IEICE The Institute of Electronics, Information and Communication Engineers

IEICE Proceeding Series

Super-efficiency and Approximate Super-efficiency in Monte Carlo Simulation Methodology

Cheng-An Yang, Kung Yao, Ken Umeno, Ezio Biglieri

Vol. 2 pp. 213-216
Publication Date: 2014/03/18
Online ISSN: 2188-5079

Downloaded from www.proceeding.ieice.org

©The Institute of Electronics, Information and Communication Engineers

Super-efficiency and Approximate Super-efficiency in Monte Carlo Simulation Methodology

Cheng-An Yang†, Kung Yao†, Ken Umeno‡, Ezio Biglieri§

†Electrical Engineering Department, UCLA, Los Angeles, United States
‡Graduate School of Informatics, Kyoto University, Kyoto, Japan
§Departament TIC, Universitat Pompeu Fabra, Barcelona, Spain
Email: {rhymer123,yao}@ucla.edu, umeno.ken.8z@kyoto-u.ac.jp, e.biglieri@ieee.org

Abstract—Monte Carlo (MC) simulation methods are widely used to solve complex engineering and scientific problems. Unlike other deterministic methods, MC produce solution with uncertainty through statistical sampling. As the sample size N growths, the uncertainty of the solution reduces. It is well-know that the variance of the error decreases as $1/N$. However, for large problems like high-dimensional integrations and computationally intensive simulations this can take months or even years to obtain the solution with acceptable tolerance. The Super-efficient Monte Carlo simulation method, originated by Umeno, produced a solution that converges as fast as $1/N^2$. However it only applies to a small class of problems that possess certain properties. We proposed an Approximate Super-efficient Monte Carlo simulation method that is applicable to a wider class of problems than the original Umeno method, where the convergence rate is as fast as $1/N^\alpha$ for $1 \le \alpha \le 2$.

1. Introduction

Ulam and von Neumann first formulated the Monte Carlo (MC) simulation methodology using random sequence to evaluate high-dimensional integrals [1]. Since then MC simulations have been used in myriads of applications to evaluate performances of various systems that are not analytically tractable.

The simplest form of MC simulation draws independent samples from a distribution $\rho(x)$ to approximate the integration. More specifically, if we want to evaluate the integral

$$I = \int_\Omega A(x)\,dx, \tag{1}$$

where the integrand $A(x)$ is on the domain $\Omega \subset \mathbb{R}^n$, we first define a function

$$B(x) := \frac{A(x)}{\rho(x)}, \tag{2}$$

and approximate (1) by calculating the N-sample average

$$\frac{1}{N}\sum_{i=1}^{N} B(X_i) \approx \mathbf{E}[B(X_1)] = I. \tag{3}$$

Here, N is the sample size, X_i's are random variables with a common probability density function (pdf) $\rho(x)$, and $\mathbf{E}[.]$ is the expectation operator.

The summation (3) converges almost surely to I if the samples X_i's are independent by the Law of Large Numbers. Furthermore, the variance of the approximation decreases to zero at the rate of $1/N$. That is,

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} B(X_i)\right] = \frac{1}{N}\mathrm{Var}\left[B(X_1)\right]. \tag{4}$$

Generating truly random sequences in a controlled manner is non-trivial. In practice, we generate pseudo-random (PR) sequences instead. A PR sequence is generated deterministically by some transformations like linear congruential recursion [1]. They appear random from the statistical point of view.

2. Chaotic MC Simulation

The chaotic MC simulation is the MC simulation with PR sequence replaced by a chaotic sequence [2]. More specifically, consider an ergodic dynamical system $(\Omega, \mathcal{A}, \rho, T)$, where $\mathcal{A}$ is the σ-algebra on Ω, $\rho(x)\,dx$ is the invariant measure and T is a measure-preserving automorphism on Ω [3]. The chaotic sequence $(x_1, x_2, x_3, \dots)$ with seed $x_0 \in \Omega$ under T is generated by

$$x_{i+1} = T(x_i), \text{ for } i = 0, 1, 2, \dots. \tag{5}$$

The chaotic MC simulation approximates the integration (1) by the "time-average"

$$\langle B(x_i)\rangle_N := \frac{1}{N}\sum_{i=1}^{N} B(x_i). \tag{6}$$

If the modified integrand $B \in L_1(\Omega, \rho)$ is Lebesgue integrable, then the approximation (6) converges to I almost surely by ergodicity [3].

2.1. Statistical and Dynamical Correlation

The greatest distinction between conventional and chaotic MC simulation is that the chaotic sequence has correlation between samples. It turned out that the correlation

has tremendous impact on the convergence rate of the approximation.

We can view $\{B_n | B_n = B(T^n x_0), x_0 \in \Omega\}_{n\in\mathbb{N}}$ as a stationary random process on $(\Omega, \mathcal{A}, \rho)$. Denote its autocorrelation function by $R(k) = \mathbf{E}[(B_{k+1} - I)(B_1 - I)]$ and denote $\langle B(x_i)\rangle_N$ by $\langle B\rangle_N$ for simplicity. The variance of the approximation error $\langle B\rangle_N - I$ is

$$\sigma_N^2 := \mathbf{E}\Big[(\langle B\rangle_N - I)^2\Big] = \mathbf{E}\Big[(\frac{1}{N}\sum_{i=1}^{N}(B_i - I)^2\Big] \tag{7}$$

$$= \frac{1}{N^2}\,\mathbf{E}\Big[\sum_{i=1}^{N}(B_i - I)^2 + 2\sum_{i>j}(B_i - I)(B_j - I)\Big] \tag{8}$$

$$= \frac{1}{N}\mathrm{Var}\,[B] + \frac{2}{N^2}\sum_{i>j} R(i-j) \tag{9}$$

$$= \frac{1}{N}\mathrm{Var}\,[B] + \frac{2}{N^2}\sum_{k=1}^{N}(N-k)R(k). \tag{10}$$

The first term in (10) is called the *statistical correlation*, which depends on the integrand B and the invariant measure $\rho(x)\,dx$. The second term is called the *dynamical correlation*, which depends on the integrand as well as the chaotic sequence [2].

2.2. Super-Efficient Chaotic MC Simulation

Rewrite the variance of the approximation error (7) as

$$\sigma_N^2 = \frac{1}{N}\underbrace{\left(\mathrm{Var}\,[B] + 2\sum_{k=1}^{N} R(k)\right)}_{\eta} - \frac{2}{N^2}\sum_{k=1}^{N} kR(k). \tag{11}$$

This shows that the convergence rate of σ_N^2 has two contributors, one decaying as $1/N$ and the other as $1/N^2$. The asymptotic convergence rate for the chaotic MC simulation is dominated by $1/N$, which has the same performance as the standard MC simulation. However, if the dynamic correlation makes the term $\eta = 0$, the convergence rate becomes $1/N^2$. We say the chaotic MC simulation is *Super-Efficient* (SE) if the variance of the approximation error decays as $1/N^2$ for $N \to \infty$. Clearly that the necessary and sufficient condition for super-efficiency is $\eta = 0$ [2].

2.3. Theoretical Basis for Super-Efficient MC Simulation

Umeno observed that some integrands under chaotic mapping make the chaotic MC simulation super-efficient [2]. We call these integrands Super-Efficient (SE) under suitable chaotic mapping.

It is not clear how super-efficiency leads to a practical algorithm until Yao [4] saw the connection between Super-Efficiency and the Lebesgue spectrum of ergodic theory [3]. This observation helps us explain the super-efficiency systematically and hopefully leads to practical algorithms as detailed in the next section.

Let Λ and F be index sets. A dynamic system is said to have Lebesgue spectrum if there exists an orthogonal basis $\{1\} \cup \{f_{\lambda,j} | \lambda \in \Lambda, j \in F\}$, where λ labels the classes and j labels the functions within each class such that

$$f_{\lambda,j} \circ T = f_{\lambda,j+1}, \tag{12}$$

for all λ and j. That is, within each class λ, the operation of T on $f_{\lambda,j}$ is invariant in the same class. The cardinality of Λ is called the multiplicity of the Lebesgue spectrum.

Since $\{1\} \cup \{f_{\lambda,j}\}$ is a complete orthogonal basis, every function in $L^2(\Omega, \rho)$ can be represented as the generalized Fourier series

$$B(x) = b_0 + \sum_{\lambda\in\Lambda}\sum_{j=0}^{\infty} b_{\lambda,j} f_{\lambda,j}(x), \tag{13}$$

where b_0 is the coefficient corresponds to the constant basis function 1. It is clear that $\mathbf{E}[B] = b_0 = I$ is the desired integral (1).

Substitute (11) by (13), we find

$$\eta = \sum_{\lambda\in\Lambda}\left(\sum_{j\geq 0} b_{\lambda,j}\right)^2 = 0. \tag{14}$$

Therefore the explicit condition for super-efficiency is that the sum of coefficients in each class λ be zero, that is

$$d_\lambda := \sum_{j=0}^{\infty} b_{\lambda,j} = 0 \text{ for all } \lambda \in \Lambda\,. \tag{15}$$

The Chebyshev dynamic system $(\Omega, \mathcal{A}, \rho, T_p)$ has Lebesgue spectrum [3]. The chaotic mapping T_p is the Chebyshev polynomial with prime degree p, defined on $\Omega = [-1, 1]$ as

$$T_p(x) = \cos(p \arccos(x)), \tag{16}$$

and the invariant measure is

$$\rho(x)\,dx = \frac{1}{\pi\sqrt{1-x^2}}\,dx. \tag{17}$$

The basis function is given by

$$f_{\lambda,j}(x) = T_{\lambda p^j}(x),\ \forall\lambda \in \Lambda, j \in F, \tag{18}$$

where Λ is the set of non-negative integer relatively prime to p and $F = \{0, 1, 2, \dots\}$. Note that $\lambda = 0$ corresponds to the class with a single function $T_0(x) = 1$.

Example 2.1. Consider the integrand [2, p. 1447]

$$A(x) = \frac{-8x^4 + 8x^2 + (1+\epsilon)x - 1}{\pi\sqrt{1-x^2}} = B_\epsilon(x)\rho(x). \tag{19}$$

Under the Chebyshev dynamical system $(\Omega, \mathcal{A}, \rho, T_p)$, $B_\epsilon(x)$ can be expanded as

$$B_\epsilon(x) = (1+\epsilon)T_1(x) - T_4(x). \tag{20}$$

If $p = 2$, the coefficients of the generalized Fourier series are $b_{1,0} = 1 + \epsilon$, $b_{1,2} = -1$ and zero otherwise. The sum of coefficients are $d_1 = \epsilon$ and $d_\lambda = 0$ for $\lambda \neq 1$. Therefore $A(x)$ is super-efficient if and only if $\epsilon = 0$. When $\epsilon \neq 0$ we have "mismatched" SE MC simulation, which appears to be super-efficient for small N but gradually loses super-efficiency as N increases [5]. See Fig. 1.

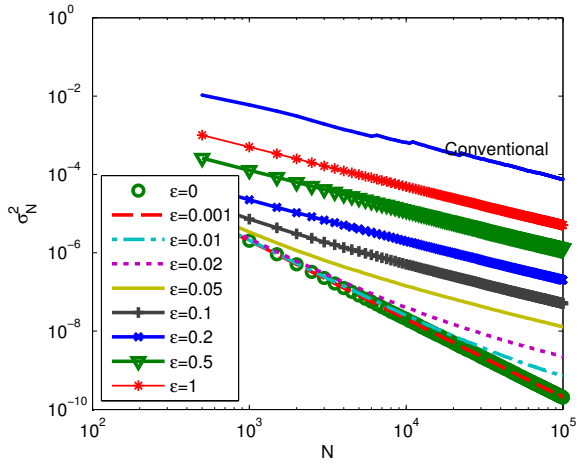

Figure 1: The variance of the approximation error σ_N^2 versus the number of samples N. The slope of the conventional MC simulation curve is -1, indicating its $1/N$ behaviour. On the other hand, the slope of the super-efficient MC simulation is -2 because σ_N^2 decays like $1/N^2$. Between these two extremes are the mismatched SE MC simulations with different size of ϵ. For $\epsilon = 0.001$, the curve is almost identical to the super-efficient curve. As ϵ becomes larger, the slope of the mismatched SE MC simulations gradually increase as N becomes larger.

3. Approximate SEMC Simulation

The collection of SE integrands can be seen as an affine subspace of $L^2(\Omega, \rho)$. Clearly most integrands are not SE. This implies chaotic MC simulation has no advantage over conventional MC simulation in general.

While most integrands do not satisfy the SE condition, Yao proposed the Approximate Super-Efficient (ASE) algorithm [4] that modifies the integrand so that it is approximately SE, and by applying chaotic MC simulation on the modified integrand, we get a much faster convergence rate of $1/N^\alpha$ for α between 1 to 2.

A crucial observation here is adding a function that has zero mean to $B(x)$ will not change the integral of $B(x)$ [6]. In [6], a concept equivalent to ASE was proposed by Umeno in 2001. Therefore, if we know the sum of coefficients d_λ in each class λ, then the new integrand

$$B'(x) = B(x) - \sum_{\lambda \in \Lambda} d_\lambda f_{\lambda,0}(x) \tag{21}$$

will be super-efficient without changing the integral of $B(x)$ (recall the basis functions $f_{\lambda,0}(x)$'s have zero mean). We call the function $d_\lambda f_{\lambda,0}(x)$ the compensator associate with class λ. By subtracting compensators from $B(x)$, we introduce negative dynamical correlation and makes the chaotic MC simulation nearly super-efficient.

In practice we do not know the sum of coefficients, and it is not possible to construct infinitely many compensators to achieve perfect super-efficiency. The idea of ASE algorithm is to approximate the sum of coefficients d_λ by its L_λ-term partial sum

$$\hat{d}_\lambda \approx b_{\lambda,0} + b_{\lambda,1} + \cdots + b_{\lambda,L_\lambda} \tag{22}$$

using conventional MC or chaotic MC simulations, where L_λ is some hopefully not too large positive integer. Then we form the modified integrand

$$\tilde{B}(x) = B(x) - \sum_{\lambda \in \Lambda_L} \hat{d}_\lambda f_{\lambda,0}(x), \tag{23}$$

where the index set Λ_L contains L classes. If the sum of coefficients $\tilde{d}_\lambda = d_\lambda - \hat{d}_\lambda$ of $\tilde{B}$ is close to zero and $\tilde{\eta} = \sum_{\lambda \in \Lambda} \tilde{d}_\lambda = \epsilon > 0$ is small, then from (11) the variance of the approximation error can be written as

$$\sigma_N^2 = \frac{\epsilon^2}{N} + \frac{\zeta}{N^2} \tag{24}$$

for some ζ. The effective convergence rate can be expressed as $1/N^\alpha$ for $\alpha \in [1, 2]$, depending on the accuracy of $\hat{d}_\lambda' s$.

The procedure of the ASE algorithm is as follows:

1. Approximate the sum of coefficients $\hat{d}_\lambda$'s in (22) using conventional or chaotic MC simulation for each $\lambda \in \Lambda_L$.

2. Subtract the compensators from the integrand $B(x)$ to form $\tilde{B}(x)$ as defined in (23), and apply chaotic MC simulation on $\tilde{B}(x)$.

ASE simulation is approximately super-efficient for moderate size of N, however, from (24) and Example 2.1 it is clear that ASE simulation will eventually lose the $1/N^2$ convergence rate as long as $\epsilon \neq 0$.

Biglieri suggested computing $\hat{d}_\lambda$'s iteratively to improve the accuracy of the estimation [5]. As opposed to the original ASE algorithm, which has fixed accuracy for the entire simulation, we proposed a Progressive ASE algorithm that keeps improving the accuracy of $\hat{d}_\lambda$'s as the chaotic MC simulation goes on. The idea is to use the samples $B(x_i)$ generated in the main chaotic MC simulation to estimate $\hat{d}_\lambda$'s continuously. Therefore we get progressively better estimates of $\hat{d}_\lambda$'s and improve the decay rate.

Example 3.1. Consider the Chebyshev dynamical system $(\Omega, \mathcal{A}, \rho, T_2)$ and the integrand [5]

$$A(x) = (1 - x^2)\exp(-x^2) = B(x)\rho(x). \tag{25}$$

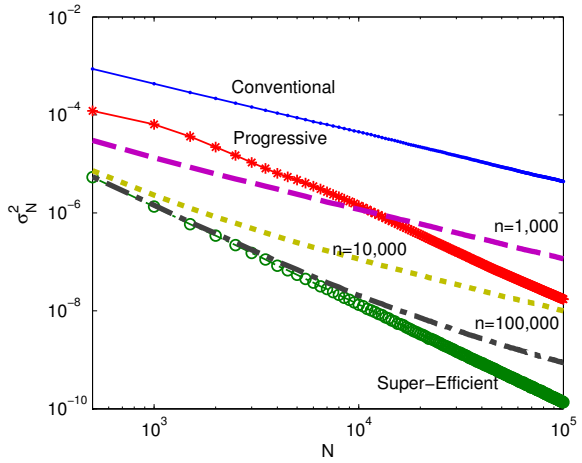

Figure 2: The variance of approximation error σ_N^2 versus the number of samples N. Throughout the entire simulation, both the conventional and super-efficient MC simulation constantly has $1/N$ and $1/N^2$ behavior respectively. The ASE MC simulations have $1/N^2$ behavior at first but gradually degraded to $1/N$. ASE simulations with larger number of n have better accuracy of the estimates $\hat{d}_\lambda$'s and lose super-efficiency later. The Progressive ASE simulation has $1/N$ behavior at first but gradually improves to $1/N^2$, because the estimates $\hat{d}_\lambda$'s get more accurate as N increases.

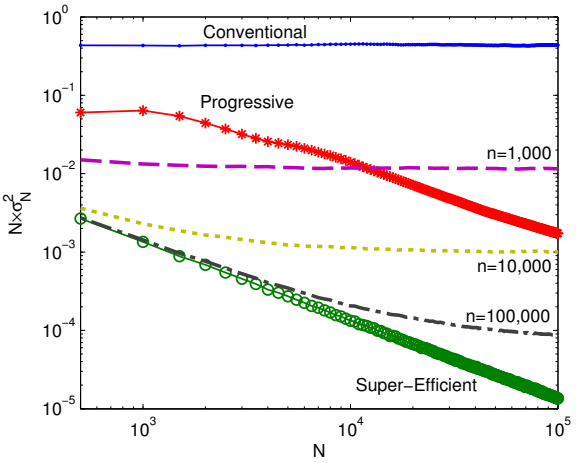

Figure 3: To see the decay exponent α more clearly, we plot $N\sigma_N^2$ versus the number of samples N. The $N\sigma_N^2$ curve for the conventional MC simulation is flat, because η is large and it dominates the convergence rate. The curve for the super-efficient MC simulation has negative slope because η is so small that the remaining $1/N$ term dominates. For ASE MC simulations with sample size $n = 1,000$ and $n = 10,000$, the super-efficiency is lost before $N < 10^3$. For $n = 100,000$, it is super-efficient up to around $N = 10^4$. The Progressive ASE simulation gains super-efficiency after $N = 10^4$.

Unlike the previous example, the modified integrand $B(x)$ has infinitely many terms in the generalized Fourier series expansion. We choose L and L_λ such that the the degree of $f_{\lambda,L_\lambda}(x)$ is at most 10. We perform chaotic MC simulation for 10^5 samples using conventional, ASE and Progressive ASE MC algorithms, see Fig.2. As a benchmark, we compute the sum of coefficients using accurate numerical integration for the super-efficient case (marked by circle). For ASE MC simulations, we use different number of samples n to estimate d_λ's to demonstrate the effect of inaccurate estimates and convergence rate. For Progressive ASE MC simulation we estimate d_λ's at the same time as the chaotic MC simulation runs.

To better visualize the decay exponent α, we plot $N\sigma_N^2$ versus N and observe its slope. From (11), if the integrand is nearly super-efficient, then the slope of $N\sigma_N^2$ will be negative. Otherwise it would be flat. See Fig.3.

4. Conclusions and Future Works

While conventional MC simulation yields the convergence rate of $1/N$, SE MC has superior convergence of $1/N^2$ for integrands of the SE type. Since most integrands are not SE, we introduce the concept of ASE. The ASE and Progressive ASE algorithms are at least as fast as conventional MC simulation and sometimes they yield near super-efficient convergence rate. Furthermore, we use the Lebesgue spectrum of the Ergodic theory to systematically study the SE MC simulation. The above discussions are applicable to multi-dimensional integrands. It is of great interest to find more applications to exploit the concept of SE and ASE.

References

[1] R. Y. Rubinstein, *Simulation and the Monte Carlo method*. Wiley series in Probability and Mathematical Statistics, Wiley, 1981.

[2] K. Umeno, "Chaotic Monte Carlo Computation: A Dynamical Effect of Random-Number Generations," *Japanese Journal of Applied Physics*, vol. 39, pp. 1442–1456, Mar. 2000.

[3] V. I. Arnol'd, *Ergodic problems of classical mechanics*, vol. 50. New York: Benjamin, 1968.

[4] K. Yao, "An Approximate Superefficient MC Method is Better Than Classical MC Method." Unpublished Report, 2009.

[5] E. Biglieri, "Some notes on "superefficient" Monte Carlo methods." Unpublished Report, 2009.

[6] K. Umeno, Japanese Patent No. 3711270 (submitted on March 30, 2002, and granted on August 19, 2005), recorded in Japanese.