IMS Collections
Borrowing Strength: Theory Powering Applications –
A Festschrift for Lawrence D. Brown
Vol. 6 (2010) 234–248

DOI: 10.1214/10-IMSCOLL616

Variance reduction via basis expansion in Monte Carlo integration

Yazhen Wang[1]

University of Wisconsin-Madison

Abstract: Monte Carlo methods are widely used in numerical integration, and variance reduction plays a key role in Monte Carlo integration. This paper investigates variance reduction for Monte Carlo integration in both finite dimensional Euclidean space and infinite dimensional Wiener space. The proposed variance reduction approaches are to use basis functions to construct control variates for finite dimensional integrals and utilize Itô-Wiener chaos expansion to design antithetic variates and control variates for Wiener integrals. We establish the variances of the proposed Monte Carlo integration estimators and show that the proposed methods can achieve dramatic variance reduction in comparison with the basic Monte Carlo estimators. Examples are used to illustrate the performance of the proposed estimators.

Contents

1. Introduction

Numerical integration has wide applications and good numerical integration schemes are always in high demand. There are many deterministic quadrature formulas for computing ordinary integrals with well behaved integrands. However, if the integrands fail to be regular such as the lack of continuous derivaive of moderate order, numerical analytic techniques, such as the trapezoidal and Simpson's rules, become less attractive. In particular the deterministic methods will run into several difficulties when applying to high dimensional integrals. For dimension $d = 1$,

*Wangs research was partially supported by NSF grant DMS-1005635.

[1]Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA

Keywords and phrases: antithetic variates, control variates, estimator, Itô-Wiener chaos expansion, orthonormal basis, simulation, Wiener process.

AMS 2000 subject classifications: Primary 65C05; secondary 62G05, 65C30.

standard deterministic integration techniques have very good accuracy for smooth integrand f. Their accuracy decreases as dimension d increases. In general for a deterministic integration method with errors $O(n^{-r})$ in one dimension, where n is data points used and r is some number usually no less than 1, the errors become $O(n^{-r/d})$ in d dimensions. At high dimensions, these deterministic integration methods are computationally infeasible or insufficiently accurate. It is often more convenient to compute such high dimensional integrals by Monte Carlo methods. The Monte Carlo approach is to represent an integral as an expectation of some random variable and then convert the integration problem into the problem of estimating a population mean. Monte Carlo methods are simple and can be used in any dimensions. The Monte Carlo estimators have errors of order $n^{-1/2}$ in probability, which is free of dimension d, although error magnitudes may deteriorate as d increases. Thus, it is very important to reduce the variances of the Monte Carlo estimators.

Variance reduction may be viewed as a means to control the variability of the Monte Carlo estimators by the use of known information about the problem. If we know nothing about the problem, variance reduction can not be achieved, while at the other extreme, if we have complete knowledge about the problem, the variance is equal to zero and there is no need for simulation. Variance reduction is often obtained from the clever use of available information about the integration problem. There exist several variance reduction techniques including antithetic variates, control variates, importance sampling, and stratified sampling. Some of these techniques may require pilot simulations, which are used to define variance reduction techniques that will refine and improve the efficiency of the whole simulations. See [2, 7].

This paper will study variance reduction for Monte Carlo integration in finite dimensional Euclid space and infinite dimensional Wiener space. Our approaches rely on basis expansion in finite dimensions and Itô-Wiener chaos expansion in Wiener space. For the finite dimensional case, we use basis functions to construct control variates for achieving variance reduction; and for Wiener integrals, we make use of Itô-Wiener chaos expansion together with orthonormal bases to design antithetic variates and control variates for variance reduction. We derive the variances of the proposed Monte Carlo estimators and demonstrate the variance reduction effects.

The rest of the paper proceeds as follows. Section 2 presents the construction of control variates by using basis functions for variance reduction in Monte Carlo integration in finite dimensions. Both orthonormal bases and a dictionary of bases are considered. Section 3 features Monte Carlo integration in Wiener space. Using Itô-Wiener chaos expansion we design schemes to construct antithetic variates and control variates for variance reduction in Monte Carlo Wiener integration. Section 4 provides two simple examples to illustrate the performance of the proposed methods.

2. Variance reduction by basis expansion

The problem we consider is to evaluate integral

$$I = \int_\Omega f(x)\,\mu(dx), \tag{2.1}$$

where μ is a probability measure on $\Omega = [0,1]^d$, and $f(x)$ is a function on Ω whose integral is analytically intractable. Standard manipulation can be applied to express integrals over domains other than the unit cube in (2.1), and integrand f in (2.1) may subsume weighting functions from importance sampling or periodization. Since all integrals will be on Ω, we will suppress Ω from integral signs.

Monte Carlo methods estimate integral I by the form

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} f(X_i), \tag{2.2}$$

where $X_1, \ldots, X_n$ are n points sampled in Ω according to the probability measure $\mu(dx)$.

2.1. Variance reduction by control variates

Suppose that $h(x)$ is an easily working function with known integral say $\int h(x) \times \mu(dx) = 0$ (if not, we replace $h(x)$ by $h(x) - \int h(x)\,\mu(dx)$). We decompose $f(x)$ into two orthogonal parts:

$$f(x) = \theta\, h(x) + f_h(x),$$

where h and f_h are orthogonal, that is,

$$\int h(x)\, f_h(x)\, \mu(dx) = 0,$$

and

$$\theta = \int f(x)\, h(x)\, \mu(dx).$$

For constant a, we define

$$\tilde{I} = \hat{I} - \frac{1}{n} \sum_{i=1}^{n} a\, h(X_i) = \frac{1}{n} \sum_{i=1}^{n} \left\{ f(X_i) - a\, h(X_i) \right\}.$$

Then

$$E(\tilde{I}) = E\left\{ f(X_1) - a\, h(X_1) \right\} = \int f(x)\, \mu(dx) - a \int h(x)\, \mu(dx) = I,$$

so $\tilde{I}$ is an unbiased estimator of I. The variance of $\tilde{I}$

$$\begin{aligned} Var(\tilde{I}) &= \frac{1}{n} Var\left\{ f(X_1) - a\, h(X_1) \right\} \\ &= \frac{1}{n} \int \left\{ f(x) - a\, h(x) - I \right\}^2 \, \mu(dx) \\ &= \frac{1}{n} (\theta - a)^2 \int [h(x)]^2 \, \mu(dx) + \frac{1}{n} \int \left\{ f_h(x) - I \right\}^2 \, \mu(dx). \end{aligned}$$

Hence, if

$$a = \theta = E\{f(X)\, h(X)\},$$

$\tilde{I}$ has the smallest variance. In order to make the control variates method practical, the integral of $h(x)$ needs to be either known or can be easily evaluated numerically. We will investigate the construction of such $h(x)$ by basis functions.

2.2. Construction of control variates by orthonormal basis

We start with an orthonormal system ψ_ℓ, $\ell = 1, 2, \ldots,$ on $L^2(\Omega, \mu(dx))$, which satisfy

$$\int [\psi_\ell(x)]^2 \, \mu(dx) = 1, \qquad \int \psi_\ell(x) \, \mu(dx) = 0,$$
$$\int \psi_\ell(x) \, \phi_j(x) \, \mu(dx) = 0, \qquad \ell \neq j.$$

We select k basis functions ψ_{ℓ_j}, $j = 1, \ldots, k$, and let $\phi_j = \psi_{\ell_j}$, $j = 1, \ldots, k$. The selected basis functions ϕ_j are used to construct $h(x)$ in the control variates method, that is,

$$f(x) = \theta \, h(x) + f_h(x), \qquad h(x) \in span\{\phi_1, \ldots, \phi_k\}, \qquad f_h(x) \perp span\{\phi_1, \ldots, \phi_k\}. \tag{2.3}$$

Let

$$\theta \, h(x) = \sum_{j=1}^{k} \theta_j \, \phi_j(x), \tag{2.4}$$

where

$$\theta_j = \int f(x) \, \phi_j(x) \, \mu(dx). \tag{2.5}$$

For constants $a_1, \ldots, a_k$, define

$$\tilde{I} = \hat{I} - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} a_j \, \phi_j(X_i) = \frac{1}{n} \sum_{i=1}^{n} \left\{ f(X_i) - \sum_{j=1}^{k} a_j \, \phi_j(X_i) \right\}. \tag{2.6}$$

Then

$$E(\tilde{I}) = E \left\{ f(X_1) - \sum_{j=1}^{k} a_j \, \phi_j(X_1) \right\} = \int f(x) \, \mu(dx) - \sum_{j=1}^{k} a_j \int \phi_j(x) \, \mu(dx) = I,$$

so $\tilde{I}$ is an unbiased estimator of I. The variance of $\tilde{I}$

$$\begin{aligned}
Var(\tilde{I}) &= \frac{1}{n} Var \left\{ f(X_1) - \sum_{j=1}^{k} a_j \, \phi_j(X_1) \right\} \\
&= \frac{1}{n} \int \left\{ f(x) - \sum_{j=1}^{k} a_j \, \phi_j(x) - I \right\}^2 \mu(dx) \\
&= \frac{1}{n} \int \left\{ \theta \, h(x) - \sum_{j=1}^{k} a_j \, \phi_j(x) \right\}^2 \mu(dx) + \frac{1}{n} \int \{f_h(x) - I\}^2 \, \mu(dx) \\
&= \frac{1}{n} \int \left\{ \sum_{j=1}^{k} (\theta_j - a_j) \, \phi_j(x) \right\}^2 \mu(dx) + \frac{1}{n} \int \{f_h(x) - I\}^2 \, \mu(dx) \\
&= \frac{1}{n} \sum_{j=1}^{k} (\theta_j - a_j)^2 + \frac{1}{n} \int \{f_h(x) - I\}^2 \, \mu(dx).
\end{aligned}$$

Hence, if

$$a_j = \theta_j = \int_\Omega f(x)\,\phi_j(x) = E\{f(X_1)\,\phi_j(X_1)\},$$

$\tilde{I}$ has the smallest variance. As θ_j are also integrals, we estimate θ_j by Monte Carlo simulation as follows. We generate a pilot sample $Y_1, \ldots, Y_m$ from $\mu(dx)$ and use the sample to estimate θ_j by

$$\tilde{\theta}_j = \frac{1}{m}\sum_{i=1}^{m} f(Y_i)\,\phi_j(Y_i). \tag{2.7}$$

Take $a_j = \tilde{\theta}_j$, plug them into $\tilde{I}$ defined by (2.6), and denote the resulting estimator by

$$\hat{I}^* = \hat{I} - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(X_i) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(X_i)\right\}. \tag{2.8}$$

The following theorem gives the variance of Monte Carlo estimator $\hat{I}^*$.

Theorem 2.1.

$$Var(\hat{I}^*) = \frac{1}{n}\left\{\int \{f_h(x) - I\}^2\,\mu(dx) + \sum_{j=1}^{k} Var(\tilde{\theta}_j)\right\},$$

where

$$Var(\tilde{\theta}_j) = \frac{1}{m}\sum_{j=1}^{k}\int \{f(y)\,\phi_j(y) - \theta_j\}^2\mu(dy).$$

Proof. Let $\mathbf{Y} = (Y_1, \ldots, Y_m)$ and note that

$$Var(\hat{I}^*) = E\{Var(\hat{I}^*|\mathbf{Y})\} + Var\{E(\hat{I}^*|\mathbf{Y})\}. \tag{2.9}$$

Simple calculations show

$$\begin{aligned} E(I^*|\mathbf{Y}) &= E(\hat{I}|\mathbf{Y}) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\tilde{\theta}_j\,E\{\phi_j(X_i)|\mathbf{Y})\} \\ &= E(\hat{I}) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\tilde{\theta}_j\,E\{\phi_j(X_i)\} = I, \end{aligned}$$

and thus,

$$Var\{E(\hat{I}^*|\mathbf{Y})\} = Var(I) = 0. \tag{2.10}$$

Also we calculate the conditional variance as follows,

$$\begin{aligned}
Var(\hat{I}^*|\mathbf{Y}) &= \frac{1}{n} Var\left\{ f(X_1) - \sum_{j=1}^{k} \tilde{\theta}_j \, \phi_j(X_1) | \mathbf{Y} \right\} \\
&= \frac{1}{n} \int \{f(x) - I - \sum_{j=1}^{k} \tilde{\theta}_j \, \phi_j(x)\}^2 \, \mu(dx) \\
&= \frac{1}{n} \int \{f_h(x) - I + \sum_{j=1}^{k} (\theta_j - \tilde{\theta}_j) \, \phi_j(x)\}^2 \, \mu(dx) \\
&= \frac{1}{n} \int \{f_h(x) - I\}^2 \, \mu(dx) + \frac{1}{n} \sum_{j=1}^{k} (\tilde{\theta}_j - \theta_j)^2,
\end{aligned}$$

and hence

$$E\{Var(\hat{I}^*|\mathbf{Y})\} = \frac{1}{n} \int \{f_h(x) - I\}^2 \, \mu(dx) + \frac{1}{n} \sum_{j=1}^{k} E\{(\tilde{\theta}_j - \theta_j)^2\},$$

where

$$E(\tilde{\theta}_j) = \frac{1}{m} \sum_{i=1}^{m} E\{f(Y_i) \, \phi_j(Y_i)\} = \theta_j,$$

$$E[(\tilde{\theta}_j - \theta_j)^2] = Var(\tilde{\theta}_j) = \frac{1}{m} \int \{f(y) \, \phi_j(y) - \theta_j\}^2 \mu(dy).$$

We complete the proof by substituting above result and (2.10) into (2.9). □

Remark 2.1. Theorem 1 shows that the variance of $\hat{I}^*$ has two parts: the variance part for $f_h(x)$ that is orthogonal to $\phi_j(x)$, and the variance part for estimating k coefficients θ_j. Note that from (2.2)-(2.5), we have

$$\begin{aligned}
Var(\hat{I}) &= \frac{1}{n} \int \{f(x) - I\}^2 \, \mu(dx) \\
&= \frac{1}{n} \left\{ \int \{f_h(x) - I\}^2 \, \mu(dx) + \int \{\theta \, h(x)\}^2 \, \mu(dx) \right\} \\
&= \frac{1}{n} \left\{ \int \{f_h(x) - I\}^2 \, \mu(dx) + \sum_{j=1}^{k} \theta_j^2 \right\}.
\end{aligned}$$

Since $f(x)$ can be expanded over the basis functions, with appropriate selection of k basis functions ϕ_j, much of $f(x)$ can be represented by $\phi_j(x)$ in the sense that $\sum_{j=1}^{k} \theta_j^2$ accounts for a large portion of $\int \{f(x) - I\}^2 \, \mu(dx)$, and thus the reduction from $\int \{f(x) - I\}^2 \, \mu(dx)$ to $\int \{f_h(x) - I\}^2 \, \mu(dx)$ can be very significant. Therefore, in comparison with the variance of $Var(\hat{I})$, $\hat{I}^*$ may achieve a big variance reduction, since the second term in the expression of $Var(\hat{I}^*)$ in Theorem 1 can be very small for reasonably large m.

Remark 2.2. To increase the variance reduction, we need to select basis functions $\phi_j = \psi_{\ell_j}$, $j = 1, \ldots, k$, such that $\sum_{j=1}^{k} \theta_j^2$ is maximized, where θ_j are defined in

(2.5). Ideally we should select k basis functions whose coefficients are k largest in terms of magnitude. That is, suppose that ψ_ℓ, $\ell = 1, 2, \ldots$, have coefficients

$$\alpha_\ell = \int f(x)\, \psi_\ell(x)\, \mu(dx).$$

We order the absolute values of α_ℓ in a decreasing order

$$|\alpha_{(1)}| \geq |\alpha_{(2)}| \geq \cdots ,$$

and denote the ordered coefficients by $\alpha_{(\ell)}$ and the corresponding basis functions by $\psi_{(\ell)}$. We select the top k basis functions $\phi_j = \psi_{(j)}$ with corresponding coefficients $\theta_j = \alpha_{(j)}$, $j = 1, \ldots, k$. Of course, the ideal selection needs to know coefficients α_ℓ, which are also integrals. As we have employed the pilot sample $Y_1, \ldots, Y_m$ to estimate the coefficients of the selected basis functions, we may as well use the sample to select basis functions. The method works as follows. From $Y_1, \ldots, Y_m$ we estimate coefficients α_ℓ

$$\hat{\alpha}_\ell = \frac{1}{m} \sum_{i=1}^{m} f(Y_i)\, \psi_\ell(Y_i), \qquad \ell = 1, \ldots, m,$$

and order them in absolute value

$$|\hat{\alpha}_{(1)}| \geq |\hat{\alpha}_{(2)}| \geq \cdots .$$

We pick up k such that

$$\frac{\sum_{j=1}^{k} \alpha_{(j)}^2}{\sum_{j=1}^{m} \alpha_j^2}$$

exceeds a pre-specified percentage such as 80% and 90%. Then we select $|\hat{\alpha}_{(j)}|$, $j = 1, \ldots, k$, and the corresponding k basis functions. Denote by $\tilde{\theta}_j = \hat{\alpha}_{(j)}$, and the corresponding selected basis functions by ϕ_j, $j = 1, \ldots, k$. We plug them into (2.8) to define the Monte Carlo estimator $\hat{I}^*$, which is totally sample dependent.

2.3. Construction of control variates by a dictionary of bases

Sometimes it is more convenient to work with a dictionary of bases that may not be orthogonal. Suppose that we have a dictionary of bases ψ_λ, $\lambda \in \Lambda$. Analog to the orthogonal case, we select k basis functions ψ_{ℓ_j}, $j = 1, \ldots, k$, set $\phi_j = \psi_{\ell_j}$, $j = 1, \ldots, k$, and take

$$f(x) = \sum_{j=1}^{k} \theta_j\, \phi_j(x) + f_h(x), \qquad f_h(x) \perp span\{\phi_1, \ldots, \phi_k\}, \tag{2.11}$$

where

$$\theta_j = \int f(x)\, \phi_j(x)\, \mu(dx). \tag{2.12}$$

From the pilot sample $Y_1, \ldots, Y_m$ we estimate θ_j by

$$\tilde{\theta}_j = \frac{1}{m} \sum_{i=1}^{m} f(X_i)\, \phi_j(X_i),$$

and build the Monte Carlo estimator of I by

$$\tilde{I}^* = \hat{I} - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(X_i) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(X_i)\right\}.$$

Similar to Theorem 1, we have the following theorem for the variance of $\tilde{I}^*$.

Theorem 2.2.

$$Var(\tilde{I}^*) = \frac{1}{n}\left\{\int \{f_h(x) - I\}^2\,\mu(dx) + \sum_{r=1}^{k}\sum_{j=1}^{k} Cov(\tilde{\theta}_r, \tilde{\theta}_j)\right\},$$

where

$$Cov(\hat{\theta}_r, \hat{\theta}_j) = \frac{1}{m}\int \{[f(y)]^2\,\phi_r(y)\,\phi_j(y) - \theta_r\,\theta_j\}\,\mu(dy).$$

Proof. Similar to the proof of Theorem 1, we use $\mathbf{Y} = (Y_1, \ldots, Y_m)$ and

$$Var(\tilde{I}^*) = E\{Var(\tilde{I}^*|\mathbf{Y})\} + Var\{E(\tilde{I}^*|\mathbf{Y})\},$$

and have

$$E(\tilde{I}^*|\mathbf{Y}) = I, \qquad Var(E(\tilde{I}^*|\mathbf{Y})) = Var(I) = 0.$$

We calculate the conditional variance as follows,

$$\begin{aligned}
Var(\tilde{I}^*|\mathbf{Y}) &= \frac{1}{n}Var\left\{f(X_1) - \sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(X_1)|\mathbf{Y}\right\} \\
&= \frac{1}{n}\int \{f(x) - I - \sum_{j=1}^{k}\tilde{\theta}_j\,\phi_j(x)\}^2\,\mu(dx) \\
&= \frac{1}{n}\int \{f_h(x) - I + \sum_{j=1}^{k}(\theta_j - \tilde{\theta}_j)\,\phi_j(x)\}^2\,\mu(dx) \\
&= \frac{1}{n}\int \{f_h(x) - I\}^2\,\mu(dx) + \frac{1}{n}\int\left\{\sum_{j=1}^{k}(\tilde{\theta}_j - \theta_j)\phi_j(x)\right\}^2\mu(dx) \\
&= \frac{1}{n}\int \{f_h(x) - I\}^2\,\mu(dx) \\
&\quad + \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}(\tilde{\theta}_i - \theta_i)\,(\tilde{\theta}_j - \theta_j)\int \phi_i(x)\,\phi_j\,\mu(dx).
\end{aligned}$$

Thus we have

$$\begin{aligned}
Var(\tilde{I}^*) &= E\{Var(\tilde{I}^*|\mathbf{Y})\} \\
&= \frac{1}{n}\int \{f_h(x) - I\}^2\,\mu(dx) \\
&\quad + \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}E(\tilde{\theta}_i - \theta_i)\,(\tilde{\theta}_j - \theta_j)\int \phi_i(x)\,\phi_j\,\mu(dx),
\end{aligned}$$

where for $j = 1, \ldots, k$,

$$E(\tilde{\theta}_j) = \frac{1}{m} \sum_{i=1}^{m} E\{f(Y_i)\, \phi_j(Y_i)\} = \theta_j,$$

$$\begin{aligned} E\{(\tilde{\theta}_i - \theta_i)\, (\hat{\theta}_j - \theta_j)\} &= Cov(\hat{\theta}_i, \hat{\theta}_j) \\ &= \frac{1}{m} \int \{f(y)\, \phi_i(y) - \theta_i\}\, \{f(y)\, \phi_j(y) - \theta_j\}\, \mu(dy) \\ &= \frac{1}{m} \int \{[f(y)]^2\, \phi_i(y)\, \phi_j(y) - \theta_i\, \theta_j\}\, \mu(dy). \end{aligned}$$ □

Remark 2.3. There exist several well known bases, and we may use these bases directly or with some modifications for the construction of the proposed Monte Carlo estimators. Univariate bases include Hermite polynomials, Laguerre polynomials, Fourier basis, splines, wavelet bases, wavelet packets and local cosine bases, and wavelet frames [3]. Multivariate bases include tensor products of univariate bases and radial bases.

3. Monte Carlo integration in Wiener space

Suppose $\{W_t \in \mathbb{R},\, t \in [0, T]\}$ is a Wiener process in probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The problem considered is to evaluate the expected value of a functional of W, that is, we need to compute $J = E[F(W)]$, where $F : C([0, T], \mathbb{R}) \to \mathbb{R}$ is a functional on the space of $\mathbb{R}$-valued continuous functions of t. The basic Monte Carlo method is to estimate J by

$$\hat{J} = \frac{1}{n} \sum_{i=1}^{n} F(W^i), \tag{3.1}$$

where $W^1, \ldots, W^n$ are independent simulations of W. Such problems are encountered in stochastic control, partial differential equations, and mathematical finance. Specifically, consider a diffusion process $\{X_t \in \mathbb{R}, t \in [0, T]\}$ governed by the following Itô stochastic differential equation,

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dW_t, \tag{3.2}$$

where $\{W_t \in \mathbb{R}, t \in [0, T]\}$ is a standard Wiener process, $b(X_t, t)$ is drift, and $\sigma(X_t, t)$ is diffusion variance [8]. We often need to evaluate $E[G(X)]$ for some functional G. As the diffusion process X_t is a functional of W_t, $G(X)$ is a functional of W, and thus we may calculate $E[G(X)]$ by Monte Carlo methods. Such Monte Carlo methods have a number of potential applications. For example, for partial differential equations like the Cauchy problem, the Feynmann-Kac formula provides stochastic representations for their solutions as functionals of W_t; in mathematical finance asset prices are often assumed to follow Itô processes, and the prices of options and derivatives often can be expressed as functionals of W_t. Monte Carlo methods are used to numerically evaluate derivative prices and solutions of the partial differential equations. The Monte Carlo evaluation is to simulate independent realizations of W_t, solve (3.2) numerically by the use of any of a number of discretization schemes for stochastic differential equations, where the simplest of which is the Euler scheme, and then evaluate the sample average given by (3.1). See [1, 2, 7, 9, 12].

In this section we investigate variance reduction methods by Itô-Wiener chaos expansion for the Monte Carlo evaluation of $J = E[F(W)]$. For $H = F(W) \in L^2(\Omega, \mathcal{F}, I\!P)$, we have the following Itô-Wiener chaos expansion [5, 6, 9],

$$\begin{aligned} H = F(W) = J &+ \int_0^T h_1(t)\, dW_t + \int_{0\le t_1<t_2\le T} h_2(t_1,t_2)\, dW_{t_1}\, dW_{t_2} \\ &+ \cdots + \int_{0\le t_1<\cdots<t_d\le T} h_d(t_1,\dots,t_d)\, dW_{t_1}\cdots dW_{t_d} + \cdots, \end{aligned} \tag{3.3}$$

where $h_d(x) \in L^2(I\!R^d, dx)$, and

$$\int_{0\le t_1<\cdots<t_d\le T} h_d(t_1,\dots,t_d)\, dW_{t_1}\cdots dW_{t_d}$$

are Itô multiple stochastic integrals which have mean zero and variance

$$\int_{0\le t_1<\cdots<t_d\le T} [h_d(t_1,\dots,t_d)]^2\, dt_1\cdots dt_d = \frac{1}{d!}\|h_d\|^2,$$

and are uncorrelated,

$$E\left[\int_{0\le t_1<\cdots<t_d\le T} h_d\, dW_{t_1}\cdots dW_{t_d} \int_{0\le t_1<\cdots<t_r\le T} h_r\, dW_{t_1}\cdots dW_{t_r}\right] = 0, \quad d \ne r.$$

Hence H has variance

$$Var(H) = \|h_1\|^2 + \frac{1}{2!}\|h_2\|^2 + \cdots + \frac{1}{d!}\|h_d\|^2 + \cdots. \tag{3.4}$$

Suppose $\phi_j^d(x)$ is an orthonormal basis for $L^2(I\!R^d, dx)$. Then we have a representation for $h_d(x) \in L^2(I\!R^d, dx)$,

$$h_d(x) = \sum_{j=1}^{\infty} \theta_j^d\, \phi_j^d(x), \qquad \|h_d\|^2 = \sum_{j=1}^{\infty} (\theta_j^d)^2.$$

Substituting the above representation into Itô-Wiener chaos expansion (3.3) we obtain a representation of H

$$\begin{aligned} H = J &+ \sum_j \theta_j^1 \int_0^T \phi_j^1(t)\, dW_t + \sum_j \theta_j^2 \int_{0\le t_1<t_2\le T} \phi^2(t_1,t_2)\, dW_{t_1}\, dW_{t_2} \\ &+ \cdots + \sum_j \theta_j^d \int_{0\le t_1<\cdots<t_d\le T} \phi_j^d(t_1,\dots,t_d)\, dW_{t_1}\cdots dW_{t_d} + \cdots, \end{aligned} \tag{3.5}$$

where

$$\theta_j^d = E\left\{H \int_{0\le t_1<\cdots<t_d\le T} \phi_j^d(t_1,\dots,t_d)\, dW_{t_1}\cdots dW_{t_d}\right\},$$

and an expression for the variance of H

$$Var(H) = \sum_j (\theta_j^1)^2 + \frac{1}{2!}\sum_j (\theta_j^2)^2 + \cdots + \frac{1}{d!}\sum_j (\theta_j^d)^2 + \cdots. \tag{3.6}$$

From (3.4) and (3.6) we obtain the variance of the basic Monte Carlo estimator $\hat{I}$ defined in (3.1)

$$Var(\hat{J}) = \frac{1}{n} Var(H) = \frac{1}{n}\left\{\|h_1\|^2 + \frac{1}{2!}\|h_2\|^2 + \cdots + \frac{1}{d!}\|h_d\|^2 + \cdots\right\}$$

$$= \frac{1}{n}\left\{\sum_j (\theta_j^1)^2 + \frac{1}{2!}\sum_j (\theta_j^2)^2 + \cdots + \frac{1}{d!}\sum_j (\theta_j^d)^2 + \cdots\right\}. \tag{3.7}$$

3.1. Variance reduction by antithetic variates

Since Wiener process W_t is symmetric, that is, $-W$ is also a Wiener process. We use $-W$ to construct antithetic variable $H_0 = F(-W)$, which has $E(H_0) = J$. The resulting Monte Carlo estimator

$$\tilde{J} = \frac{1}{n}\sum_{i=1}^{n} \{F(W^i) + F(-W^i)\}/2. \tag{3.8}$$

Replace W by $-W$ in (3.3) we obtain

$$H_0 = F(-W) = J - \int_0^T h_1(t)\, dW_t + \int_{0 \le t_1 < t_2 \le T} h_2(t_1, t_2)\, dW_{t_1}\, dW_{t_2} + \cdots$$
$$+ (-1)^d \int_{0 \le t_1 < \cdots < t_d \le T} h_d(t_1, \ldots, t_d)\, dW_{t_1} \cdots dW_{t_d} + \cdots .$$

Combining the above expression with (3.3) and (3.5) we conclude

$$[H + H_0]/2 = [F(W) + F(-W)]/2 = J + \int_{0 \le t_1 < t_2 \le T} h_2(t_1, t_2)\, dW_{t_1}\, dW_{t_2} + \cdots$$
$$+ \int_{0 \le t_1 < \cdots < t_{2d} \le T} h_{2d}(t_1, \ldots, t_{2d})\, dW_{t_1} \cdots dW_{t_{2d}} + \cdots$$
$$= J + \sum_j \theta_j^2 \int_{0 \le t_1 < t_2 \le T} \phi^2(t_1, t_2)\, dW_{t_1}\, dW_{t_2} + \cdots$$
$$+ \sum_j \theta_j^{2d} \int_{0 \le t_1 < \cdots < t_{2d} \le T} \phi_j^{2d}(t_1, \ldots, t_{2d})\, dW_{t_1} \cdots dW_{t_{2d}} + \cdots . \tag{3.9}$$

Using (3.8) and (3.9) we can show that $E(\tilde{J}) = J$ and $\tilde{J}$ has variance

$$Var(\tilde{J}) = \frac{1}{n}\left\{\frac{1}{2!}\|h_2\|^2 + \frac{1}{4!}\|h_4\|^2 \cdots + \frac{1}{(2d)!}\|h_{2d}\|^2 + \cdots\right\}$$

$$= \frac{1}{n}\left\{\frac{1}{2!}\sum_j (\theta_j^2)^2 + \frac{1}{4!}\sum_j (\theta_j^4)^2 + \cdots + \frac{1}{(2d)!}\sum_j (\theta_j^{2d})^2 + \cdots\right\}. \tag{3.10}$$

Remark 3.1. In comparison with the variance of $\hat{J}$ given by (3.7), the variance of $\tilde{J}$ in (3.10) has only terms with even d, thus $\tilde{J}$ should achieve a substantial variance reduction. From now on we consider Itô-Wiener expansion with only terms indexed by even d.

3.2. Variance reduction by control variates

Of θ_j^d with even d, we order $|\theta_j^d|/d!$ in a decreasing order and pick up the k largest ones. Denote the selected k coefficients by $\theta_{j_\ell}^{d_\ell}$, $\ell = 1, \ldots, k$. We use the terms with $\theta_{j_\ell}^{d_\ell}$ in (3.9) as control variates to adjust Monte Carlo estimator $\tilde{J}$ defined in (3.8) for variance reduction. First we simulate another m independent samples, $B^1, \ldots, B^m$, of Wiener process, and use the samples to estimate $\theta_{j_\ell}^{d_\ell}$ by

$$\hat{\theta}_{j_\ell}^{d_\ell} = \frac{1}{m} \sum_{i=1}^{m} F(B^i) \int_{0 \le t_1 < \cdots < t_{d_\ell} \le T} \phi_{j_\ell}^{d_\ell}(t_1, \ldots, t_{d_\ell})\, dB_{t_1}^i \cdots dB_{t_{d_\ell}}^i .$$

Then we build the Monte Carlo estimator

$$\begin{aligned} \hat{J}^* = \frac{1}{n} \sum_{i=1}^{n} \Bigg\{ & \frac{F(W^i) + F(-W^i)}{2} \\ & - \sum_{\ell=1}^{k} \hat{\theta}_{j_\ell}^{d_\ell} \int_{0 \le t_1 < \cdots < t_{d_\ell} \le T} \phi_{j_\ell}^{d_\ell}(t_1, \ldots, t_{d_\ell})\, dW_{t_1}^i \cdots dW_{t_{d_\ell}}^i \Bigg\}. \end{aligned} \tag{3.11}$$

The variance of $\hat{J}^*$ is given in the following theorem.

Theorem 3.1. *$\hat{J}^*$ is an unbiased estimator of J with variance*

$$Var(\hat{J}^*) = \frac{1}{n} \left\{ \sum_{(d,j) \ne (d_\ell, j_\ell)} \left(\theta_j^d/d!\right)^2 + \sum_{\ell=1}^{k} Var\left(\hat{\theta}_{j_\ell}^{d_\ell}\right) \right\},$$

where

$$Var\left(\hat{\theta}_{j_\ell}^{d_\ell}\right) = \frac{1}{m} Var\left(F(B) \int_{0 \le t_1 < \cdots < t_{d_\ell} \le T} \phi_{j_\ell}^{d_\ell}(t_1, \ldots, t_{d_\ell})\, dB_{t_1} \cdots dB_{t_{d_\ell}} \right).$$

Proof. Let $\mathbf{B} = (B^1, \ldots, B^m)$, and note that

$$Var(\hat{J}^*) = E\{Var(\hat{J}^*|\mathbf{B})\} + Var\{E(\hat{J}^*|\mathbf{B})\}. \tag{3.12}$$

Since multiple stochastic integrals have zero mean, simple conditional calculations show

$$\begin{aligned} E(J^*|\mathbf{B}) &= E(\tilde{J}) - \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{k} \tilde{\theta}_{j_\ell}^{d_\ell} E\left[\int_{0 \le t_1 < \cdots < t_{d_\ell} \le T} \phi_{j_\ell}^{d_\ell}(t_1, \ldots, t_{d_\ell})\, dW_{t_1}^i \cdots dW_{t_{d_\ell}}^i \right] \\ &= J, \end{aligned}$$

and thus $E[\hat{J}^*] = J$, and

$$Var\{E(\hat{J}^*|\mathbf{B})\} = Var(J) = 0. \tag{3.13}$$

On the other hand, from (3.9) and (3.11) we obtain

$$
\begin{aligned}
\hat{J}^* = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ & \sum_{\text{even } d}\sum_{j}\theta_j^d \int_{0\le t_1<\cdots<t_d\le T}\phi_j^d(t_1,\ldots,t_d)\,dW_{t_1}^i\cdots dW_{t_d}^i \\
& - \sum_{\ell=1}^{k}\hat{\theta}_{j_\ell}^{d_\ell}\int_{0\le t_1<\cdots<t_{d_\ell}\le T}\phi_{j_\ell}^{d_\ell}(t_1,\ldots,t_{d_\ell})\,dW_{t_1}^i\cdots dW_{t_{d_\ell}}^i\Bigg\} \\
= \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ & \sum_{(d,j)\ne(d_\ell,j_\ell)}\theta_j^d\int_{0\le t_1<\cdots<t_d\le T}\phi_j^d(t_1,\ldots,t_d)\,dW_{t_1}^i\cdots dW_{t_d}^i \\
& - \sum_{\ell=1}^{k}(\hat{\theta}_{j_\ell}^{d_\ell}-\theta_{j_\ell}^{d_\ell})\int_{0\le t_1<\cdots<t_{d_\ell}\le T}\phi_{j_\ell}^{d_\ell}(t_1,\ldots,t_{d_\ell})\,dW_{t_1}^i\cdots dW_{t_{d_\ell}}^i\Bigg\}.
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
Var(\hat{J}^*|\mathbf{B}) &= \frac{1}{n}\Bigg\{\sum_{(d,j)\ne(d_\ell,j_\ell)}\theta_j^d\int_{0\le t_1<\cdots<t_d\le T}\phi_j^d(t_1,\ldots,t_d)\,dt_1\cdots dt_d \\
&\qquad - \sum_{\ell=1}^{k}(\hat{\theta}_{j_\ell}^{d_\ell}-\theta_{j_\ell}^{d_\ell})\int_{0\le t_1<\cdots<t_{d_\ell}\le T}\phi_{j_\ell}^{d_\ell}(t_1,\ldots,t_{d_\ell})\,dt_1\cdots dt_{d_\ell}\Bigg\} \\
&= \frac{1}{n}\left\{\sum_{(d,j)\ne(d_\ell,j_\ell)}(\theta_j^d)^2/d! + \sum_{\ell=1}^{k}(\hat{\theta}_{j_\ell}^{d_\ell}-\theta_{j_\ell}^{d_\ell})^2/d_\ell!\right),
\end{aligned}
$$

and hence

$$
E\{Var(\hat{J}^*|\mathbf{B})\} = \frac{1}{n}\left\{\sum_{(d,j)\ne(d_\ell,j_\ell)}(\theta_j^d)^2/d! + \sum_{\ell=1}^{k}E[(\hat{\theta}_{j_\ell}^{d_\ell}-\theta_{j_\ell}^{d_\ell})^2]/d_\ell!\right),
$$

where $E(\hat{\theta}_{j_\ell}^{d_\ell})=\theta_{j_\ell}^{d_\ell}$, and

$$
\begin{aligned}
E[(\hat{\theta}_{j_\ell}^{d_\ell}-\theta_{j_\ell}^{d_\ell})^2] &= Var(\hat{\theta}_{j_\ell}^{d_\ell}) \\
&= \frac{1}{m}Var\left(F(B)\int_{0\le t_1<\cdots<t_{d_\ell}\le T}\phi_{j_\ell}^{d_\ell}(t_1,\ldots,t_{d_\ell})\,dB_{t_1}\cdots dB_{t_{d_\ell}}\right).
\end{aligned}
$$

We complete the proof by substituting above result and (3.13) into (3.12). □

Remark 3.2. From Theorem 3 we can see that with appropriate selection of k terms, $\sum_{(d,j)\ne(d_\ell,j_\ell)}(\theta_j^d)^2/d!$ can be very small, and also the sum of $Var(\hat{\theta}_{j_\ell}^{d_\ell})$ over $1\le\ell\le k$ may be very small for reasonably large m. Thus, similar to the finite dimensional case in Section 2, in comparison with the variance of $\hat{J}$, $\hat{J}^*$ can achieve a significant variance reduction. Also the same selection scheme described in Section 2.2 can be adopted here for the selection of k and the k terms $\theta_{j_\ell}^{d_\ell}$ used in the construction of $\hat{J}^*$ based on simulated $B^1,\ldots,B^m$.

4. Examples

We illustrate two simple examples to demonstrate the performance of the proposed methods. The first example is the integral of a non-smooth function. Consider

$$I = \int_{-1}^{1} f(x)\, dx = \int_{0}^{1} 2\, f(2\, u - 1)\, du,$$

where

$$f(x) = \sum_{j=0}^{\infty} 2^{-0.25\, j}\, cos(2^j\, \pi\, x)$$

is continuous but nowhere differentiable. The theoretical value of I is 0. Basic Monte Carlo estimator $\hat{I}$ has variance σ^2/n, where

$$\sigma^2 = \int_{-1}^{1} [f(x)]^2\, dx = \frac{1}{1 - 2^{-0.5}}.$$

Take cosine basis

$$cos(\pi\, j\, x), \qquad j = 1, \ldots,$$

and select the four basis functions with $j = 2, 4, 8, 16$. The corresponding Monte Carlo estimator $\hat{I}^*$ in (2.8) has variance $\sigma_*^2/n + \tau^2/(n\, m)$, where

$$\sigma_*^2 = \sum_{j=5}^{\infty} 2^{-0.5\, j} = \frac{2^{-2.5}}{1 - 2^{-0.5}} = 0.177\, \sigma^2,$$

and τ^2 is given by the second term in the expression of $Var(\hat{I}^*)$ in Theorem 1. For reasonably large m, the term $\tau^2/(n\, m)$ is negligible, so with $\hat{I}$ as benchmark, $\hat{I}^*$ can make 82% variance reduction.

The second example is a Wiener integral. Consider Itô process X_t, where

$$dX_t = X_t\, dt + h(t)\, X_t\, dW_t, \qquad t \in [0, 1], \qquad X_0 = 0.$$

where $h(t) = 1$ for $t \in [0, 1/2]$ and $= 2$ for $t \in (1/2, 1]$. The problem is to evaluate $J = E\{\|X_1\|\}$. The theoretical value of $J = e = 2.7183$. Basic Monte Carlo estimator $\hat{J}$ has variance σ^2/n, where

$$\sigma^2 = e^2\, \{e^{2.5} - 1\} = 82.6281.$$

Antithetic variates estimator $\tilde{J}$ has variance σ_0^2/n, where

$$\sigma_0^2 = e^2\, \{(e^{2.5} + e^{-2.5})/2 - 1\} = 37.9223,$$

which represents 54% variance reduction from $\hat{J}$ to $\tilde{J}$.

Next we consider control variates estimator $\hat{J}^*$. Take Haar wavelets on $[0, 1]$: $\phi(x) = 1$, $\psi(x) = -1$ for $x \in [0, 1/2]$ and 1 for $x \in (1/2, 1]$, and

$$\psi_{j,k}(x) = 2^{j/2}\, \psi(2^j\, x - k), \qquad j = 1, 2, \ldots, k = 0, \ldots, 2^j - 1,$$

and use them to form wavelets on $[0, 1]^d$ (see [3]).

Four terms are selected to form control variates and build the Monte Carlo estimator $\hat{J}^*$. The four terms are selected as follow: three for $h_2(x)$ in $[0, 1]^2$:

$$2.25\, \phi(t_1)\, \phi(t_2), \;\; 0.75\, \phi(t_1)\, \psi(t_2), \;\; 0.75\, \psi(t_1)\, \phi(t_2),$$

and one for $h_4(x)$ in $[0,1]^4$:

$$5.0625\,\phi(t_1)\,\phi(t_2)\,\phi(t_3)\,\phi(t_4).$$

The control variates Monte Carlo estimator $\hat{J}^*$ has variance $\sigma_*^2/n+\tau^2/(m\,n)$, where

$$\sigma_*^2 = 0.9707\,e^2 = 7.1723,$$

and τ^2 is given by the second term in the expression of $Var(\hat{J}^*)$ in Theorem 3. For reasonably large m, the term $\tau^2/(n\,m)$ is negligible, so the variance can be further reduced by about 81% from $\tilde{J}$ to $\hat{J}^*$. In fact, if we take $\hat{J}$ as benchmark, $\hat{J}^*$ can reduce variance by about 91%, or equivalently, $\hat{J}$ has variance 11 times higher than $\hat{J}^*$.

Acknowledgements

Wangs research was partially supported by NSF grant DMS-1005635.

References

[1] Barraquand, J. (1995). Numerical valuation of high dimensional multivariate European securities. *Management Science* **41** 1882–1891.
[2] Boyle, P., Broadie, M. and Glasserman, P. (1997). Monte Carlo methods for security pricing. *J. Econ. Dynamics and Control* **21** 1267–1321.
[3] Daubechies, I. (1992). *Ten Lectures on Wavelets. CBMS-NSF Series in Applied Mathematics* **61**. SIAM, Philadelphia.
[4] Hald, O. (1987). Approximation of Wiener integrals. *J. Comput. Phys.* **69** 460–470.
[5] Hu, Y. (1997). Itô-Wiener chaos expansion with exact residual and correlation, variance inequalities. *J. Theoretical Probab.* **10** 835–848.
[6] Itô, K. (1951). Multiple Wiener integrals. *J. Math. Soc. Japan* **3** 157–169.
[7] Jäckel, P. (2002). *Monte Carlo Methods in Finance.* Wiley, England.
[8] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus,* 2nd ed. Springer, New York.
[9] Kloeden, P. E. and Platen, E. (1999). *Numerical Solution for Stochastic Differential Equations.* Springer, New York.
[10] Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer, New York.
[11] Plapp, M. and Karma, A. (2000). Multiscale finite-difference-diffusion-Monte-Carlo method for simulating dendritic solidification. *J. Comput. Phys.* **165** 592–619.
[12] Plaskota, L., Wasilkowski, G. W. and Woźniakowski, H. (2000). A new algorithm and worst case complexity for Feynman-Kac path integration. *J. Comput. Phys.* **164** 335–353.