Extended Latin Hypercube Sampling for Integration and Simulation

Rami El Haddad, Rana Fakhereddine, Christian Lécot, and Gopalakrishnan Venkiteswaran

Abstract We analyze an extended form of Latin hypercube sampling technique that can be used for numerical quadrature and for Monte Carlo simulation. The technique utilizes random point sets with enhanced uniformity over the s-dimensional unit hypercube. A sample of $N = n^s$ points is generated in the hypercube. If we project the N points onto their ith coordinates, the resulting set of values forms a stratified sample from the unit interval, with one point in each subinterval $[(k-1)/N, k/N)$. The scheme has the additional property that when we partition the hypercube into N subcubes $\prod_{i=1}^{s}[(\ell_i - 1)/n, \ell_i/n)$, each one contains exactly one point. We establish an upper bound for the variance, when we approximate the volume of a subset of the hypercube, with a regular boundary. Numerical experiments assess that the bound is tight. It is possible to employ the extended Latin hypercube samples for Monte Carlo simulation. We focus on the random walk method for diffusion and we show that the variance is reduced when compared with classical random walk using ordinary pseudo-random numbers. The numerical comparisons include stratified sampling and Latin hypercube sampling.

R. El Haddad (✉)
Faculté des Sciences, Département de Mathématiques, Université Saint-Joseph,
BP 11-514 Riad El Solh, Beyrouth 1107 2050, Liban
e-mail: rami.haddad@fs.usj.edu.lb

R. Fakhereddine · C. Lécot
Laboratoire de Mathématiques, Université de Savoie, UMR 5127 CNRS,
Campus scientifique, 73376 Le Bourget-du-Lac, France
e-mail: Rana.Fakhereddine@univ-savoie.fr; Christian.Lecot@univ-savoie.fr

G. Venkiteswaran
Department of Mathematics, Birla Institute of Technology and Science, Pilani,
Jhunjhunu 333 031, Rajasthan, India
e-mail: gvenki.bits@gmail.com

J. Dick et al. (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2012*,
Springer Proceedings in Mathematics & Statistics 65,
DOI 10.1007/978-3-642-41095-6_13, © Springer-Verlag Berlin Heidelberg 2013

1 Introduction

Approximating integrals is a basic problem of numerical analysis and may be a component in a more complex computation. Two families of techniques have been developed: deterministic methods and Monte Carlo. We only consider here random algorithms, which are parts of stochastic simulation methods used in applied sciences. Monte Carlo (MC) methods are known to converge slowly, with respect to the number of random points used. Various techniques have been developed, in order to reduce the variance of the approximation, including stratified sampling and Latin hypercube sampling [5,7,9].

Let $s \geq 1$ be a given dimension; then $I^s := [0,1)^s$ is the s-dimensional half-open unit hypercube and λ_s denotes the s-dimensional Lebesgue measure. If g is a square-integrable function defined on I^s, we want to approximate

$$\mathscr{I} := \int_{I^s} g(x) d\lambda_s(x). \tag{1}$$

For the usual MC approximation, $\{U_1, \ldots, U_N\}$ are independent random variables uniformly distributed over I^s. Then

$$X := \frac{1}{N} \sum_{k=1}^{N} g(U_k) \tag{2}$$

is an unbiased estimator of $\mathscr{I}$. A simple stratified sampling (SSS) method was proposed in [10]. Let $\{D_1, \ldots, D_N\}$ be a partition of I^s, so that $\lambda_s(D_1) = \cdots = \lambda_s(D_N) = 1/N$. Let $\{V_1, \ldots, V_N\}$ be independent random variables, with V_ℓ uniformly distributed over D_ℓ. Then

$$Y := \frac{1}{N} \sum_{\ell=1}^{N} g(V_\ell) \tag{3}$$

is another unbiased estimator of $\mathscr{I}$ and for a regular g, one has $\mathrm{Var}(Y) \leq \mathrm{Var}(X)$: we refer to [1, 2, 10] for variance reduction analyses. Latin hypercube sampling (LHS) was introduced in [15]. Let $I_\ell := [(\ell-1)/N, \ell/N)$ for $1 \leq \ell \leq N$ and $\{V_1^i, \ldots, V_N^i\}$ be independent random variables, where V_ℓ^i is uniformly distributed over I_ℓ. If $\{\pi^1, \ldots, \pi^s\}$ are independent random permutations of $\{1, \ldots, N\}$, put $W_\ell := (V^1_{\pi^1(\ell)}, \ldots, V^s_{\pi^s(\ell)})$. Then

$$Z := \frac{1}{N} \sum_{\ell=1}^{N} g(W_\ell) \tag{4}$$

is another unbiased estimator of $\mathscr{I}$. McKay et al. [15] showed that if g is a monotonic function of each of its argument then one has $\mathrm{Var}(Z) \leq \mathrm{Var}(X)$. The analysis in [21] established that for any square-integrable g, LHS does reduce the variance relative to simple random sampling in an asymptotic sense ($N \to \infty$). A proposition in [20] implied that an N-point Latin hypercube sample never leads to a variance greater than that of simple MC with $N - 1$ points. LHS stratifies only the one-dimensional marginals of the uniform distribution over the unit hypercube. Orthogonal array (OA)-based LHS was proposed in [18, 22]. This method generalizes LHS by stratifying low-dimensional ($\leq r$ for OA-based LHS with a corresponding orthogonal array of strength r) marginal distributions. Variance formulas of order $\mathscr{O}(N^{-1})$ were given by Owen [19, 20].

We analyze here a hybrid of SSS and LHS, where the random samples retain some uniformity properties of the nets used in quasi-Monte Carlo methods [17]. More precisely, we construct $N = n^s$ random points in I^s such that in every interval

$$I^{i-1} \times \left[\frac{k-1}{N}, \frac{k}{N}\right) \times I^{s-i} \text{ (for } 1 \leq i \leq s \text{ and } 1 \leq k \leq N)$$

or

$$I_\ell := \prod_{i=1}^{s} \left[\frac{\ell_i - 1}{n}, \frac{\ell_i}{n}\right) \text{ (for } 1 \leq i \leq s \text{ and } 1 \leq \ell_i \leq n)$$

lies only one point of the set (property $\mathscr{P}$): an example is shown on Fig. 1. We call this approach *extended Latin hypercube sampling* (ELHS). In contrast with OA-based LHS, ELHS achieves full (s-dimensional) stratification and also stratifies the one-dimensional marginals. The construction of extended Latin hypercube samples is elementary and requires only random permutations. Both methods are similar in the two-dimensional case.

In Sect. 2 we analyze a MC method using ELHS for numerical integration. Since we have experienced that some simulation methods can be reduced to numerical integration of indicator functions of subdomains of I^s, we focus here on the approximation of the volume of subsets of the unit hypercube. We prove a bound for the variance and we show through numerical experiments that the orders obtained are precise. We compare the variance of the following methods: usual MC, SSS, LHS and ELHS. In Sect. 3, we propose a random walk algorithm for one-dimensional diffusion. Each step of the simulation is formulated as a numerical integration in I^2. In order to benefit from the great uniformity of extended Latin hypercube samples, the particles are sorted by position before performing MC quadrature. The results of a numerical experiment show that the use of ELHS leads to reduced variance, when compared with usual MC, SSS or LHS.

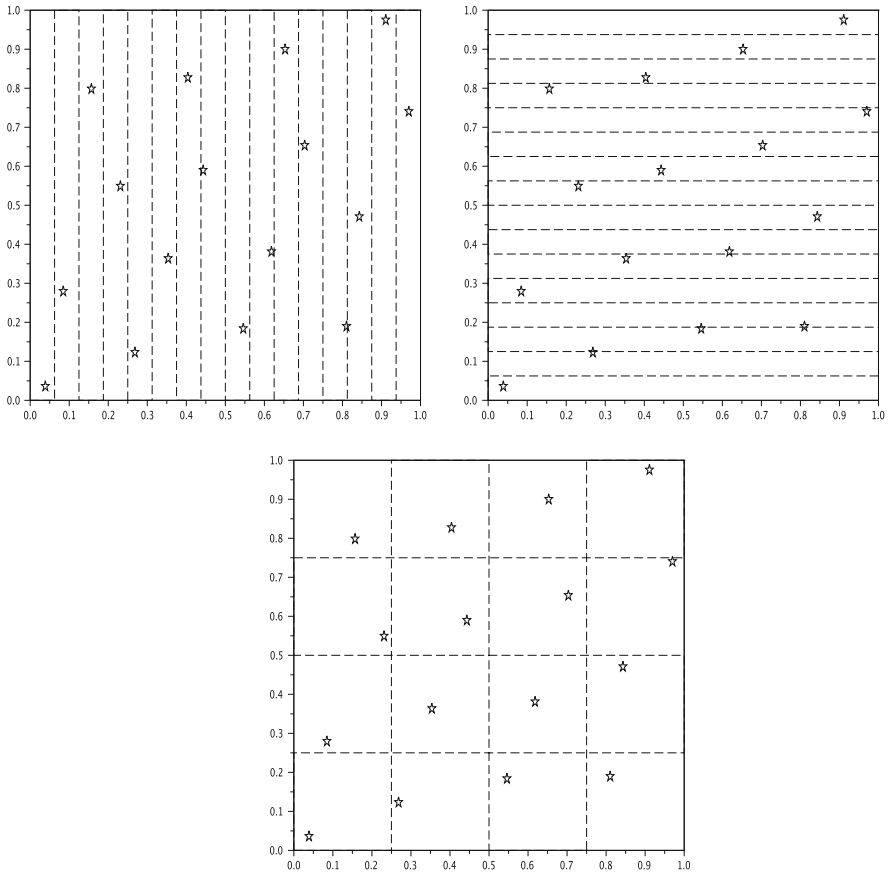

Fig. 1 An extended Latin hypercube sample of 4^2 points ($\star$) in dimension $s = 2$.

2 Numerical Integration

We consider the problem of evaluating integrals like (1) when $g = 1_A$, for some measurable $A \subset I^s$. For usual MC approximation (2), one has

$$\mathrm{Var}(X) = \frac{1}{N}\lambda_s(A)\big(1 - \lambda_s(A)\big) \leq \frac{1}{4N}. \tag{5}$$

We analyze here ELHS using samples of $N = n^s$ points. If $x := (x_1, \ldots, x_s)$, we put $\hat{x}_i := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_s)$. Let $\sigma^1, \ldots, \sigma^s$ be random bijections

$$\{1, \ldots, n\}^{s-1} \to \{1, \ldots, n^{s-1}\}$$

and $u^1, \ldots, u^s$ be random variables uniformly distributed on I^N; we assume that all these variables are mutually independent. Then we put

$$W_\ell = (W_\ell^1, \ldots, W_\ell^s) \quad \text{with} \quad W_\ell^i := \frac{\ell_i - 1}{n} + \frac{\sigma^i(\hat{\ell}_i) - 1 + u_\ell^i}{N}, \tag{6}$$

for $\ell := (\ell_1, \ldots, \ell_s)$ with $1 \le \ell_i \le n$. Then the point set $\{W_\ell : 1 \le \ell_i \le n\}$ has property $\mathscr{P}$. For $\ell = (\ell_1, \ldots, \ell_s)$ with $1 \le \ell_i \le n$ and $m = (m_1, \ldots, m_s)$ with $1 \le m_i \le n^{s-1}$, let

$$I_{\ell,m} := \prod_{i=1}^{s} \left[\frac{\ell_i - 1}{n} + \frac{m_i - 1}{N}, \frac{\ell_i - 1}{n} + \frac{m_i}{N} \right);$$

then $I_\ell = \bigcup_m I_{\ell,m}$. We have

$$E[1_A(W_\ell)] = \frac{1}{n^{s(s-1)}} \sum_m \int_{I^s} 1_A \left(\frac{\ell_1 - 1}{n} + \frac{m_1 - 1 + u_1}{N}, \ldots, \frac{\ell_s - 1}{n} + \frac{m_s - 1 + u_s}{N} \right) du$$

where the sum extends over all $m = (m_1, \ldots, m_s)$ with $1 \le m_i \le n^{s-1}$. Hence

$$E[1_A(W_\ell)] = N \int_{I_\ell} 1_A(u) du = N \lambda_s(I_\ell \cap A). \tag{7}$$

Consequently, if Z is defined by (4), it is an unbiased estimator of $\mathscr{I}$; we want to estimate $\mathrm{Var}(Z)$.

Proposition 1. *Let $A \subset I^s$ be such that, for all i, with $1 \le i \le s$,*

$$A = \{(u_1, \ldots, u_s) \in I^s : u_i < f_i(\hat{u}_i)\},$$

where $\hat{u}_i := (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_s)$ and f_i are Lipschitz continuous functions $\overline{I}^{s-1} \to \overline{I}$. Let $\{W_\ell : 1 \le \ell_i \le n\}$ be defined by (6). If

$$Z := \frac{1}{N} \sum_\ell 1_A(W_\ell),$$

then

$$\mathrm{Var}(Z) \le \left(\frac{k+2}{4} + 2s(k+2)^2 \right) \frac{1}{N^{1+1/s}},$$

where k is a Lipschitz constant (for the maximum norm) for all the f_i.

Proof. We may write

$$\mathrm{Var}(Z) = \frac{1}{N^2} \sum_\ell \mathrm{Var}(1_A(W_\ell)) + \frac{1}{N^2} \sum_{\ell \ne \ell'} \mathrm{Cov}(1_A(W_\ell), 1_A(W_{\ell'})).$$

From (7) we obtain

$$\frac{1}{N^2}\sum_{\ell}\mathrm{Var}(1_A(W_\ell)) = \sum_{\ell}\mathscr{V}_0(\ell),$$

where

$$\mathscr{V}_0(\ell) = \frac{1}{n^s}\lambda_s(I_\ell \cap A) - (\lambda_s(I_\ell \cap A))^2.$$

Since $\mathscr{V}_0(\ell) = 0$ whenever $I_\ell \subset A$ or $I_\ell \cap A = \emptyset$ and since $0 \le n^s\lambda_s(I_\ell \cap A) \le 1$, we have

$$\sum_{\ell}|\mathscr{V}_0(\ell)| \le \frac{1}{4n^{2s}}\#\{\ell : I_\ell \not\subset A \text{ and } I_\ell \cap A \neq \emptyset\}.$$

Here, $\#E$ denotes the number of elements of a set E. Similarly, we have

$$\frac{1}{N^2}\sum_{\ell\neq\ell'}\mathrm{Cov}(1_A(W_\ell), 1_A(W_{\ell'})) = \sum_{i=1}^{s}\sum_{\substack{\hat{\ell}_i=\hat{\ell}'_i\\ \ell_i\neq\ell'_i}}\mathscr{V}_i(\ell,\ell') + \sum_{\hat{\ell}_j\neq\hat{\ell}'_j}\mathscr{V}_{s+1}(\ell,\ell'),$$

where

$$\mathscr{V}_i(\ell,\ell') = \frac{n^{s(s-1)}}{(n^{s-1}-1)^{s-1}}\sum_{\substack{m_i=m'_i\\ m_j\neq m'_j}}\lambda_s(I_{\ell,m}\cap A)\lambda_s(I_{\ell',m'}\cap A) - \lambda_s(I_\ell\cap A)\lambda_s(I_{\ell'}\cap A),$$

$$\mathscr{V}_{s+1}(\ell,\ell') = \frac{n^{s(s-1)}}{(n^{s-1}-1)^{s}}\sum_{m_j\neq m'_j}\lambda_s(I_{\ell,m}\cap A)\lambda_s(I_{\ell',m'}\cap A) - \lambda_s(I_\ell\cap A)\lambda_s(I_{\ell'}\cap A).$$

And so

$$\sum_{\substack{\hat{\ell}_i=\hat{\ell}'_i\\ \ell_i\neq\ell'_i}}|\mathscr{V}_i(\ell,\ell')| \le$$

$$\frac{1}{n^{2s}}\#\{(\ell,\ell') : \hat{\ell}_i = \hat{\ell}'_i, \ell_i \neq \ell'_i,\ I_\ell \not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\},$$

$$\sum_{\hat{\ell}_j\neq\hat{\ell}'_j}|\mathscr{V}_{s+1}(\ell,\ell')| \le$$

$$\frac{s}{n^{3s-1}}\#\{(\ell,\ell') : \hat{\ell}_j \neq \hat{\ell}'_j,\ I_\ell \not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\}.$$

Consequently

$$\begin{aligned}\mathrm{Var}(Z) \;&\le \frac{1}{4n^{2s}}\#\{\ell : I_\ell \not\subset A \text{ and } I_\ell \cap A \neq \emptyset\}\\ &+\frac{1}{n^{2s}}\sum_{i=1}^{s}\#\{(\ell,\ell') : \hat{\ell}_i = \hat{\ell}'_i, \ell_i \neq \ell'_i,\ I_\ell \not\subset A,\ I_\ell \cap A \neq \emptyset,\ I_{\ell'} \not\subset A,\ I_{\ell'} \cap A \neq \emptyset\}\\ &+\frac{s}{n^{3s-1}}\#\{(\ell,\ell') : \hat{\ell}_j \neq \hat{\ell}'_j,\ I_\ell \not\subset A,\ I_\ell \cap A \neq \emptyset,\ I_{\ell'} \not\subset A,\ I_{\ell'} \cap A \neq \emptyset\}.\end{aligned}$$

Let us note

$$\hat{I}_{\ell,i} = \prod_{\substack{j=1\\ j\neq i}}^{s}\left[\frac{\ell_j - 1}{n}, \frac{\ell_j}{n}\right).$$

We have the following inferences:

- If $I_\ell \not\subset A$, there exists $\hat{u}_{\ell,i} \in \hat{I}_{\ell,i}$ such that $nf_i(\hat{u}_{\ell,i}) < \ell_i$,
- If $I_\ell \cap A \neq \emptyset$, there exists $\hat{v}_{\ell,i} \in \hat{I}_{\ell,i}$ such that $\ell_i < nf_i(\hat{v}_{\ell,i}) + 1$.

Hence

$$\begin{aligned}&\#\{\ell : I_\ell \not\subset A \text{ and } I_\ell \cap A \neq \emptyset\} \le n^{s-1}(k+2),\\ &\#\{(\ell,\ell') : \hat{\ell}_i = \hat{\ell}'_i, \ell_i \neq \ell'_i,\ I_\ell \not\subset A,\ I_\ell \cap A \neq \emptyset,\ I_{\ell'} \not\subset A,\ I_{\ell'} \cap A \neq \emptyset\} \le n^{s-1}(k+2)^2,\\ &\#\{(\ell,\ell') : \hat{\ell}_j \neq \hat{\ell}'_j,\ I_\ell \not\subset A,\ I_\ell \cap A \neq \emptyset,\ I_{\ell'} \not\subset A,\ I_{\ell'} \cap A \neq \emptyset\} \le n^{2(s-1)}(k+2)^2,\end{aligned}$$

and the result follows. □

The variance bound represents a gain in accuracy of the factor $N^{-1/s} = 1/n$ as compared with simple MC. The gain is of diminishing importance as s becomes large and limits the use of the present approach to problems of moderate dimension. This is precisely the case in some MC particle simulations, such as the random walk proposed in Sect. 3. A variance bound with the same order was established in [14]. The differences are as follows. Firstly a two-dimensional analysis in the context of the simulation of Markov chains was conducted in [14] and a possible generalization to higher-dimensional settings was discussed. Secondly the point set used in [14] was obtained by simple stratified sampling over the unit square, with one point in each subsquare $\prod_{i=1}^{2}[(\ell_i - 1)/n, \ell_i/n)$ (without the LHS property).

We use a simple example to illustrate the previous analysis. We consider the subset of the unit ball:

$$Q := \{u \in I^s : \|u\|_2 < 1\},$$

where $\|u\|_2$ denotes the Euclidean norm of u. In order to estimate the variance of the MC, SSS, LHS and ELHS approximations, we replicate the quadrature independently M times and compute the sample variance. We use $M = 100, \ldots, 1{,}000$ and

we only see small differences between the estimates. The results (for $M = 1{,}000$) are displayed in Fig. 2. It appears that the better accuracy due to ELHS goes beyond an improved convergence order: the slope of the curve given by the series of ELHS points is steeper than the slope of the corresponding curve for the MC or LHS points; in addition, the starting ELHS point is below the starting MC or LHS points. The computation times are given in the same figure; one can see that for obtaining the smallest variance achieved by usual MC or LHS, the ELHS approach needs less time. Assuming $\text{Var} = \mathcal{O}(N^{-\alpha})$, linear regression can be used to evaluate α and the outputs are listed in Table 1. The values obtained for ELHS are very close to the orders of the bounds given in Proposition 1, despite the fact that the hypothesis on the boundary of A is not satisfied, since the functions

$$f_i : \hat{u}_i \to \sqrt{1 - u_1^2 - \cdots - u_{i-1}^2 - u_{i+1}^2 - \cdots - u_s^2}$$

are not Lipschitz continuous on $\overline{I}^{s-1}$. This suggests that the hypothesis is too strong.

Table 1 Order α of the variance of the calculation of $\lambda_s(Q)$.

Dimension	MC	LHS	SSS	ELHS
$s = 2$	0.99	1.00	1.48	1.50
$s = 3$	1.00	1.00	1.34	1.33
$s = 4$	1.01	1.00	1.26	1.24

3 Simulation of Diffusion

In many physical applications, there is a need to simulate plain diffusion problems. These problems are frequently encountered as sub-problems while solving more complicated ones. MC simulation has proved a valuable tool for investigating processes involving the diffusion of substances [6,8,23]. In this section we consider a particle method for solving the initial value problem

$$\frac{\partial c}{\partial t}(x,t) = D\frac{\partial^2 c}{\partial x^2}(x,t), \quad x \in \mathbf{R},\ t > 0, \tag{8}$$

$$c(x,0) = c_0(x), \quad x \in \mathbf{R}, \tag{9}$$

with diffusion coefficient $D > 0$. We assume that the initial data satisfies

$$c_0 \geq 0, \quad \int_{\mathbf{R}} c_0(x)dx = 1. \tag{10}$$

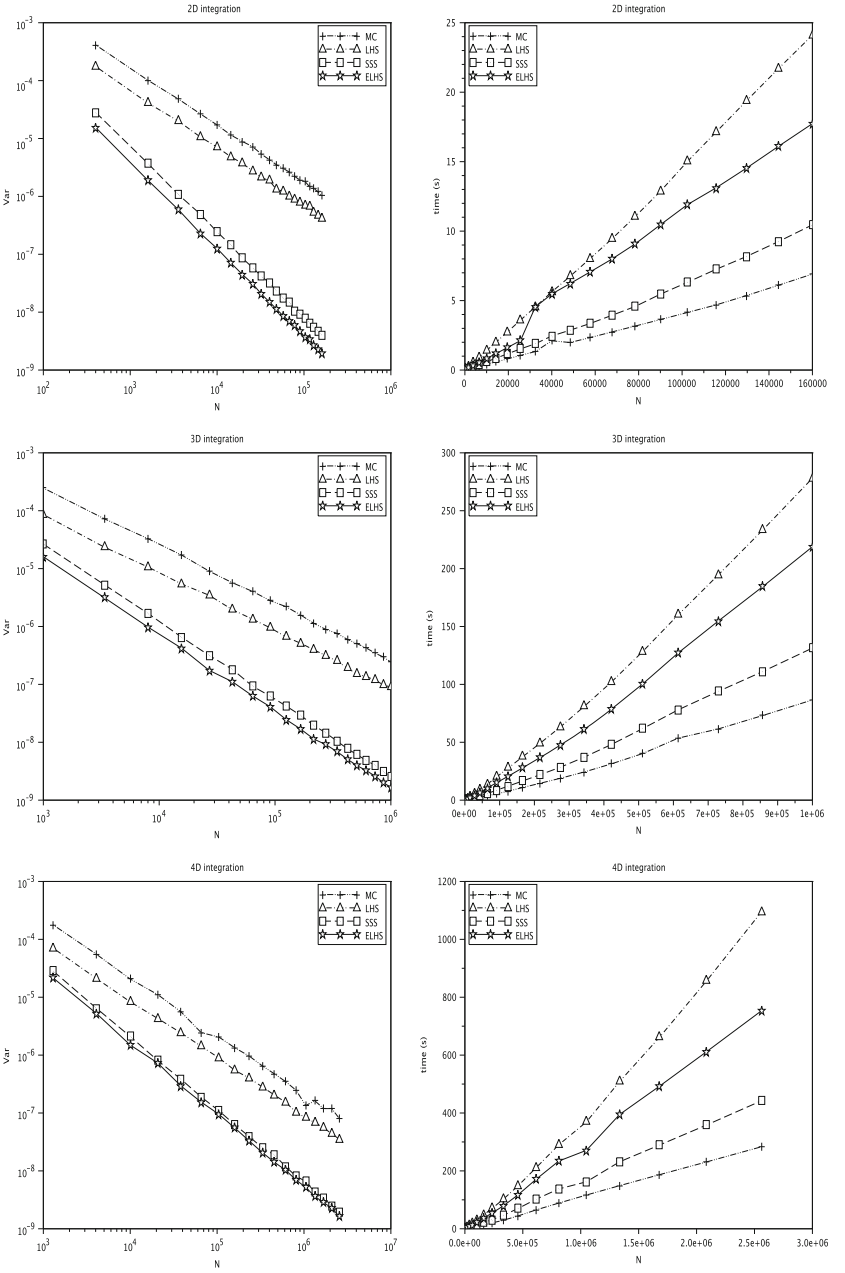

Fig. 2 Sample variance of $M = 1{,}000$ independent copies of the calculation of $\lambda_s(Q)$ as a function of N (*left*, log–log plot) and CPU time in seconds for the sample variance (*right*). Comparison of MC (+), LHS (△), SSS (□) and ELHS methods (⋆) outputs for $s = 2$ and $20^2 \leq N \leq 400^2$ (*top*), $s = 3$ and $10^3 \leq N \leq 100^3$ (*middle*), $s = 4$ and $6^4 \leq N \leq 40^4$ (*bottom*).

The solution possesses the conservation property

$$\forall t > 0 \quad \int_{\mathbf{R}} c(x,t)dx = 1. \tag{11}$$

The fundamental solution for the heat operator $\dfrac{\partial}{\partial t} - D\dfrac{\partial^2}{\partial x^2}$ is

$$E(x,t) := \frac{1}{\sqrt{4\pi Dt}} \mathrm{e}^{-x^2/4Dt}, \quad x \in \mathbf{R},\ t > 0.$$

For any $\tau \geq 0$ the solution of (8) satisfies

$$c(x,t) = \int_{\mathbf{R}} E(x-y, t-\tau)c(y,\tau)dy, \quad x \in \mathbf{R},\ t > \tau. \tag{12}$$

For the numerical approximation of the solution we choose an integer n and we put $N = n^2$. The first step of the simulation involves approximating the initial data u_0 with a sum of Dirac delta functions (particles),

$$c^0(x) := \frac{1}{N}\sum_{k=1}^{N} \delta(x - x_k^0).$$

One has to sample $x_1^0, \ldots, x_N^0$ according to the density function c_0; this may be done by inversion method

$$x_k^0 := C_0^{-1}\left(\frac{2k-1}{2N}\right), \quad 1 \leq k \leq N,$$

where C_0 is the cumulative distribution function associated with c_0. Let Δt be a time step, put $t_p := p\Delta t$ and $c_p(x) := c(x, t_p)$. Given particles at positions x_k^p and the approximate solution

$$c^p(x) := \frac{1}{N}\sum_{k=1}^{N} \delta(x - x_k^p)$$

at time t_p, the solution at time t_{p+1} is obtained as follows.

Generate an extended Latin hypercube sample, as is done in Sect. 2

$$\{W_\ell : 1 \leq \ell_1 \leq n, 1 \leq \ell_2 \leq n\} \subset I^2.$$

Relabel the particles. We order the particles by position:

$$x_1^p \leq x_2^p \leq \cdots \leq x_N^p. \tag{13}$$

This type of sorting was initiated in [11] and used in the context of simulation of diffusion in [12, 16]. Since each step of the random walk algorithm may be described by a numerical integration (see below), the sorting reverts to minimizing the amplitude of the jumps of the function to be integrated.

Diffusion of particles. Using (12), one obtains an approximation to the solution at time t_{p+1}:

$$\tilde{c}^{p+1}(x) := \frac{1}{N}\sum_{k=1}^{N} E(x - x_k^p, \Delta t).$$

Let

$$f(u) := \sqrt{2D\Delta t}\Phi^{-1}(u), \quad u \in (0,1),$$

where Φ denotes the standard normal cumulative distribution function. If $A \subset \mathbf{R}$, denote by 1_A the indicator function. For any measurable $A \subset \mathbf{R}$, one has

$$\int_{\mathbf{R}} \tilde{c}^{p+1}(x) 1_A(x) dx = \frac{1}{N}\sum_{k=1}^{N} \int_I 1_A(x_k^p + f(u)) du. \tag{14}$$

For $1 \le k \le N$, let 1_{I_k} denote the indicator function of $I_k := [(k-1)/N, k/N)$. We associate to any measurable $A \subset \mathbf{R}$ the following indicator function:

$$C_A^{p+1}(u) := \sum_{k=1}^{N} 1_{I_k}(u_1) 1_A(x_k^p + f(u_2)), \quad u = (u_1, u_2) \in I \times (0,1).$$

It is easy to verify that

$$\int_{\mathbf{R}} \tilde{c}^{p+1}(x) 1_A(x) dx = \int_{I^2} C_A^{p+1}(u) du. \tag{15}$$

We recover an approximate solution at time t_{p+1} by performing a MC quadrature using the extended Latin hypercube sample defined above: for any measurable $A \subset \mathbf{R}$

$$\int_{\mathbf{R}} 1_A(x) c^{p+1}(x) = \frac{1}{N}\sum_{\ell_1=1}^{n}\sum_{\ell_2=1}^{n} C_A^{p+1}(W_\ell).$$

The algorithm may be summarized as follows. Let $\lfloor x \rfloor$ denote the greatest integer $\le x$ and put $k(u) := \lfloor Nu \rfloor$. The positions of the particles are updated according to

$$x_{k(W_\ell^1)}^{p+1} = x_{k(W_\ell^1)}^{p} + f(W_\ell^2), \quad 1 \le \ell_1 \le n,\ 1 \le \ell_2 \le n. \tag{16}$$

For any $\ell := (\ell_1, \ell_2)$, the first projection W_ℓ^1 selects the particle number $k(W_\ell^1)$ and the second projection W_ℓ^2 gives the random displacement $f(W_\ell^2)$ of the selected

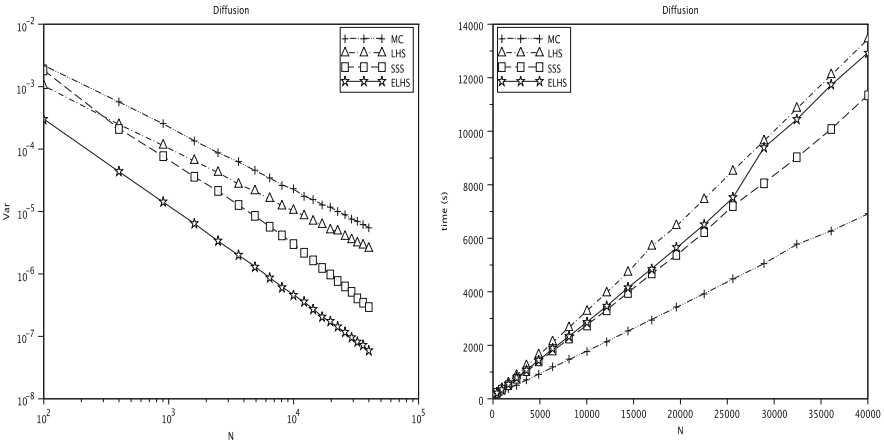

Fig. 3 Sample variance of $M = 5{,}000$ independent copies of the calculation of $\int_0^a c(x,T)dx$ as a function of N (*left*, log–log plot) and CPU time in seconds for the sample variance (*right*). Comparison of MC (+), LHS (△), SSS (□) and ELHS methods (⋆) outputs for $10^2 \leq N \leq 200^2$.

particle. In this algorithm, we may replace extended Latin hypercube samples with simple stratified samples or Latin hypercube samples. The classical random walk algorithm works as follows: there is no reordering of the particles and

$$x_k^{p+1} = x_k^p + f(U_k), \quad 1 \leq k \leq N. \tag{17}$$

Here $U_1, \ldots, U_N$ are independent random samples drawn from the uniform distribution on I.

We compare the approaches in a simple situation. We solve (8)–(9) with $D = 1.0$ and

$$c_0(x) := \frac{1}{\sqrt{\pi}} \mathrm{e}^{-x^2}, \quad x \in \mathbf{R}.$$

We approximate the integral

$$\int_0^a c(x,T)dx,$$

for $a = 4.0$ and $T = 1.0$. The time step is chosen to be $\Delta t := 1/100$. We replicate the computation independently $M = 5{,}000$ times to calculate the sample variance of the MC, SSS, LHS and ELHS approximations. The results are displayed in Fig. 3. As before (Sect. 2), the ELHS method produces better accuracy and improved convergence rate for the variance. The computation times are given in the same figure; one can see that, for the same calculation time, the ELHS technique has a

smaller variance than the other methods. If we assume Var $= \mathcal{O}(N^{-\beta})$, we can estimate β using linear regression; the outputs are listed in Table 2.

Table 2 Order β of the variance of the calculation of $\int_0^a c(x,T)dx$.

MC	LHS	SSS	ELHS
1.00	1.00	1.44	1.43

Although not reported, supplementary results lead to the following remarks. Firstly, it is useless to reorder the particles by position, when using the simple random walk algorithm (with ordinary pseudo-random numbers). Secondly, if we employ ELHS without reordering the particles, the variance is larger than the variance of the simple random walk, and the convergence order (estimated by linear regression) is the same.

4 Conclusion

We have analyzed an extended LHS technique that produces random points which are evenly spread over the unit cube. We have established that for approximate calculation of the measure of some subsets of the hypercube, the technique has a reduced variance, when compared to usual Monte Carlo, simple stratified sampling or Latin hypercube sampling, and a better convergence order.

Then we have modified the classical random walk method for simulation of diffusion. We reorder the particles by position in every time step, and we replace pseudo-random numbers with simple stratified samples, Latin hypercube samples or extended Latin hypercube samples. In an example, we have shown that the method using extended Latin hypercube samples produces lower variance with improved convergence order than the other strategies.

For approximate integration, the hypothesis made on the subsets of the unit hypercube could be relaxed. For the simulation procedure, a bound of the variance is not available: it certainly deserves future work. Another way of progress is in applications of the method to more complex diffusion problems [4] or to Markov chains, as it was done for QMC [3] or randomized QMC methods [13, 14].

References

1. Cheng, R.C.H., Davenport, T.: The problem of dimensionality in stratified sampling. Management Science **35**, 1278–1296 (1989)
2. El-Haddad, R., Fakhreddine, R., Lécot, C.: Stratified Monte Carlo integration. In: Sabelfeld, K.K., Dimov, I. (eds.) Monte Carlo Methods and Applications, pp. 105–113. De Gruyter, Berlin (2013)

3. El-Haddad, R., Lécot, C., L'Ecuyer, P.: Quasi-Monte Carlo simulation of discrete-time Markov chains on multidimensional state spaces. In: Keller, A., Heinrich, S., Niederreiter H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006, pp. 413–429. Springer, Berlin/Heidelberg (2008)
4. El-Haddad, R., Lécot, C., Venkiteswaran, G.: Quasi-Monte Carlo simulation of diffusion in a spatially nonhomogeneous medium. In: L'Ecuyer, P., Owen, A.B. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2008, pp. 339–354. Springer, Berlin/Heidelberg (2010)
5. Evans, M., Swartz, T.: Approximating Integrals via Monte Carlo and Deterministic Methods. Oxford University Press, Oxford (2000)
6. Farnell, L., Gibson, W.G.: Monte Carlo simulation of diffusion in a spatially nonhomogeneous medium: correction to the Gaussian steplength. J. Comput. Phys. **198**, 65–79 (2004)
7. Fishman, G.S.: Monte Carlo. Springer, New York (1996)
8. Ghoniem, A.F., Sherman, F.S.: Grid-free simulation of diffusion using random walk methods. J. Comput. Phys. **61**, 1–37 (1985)
9. Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer, New York (2004)
10. Haber, S.: A modified Monte-Carlo quadrature. Math. Comp. **20**, 361–368 (1966)
11. Lécot, C.: A Direct Simulation Monte Carlo scheme and uniformly distributed sequences for solving the Boltzmann equation. Computing **41**, 41–57 (1988)
12. Lécot, C., El-Khettabi, F.: Quasi-Monte Carlo simulation of diffusion. J. Complexity **15**, 342–359 (1999)
13. 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, pp. 329–343. Springer, Berlin/Heidelberg (2004)
14. L'Ecuyer, P., Lécot, C., Tuffin, B.: A randomized quasi-Monte Carlo simulation method for Markov chains. Oper. Res. **56**, 958–975 (2008)
15. McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239–245 (1979)
16. Morokoff, W.J., Caflisch, R.E.: A Quasi-Monte Carlo approach to particle simulation of the heat equation. SIAM J. Numer. Anal. **30**, 1558–1573 (1993)
17. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)
18. Owen, A.B.: Orthogonal arrays for computer experiments, integration and visualization. Statist. Sinica **2**, 439–452 (1992)
19. Owen, A.B.: Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. Ann. Statist. **22**, 930–945 (1994)
20. Owen, A.B.: Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal. **34**, 1884–1910 (1997)
21. Stein, M.: Large sample properties of simulations using Latin hypercube sampling. Technometrics **29**, 143–151 (1987)
22. Tang, B.: Orthogonal array-based Latin hypercubes. J. Amer. Statist. Assoc. **88**, 1392–1397 (1993)
23. Venkiteswaran, G., Junk, M.: A QMC approach for high dimensional Fokker-Planck equations modelling polymeric liquids. Math. Comput. Simulation **68**, 45–56 (2005)