]$$

for some constants $K > 0$ (the strike price) and $T > 0$ (the maturity), where r is the riskless rate. The expected value C of this payoff, which is the exact value of the option, can be computed by the formulas given by Stulz [Stu82] and Johnson [Joh87].

Instead of simulating the BGM directly (which can be done by simulating the Brownian motion (W_1, W_2) at the desired observation times), here we simulate a numerical approximation based on the multivariate binomial lattice method developed in [BEG89]. This method is an extension of the Cox-Ross-Rubinstein approach [CRR79]. It proceeds as follows.

A discrete-time model with discrete probability distribution is constructed to approximate the bivariate lognormal distribution. In this model, at each time step, each asset price can only move up or down (only two possibilities), so there is four possible transitions for the process. Let $h = T/P$ be the

length of the time step, where P is the number of steps. The value of $S_i(t)$ is multiplied by $\exp(\sigma_i\sqrt{h})$ in a move up and divided by this same value in a down move. The formulas for the transition probabilities of up and down moves (and other details) can be found in [BEG89]. They are selected in a way that the characteristic function of the discrete distribution at any fixed time point converges to that of the lognormal at that point, when $h \rightarrow 0$. This model does not satisfy condition (5).

We use the following parameter values (time is measured in years): $S_1(0) = S_2(0) = 40$, $\sigma_1 = 0.2$, $\sigma_2 = 0.3$, $\rho = 0.5$, $r = 0.05$, $T = 7/12$, $K = 35$. Let $C_{N,P}$ be the QMC approximation of C with N paths and P time steps. We measure the error with the following discrete L^1 norm:

$$\text{Err}_{N,P} = \frac{1}{20} \sum_{m=1}^{20} |C - C_{N,Pm/20}|.$$

Figure 3 shows the value of $\text{Err}_{N,P}$ for P varying from $2^2 \times 20$ to $2^6 \times 20$, and N varying from 2^5 to 2^{20} , for both the MC method (left panel), and QMC (right panel), in a log-log scale. Note that there are two sources of error here: (1) the discretization error and (2) the additional error due to using MC or QMC instead of solving the binomial lattice model exactly. The discretization error vanishes when $P \rightarrow \infty$, whereas the other source of error converges to zero when $N \rightarrow \infty$.

For MC, the error does not seem to depend much on P , which means that the discretization error is small compared with the MC (statistical) error. For QMC, this is also true for the small values of N , but not for the large ones. For MC, we have a slope of about $-1/2$, so the error converges as $\mathcal{O}(N^{-1/2})$ as a function of N , as expected. For QMC, when P is large, the slope is steeper than $-1/2$, which indicates a faster convergence rate than for MC. For small values of P , the error eventually reaches a plateau and stops decreasing when we keep increasing N ; this indicates that the QMC error eventually becomes negligible compared with the discretization error. This shows that a good strategy in this type of situation is to increase both P and N simultaneously. QMC clearly dominates MC in that case. For instance, for $P = 1200$ and $N = 2^{20}$, $\text{Err}_{N,P}$ is about 2^{-7} for MC and 2^{-11} for QMC. As another example, for $P = 320$, the same error level is attained by QMC with $N = 8192$ and by MC with $N = 524288$.

6 Conclusion

We have proposed and analyzed a QMC method for the simulation of discrete-time Markov chains on a multi-dimensional state space. The method simulates several copies of the chain in parallel and reduces the error by a technique that sorts the chain in a special way, based on the several coordinates of their states, at each step. We have proved a convergence result for the worst-case

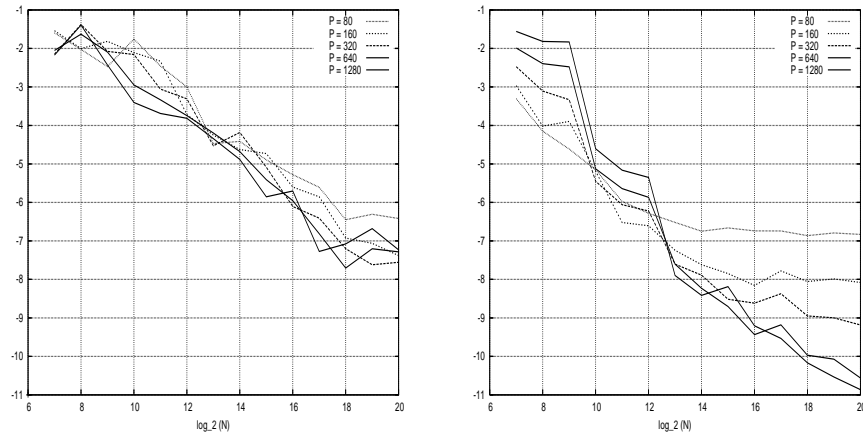


Fig. 3. Pricing an European call on the maximum of two risky assets by a binomial lattice model. The error $\text{Err}_{N,P}$ as a function of N in log-log scale (in base 2) for different numbers of time steps P , for MC (left) and QMC (right).

error as the number of simulated paths increases, under a special condition. In our empirical experiments, the performance of the proposed method was clearly superior to MC. Directions for future research include the theoretical analysis of the method in more general settings, experiments with larger and more complicated models, and the analysis of a randomized version of the method to produce unbiased low-variance estimators.

References

- [BEG89] Boyle, P. P., Evnine J., Gibbs, S.: Numerical evaluation of multivariate contingent claims. *The Review of Financial Studies*, **2**, 241–250 (1989)
- [CRR79] Cox, J.C., Ross, S.A., Rubinstein, M.: Option pricing : a simplified approach. *Journal of Financial Economics*, **7**, 229–263 (1979)
- [HKY03] Hayfavi, A., Körezlioğlu, H., Yildirak, K.: Bivariate extension of Cox-Ross-Rubinstein model: model identification. 16th Annual Conference of Greek Statistical Institute. Kavala, Greece, 30 April – 3 May (2003)
- [Joh87] Johnson, H.: Options on the maximum or the minimum of several assets. *Journal of Financial and Quantitative Analysis*, **22**, 277–283 (1987)
- [Lec96] Lécot, C.: Error bounds for quasi-Monte Carlo integration with nets. *Mathematics of Computation*, **65**, 179–187 (1996)
- [LT04] Lécot, C., Tuffin, B.: Quasi-Monte Carlo methods for estimating transient measures of discrete time Markov chains. In: Niederreiter, H. (ed.) *Monte*

- Carlo and Quasi Monte Carlo Methods 2002. Springer-Verlag, Berlin, 329–343 (2004)
- [L’Ec99] L’Ecuyer, P.: Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators. *Operations Research*, **47**, 159–164 (1999)
- [LLT06] L’Ecuyer, P., Lécot, C., and Tuffin B.: A randomized quasi-Monte Carlo simulation method for Markov chains. *Operations Research*, 2007, to appear.
- [Nie87] Niederreiter, H.: Point sets and sequences with small discrepancy. *Monatshefte für Mathematik*, **104**, 273–337 (1987)
- [Nie92] Niederreiter, H.: *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, (1992)
- [Stu82] Stulz, R. M.: Options on the minimum or the maximum of two risky assets: analysis and applications. *Journal of Financial Economics*, **10**, 161–185, (1982)
- [Zar68] Zaremba, S.K.: Some applications of multidimensional integration by parts. *Ann. Polon. Math*, **21**, 85–96 (1968)