Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

FINITE VARIANCE UNBIASED ESTIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

Ankush Agarwal

Centre de Mathématiques Appliquées (CMAP)
École Polytechnique and CNRS
Route de Saclay
91128 Palaiseau Cedex, FRANCE

Emmanuel Gobet

Centre de Mathématiques Appliquées (CMAP)
École Polytechnique and CNRS
Route de Saclay
91128 Palaiseau Cedex, FRANCE

ABSTRACT

We develop a new unbiased estimation method for Lipschitz continuous functions of multi-dimensional stochastic differential equations with Lipschitz continuous coefficients. This method provides a finite variance estimator based on a probabilistic representation which is similar to the recent representations obtained through the parametrix method and recursive application of the automatic differentiation formula. Our approach relies on appropriate change of variables to carefully handle the singular integrands appearing in the iterated integrals of the probabilistic representation. It results in a scheme with randomized intermediate times where the number of intermediate times has a Pareto distribution.

1 INTRODUCTION

For $d \geq 1$, we consider the process $\{X_s^{t,x}; s \geq t\}$ which is the solution of the following stochastic differential equation (SDE):

$$\mathrm{d}X_s^{t,x} = b(s, X_s^{t,x})\mathrm{d}s + \sigma(s, X_s^{t,x})\mathrm{d}W_s, \quad s \geq t, \quad X_t^{t,x} = x,$$

where W is a d-dimensional Wiener process and for fixed time $T > 0$, the coefficients $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathscr{S}^d$ satisfy some standard assumptions. Here, $\mathscr{S}^d$ denotes the set of $d \times d$ dimensional matrices. We are interested to calculate for some bounded measurable function g the following expectation

$$u(t,x) = \mathbb{E}[g(X_T^{t,x})].$$

When no explicit solution is available for $X_T^{t,x}$, the classical approach is to estimate the expectation by using time discretization schemes such as Euler (Platen and Kloeden 1992) and Milstein (Milstein 1975) schemes. We can also use Gaussian approximations with some corrections (Bompis and Gobet 2014). However, due to the time discretization step, these methods inherently contain a bias which vanishes asymptotically. A way to address this issue is through bias reduction schemes such as the *randomization method* (random number of discretization steps) of Rhee and Glynn (2015) which can be seen as a randomized version of the multilevel Monte Carlo method of Giles (2008). However, the *randomization method* has a finite cost but infinite variance in full generality (when b, σ, g are Lipschitz continuous). The variance of the *randomization method* based estimator is finite when σ is constant and b is C_b^2, indeed, as the Euler scheme then converges strongly at order 1 (for $k \in \mathbb{N}$, C_b^k stands for the set of bounded continuous functions φ from one Euclidean space to another, k times continuously differentiable with bounded derivatives). Our aim is to propose a new unbiased estimation scheme for $u(t,x)$, in the multidimensional setting $d \geq 1$ with finite cost such that the estimator has finite variance in the case when b, σ and g are Lipschitz continuous.

Some of the recent approaches in this direction involve generating unbiased samples of $g(X_T^{t,x})$ using the method of exact simulation by Beskos and Roberts (2005) and Chen and Huang (2013). Note that these schemes essentially handle one-dimensional situations. More recently, for *time-independent* coefficients b and σ, Kohatsu-Higa and co-authors (Bally and Kohatsu-Higa 2015, Andersson and Kohatsu-Higa 2017) have proposed two new schemes (the so-called *forward* (F) and *backward* (B) methods) based on parametrix techniques (infinite series expansion). After reinterpreting the infinite series using a Poisson process $(N_t : t \geq 0)$ (with parameter $\lambda > 0$), these *forward* and *backward* methods take the following form

$$\mathbb{E}\left[g(X_T^{0,x_0})\right] = e^{\lambda T}\mathbb{E}\left[\Phi(X_T^\pi)\prod_{i=0}^{N_T-1}\lambda^{-1}\theta_{\tau_{i+1}-\tau_i}(X_{\tau_i}^\pi, X_{\tau_{i+1}}^\pi)\right] \tag{1}$$

where $\pi := (\tau_i : i \geq 1)$ are the Poisson jump times on $[0,T]$. In the above X^π is a Euler scheme based on the grid π, with drift μ and diffusion σ, with some initialization according to a probability measure $\nu(\mathrm{d}x)$, and $\theta_t(x,y) := \frac{1}{2}\sum_{i,j=1}^d \kappa_t^{i,j}(x,y) - \sum_{i=1}^d \rho_t^i(x,y)$, for some functions κ and ρ. More precisely, we have the following:

The **forward method** is defined by (see Bally and Kohatsu-Higa (2015), Theorems 5.7 and 7.1)

$$\begin{cases} \nu(\mathrm{d}x) := \delta_{x_0}(x), \qquad \Phi(x) := g(x), \qquad \mu(x) := b(x), \qquad a(y) := \sigma\sigma^\top(y), \\ \kappa_t^{i,j}(x,y) := \partial_{i,j}^2 a^{i,j}(y) + \partial_j a^{i,j}(y) H_{ta(x)}^i(y-x-b(x)t) \\ \qquad\qquad +\partial_i a^{i,j}(y) H_{ta(x)}^j(y-x-b(x)t) + (a^{i,j}(y)-a^{i,j}(x)) H_{ta(x)}^{i,j}(y-x-b(x)t), \\ \rho_t^i(x,y) := \partial_i b^i(y) + (b^i(y)-b^i(x)) H_{ta(x)}^i(y-x-b(x)t) \end{cases} \tag{F}$$

where $H^i, H^{i,j}$ are related to the two first Hermite polynomials:

$$H_{ta(x)}^i(z) := -t^{-1}(a^{-1}z)^i, \qquad H_{ta(x)}^{i,j}(z) = t^{-2}(a^{-1}z)^i(a^{-1}z)^j - t^{-1}(a^{-1}z)^{i,j}.$$

The representation (1) with (F) holds under the assumption that $b \in C_b^1$ and $a = \sigma\sigma^\top \in C_b^2$ and that a is uniformly elliptic (Andersson and Kohatsu-Higa (2017), Theorem 5.1). Assuming that function g is bounded measurable is sufficient. Furthermore, if a is constant, the variance of the random variable in (1) is finite, otherwise in general, it is infinite. However, there is a modification of (1) using $\tau_{i+1} - \tau_i$ with Beta distribution (Andersson and Kohatsu-Higa (2017), proposition 7.3) such that the new random variable has finite polynomial moments of any order.

The **backward method** is defined, upon the additional condition that g is a density function, by

$$\begin{cases} \nu(\mathrm{d}x) := g(x)\mathrm{d}x, \qquad \Phi(x) := \delta_{x_0}(x), \qquad \mu(x) := -b(x), \\ \kappa_t^{i,j}(x,y) := (a^{i,j}(y)-a^{i,j}(x)) H_{ta(x)}^{i,j}(y-x+b(x)t), \\ \rho_t^i(x,y) := (b^i(y)-b^i(x)) H_{ta(x)}^i(y-x+b(x)t). \end{cases} \tag{B}$$

The representation (1) with (B) holds under the assumption that $b \in H_b^\alpha$ and $\sigma \in H_b^\alpha$ (for $\alpha \in (0,1]$, H_b^α stands for bounded and uniform Hölder continuous functions) and that a is uniformly elliptic (Andersson and Kohatsu-Higa (2017), Theorem 5.1). Furthermore, when σ is constant, the variance of the random variable in (1) is finite and for non-constant σ, it is in general infinite. The aforementioned modification of (1) using Beta sampling leads to finite variance only in dimension 1.

Another unbiased estimation method has been developed recently by Henry-Labordere et al. (2017). It makes use of a decomposition method of the expectation $u(t,x)$ using stochastic calculus and the underlying PDE, with recursive computations of the gradient and the Hessian matrix of u through the integration by parts formulas of Malliavin calculus. Similar ideas have been performed for sensitivity analysis of diffusion processes by Monte-Carlo methods (Gobet and Munos 2005, Theorem 2.11). The crucial point in the analysis of (Henry-Labordere et al. 2017) is to use piece-wise constant Gaussian proxy so that Malliavin weights are explicit. Under Lipschitz assumptions on b,σ,g and uniform ellipticity, the authors derive a representation similar to (1) using a Poisson process and prove that in the case of constant σ, the estimator variance is finite (Henry-Labordere et al. 2017, Theorem 3.2). The case of non constant σ with finite variance is handled in dimension $d=1$, with $b=0,\sigma\in C_b^2, g\in C_b^2$ (Henry-Labordere et al. 2017, Theorem 4.2). Doumbia et al. (2017) proposed a variant of the previous Malliavin method, using $\tau_{i+1}-\tau_i$ distributed as Gamma random variables. It allows to prove that the new estimator has finite variance for any $d\geq 1$, as soon as $b,\sigma\in H_b^1$, $g\in C_b^1$ and $\nabla g\in H_b^1$. In contrast with the previous parametrix-based methods, these Malliavin methods allow time-dependent coefficients b and σ with appropriate Hölder regularity in time.

1.1 Contributions

It appears that the open problems, for which finite variance is not available so far, correspond to the cases σ non constant, with low regularity conditions both on (b,σ) and g. In this paper, we propose a method which is able to undertake this general setting. We follow the Malliavin approach of Henry-Labordere et al. (2017) but with substantial improvement to allow such level of generality. Mainly, we incorporate smart sampling of the grid $\pi=(\tau_i : i\geq 0)$ to suitably handle the singular integrands as first suggested by Helluy, Maire, and Ravel (1998). The representation we obtain (see Equation (17)) takes a form close to (1) with a random variable N_T that has a Pareto-type distribution (instead of Poisson).

1.2 Assumptions

Assumption 1 $(b,\sigma):[0,T]\times\mathbb{R}^d\to\mathbb{R}^d\times\mathscr{S}^d$ are uniformly bounded, $\frac{1}{2}$-Hölder continuous in time and Lipschitz continuous in space variable. In particular, we have for some constant L,

$$|(b,\sigma)(t,x)-(b,\sigma)(s,y)|\leq L\Big(\sqrt{|t-s|}+|x-y|\Big),\ \text{for all}\ (t,x),(s,y)\in[0,T]\times\mathbb{R}^d.$$

Assumption 2 Let $a:=\frac{1}{2}\sigma\sigma^\top$. Then, the diffusion process $X^{t,x}$ is non-degenerate, i.e. there exists $\delta>0$ such that for $y\in\mathbb{R}^d, \langle a(t,x)y,y\rangle\geq\delta|y|^2,$ for all $(t,x)\in[0,T]\times\mathbb{R}^d$.

Assumption 3 $g\in C_b^2$.

Assumption 4 $g\in H_b^1$ (Lipschitz continuous).

2 PRELIMINARY CALCULATIONS

2.1 Derivation of Representation for $g\in C^2$

The principal difficulty in the derivation of our unbiased estimation method arises due to the irregularity of coefficients (b,σ) and function g and remains indifferent to the dimensionality of the SDE. Furthermore, the problem remains unchanged even when $b=0$. Therefore, for simplicity of exposition, we present our results in the case of SDE without any drift term and $d=1$. Then, we have

$$X_s^{t,x}=x+\int_t^s\sigma(r,X_r^{t,x})\mathrm{d}W_r,\, s\geq t,$$

where $\{W_r; r\geq 0\}$ is a standard Brownian motion. Consider the following linear partial differential equation (PDE)

$$\partial_t u + a\mathrm{D}^2 u = 0,\qquad u(T,\cdot)=g(\cdot). \tag{2}$$

Here, $\mathrm{D}^2(\cdot)$ denotes second order differential operator w.r.t. state variable. We know that under Assumptions 1, 2 and 3, if there exists a unique classical solution $v\in C_b^{1,2}([0,T]\times\mathbb{R},\mathbb{R})$ for equation (2), it can also be written as

$$v(t,x) = u(t,x) = \mathbb{E}\left[g(\tilde{X}_T^{t,x}) + \int_t^T H^{t,x}(s_1,\tilde{X}_{s_1}^{t,x})\mathrm{D}^2 u(s_1,\tilde{X}_{s_1}^{t,x})\mathrm{d}s_1\right], \tag{3}$$

where for any $(t,x) \in [0,T]\times\mathbb{R}$, the function $H^{t,x}:[0,T]\times\mathbb{R}\to\mathbb{R}$ and the process $\{\tilde{X}_s^{t,x}; s\geq t\}$ are given as

$$H^{t,x}(s,y) := \big(a(s,y)-a(t,x)\big), \text{ and, } \tilde{X}_s^{t,x} := x+\sigma(t,x)(W_s-W_t).$$

$\tilde{X}^{t,x}$ is also commonly known as Gaussian *proxy process*. Next, we have the following automatic differentiation formula by applying Elworthy's formula (Fournié et al. 1999) (see (Doumbia et al. 2017, Lemma 3.1) or (Henry-Labordere et al. 2017, Lemma A.3) for technical details)

$$\mathrm{D}^2 u(s_1,\tilde{X}_{s_1}^{t,x}) = \mathbb{E}\left[g(\tilde{X}_T^{s_1,\tilde{X}_{s_1}^{t,x}})\mathscr{V}_{s_1,T}^{s_1,\tilde{X}_{s_1}^{t,x}} + \int_{s_1}^T H^{s_1,\tilde{X}_{s_1}^{t,x}}(s_2,\tilde{X}_{s_2}^{s_1,\tilde{X}_{s_1}^{t,x}})\mathrm{D}^2 u(s_2,\tilde{X}_{s_2}^{s_1,\tilde{X}_{s_1}^{t,x}})\mathscr{V}_{s_1,s_2}^{s_1,\tilde{X}_{s_1}^{t,x}}\mathrm{d}s_2\middle|\mathscr{F}_{s_1}\right], \tag{4}$$

where $(\mathscr{F}_s)_{s\geq t}$ is the natural filtration associated to $\tilde{X}^{t,x}$ and the second order Malliavin weight is given as

$$\mathscr{V}_{s',t'}^{\tilde{t},\tilde{x}} := \sigma(\tilde{t},\tilde{x})^{-2}\cdot\big((t'-s')^{-2}(\delta W_{s',t'})^2-(t'-s')^{-1}\big), \quad \delta W_{s',t'} := W_{t'}-W_{s'}. \tag{5}$$

In this case, formula (4) is a result of the likelihood ratio method of Broadie and Glasserman (1996). We use the result in (4) to write (3) as

$$\begin{aligned} u(t,x) = \mathbb{E}\Bigg[&g(\tilde{X}_T^{t,x}) + \int_t^T H^{t,x}(s_1,\tilde{X}_{s_1}^{t,x})\mathscr{V}_{s_1,T}^{s_1,\tilde{X}_{s_1}^{t,x}} g(\tilde{X}_T^{s_1,\tilde{X}_{s_1}^{t,x}})\mathrm{d}s_1 \\ &+ \int_t^T H^{t,x}(s_1,\tilde{X}_{s_1}^{t,x})\mathrm{d}s_1\int_{s_1}^T H^{s_1,\tilde{X}_{s_1}^{t,x}}(s_2,\tilde{X}_{s_2}^{s_1,\tilde{X}_{s_1}^{t,x}})\mathrm{D}^2 u(s_2,\tilde{X}_{s_2}^{s_1,\tilde{X}_{s_1}^{t,x}})\mathscr{V}_{s_1,s_2}^{s_1,\tilde{X}_{s_1}^{t,x}}\mathrm{d}s_2\Bigg]. \end{aligned}$$

This procedure can be iteratively repeated to obtain

$$u(t,x) = \sum_{k=0}^n u^{(k)}(t,x)+\tilde{u}^{(n+1)}(t,x), \tag{6}$$

where

$$\begin{aligned} u^{(0)}(t,x) &:= \mathbb{E}\big[g(\tilde{X}_T^{t,x})\big], \\ u^{(k)}(t,x) &:= \mathbb{E}\left[\int_t^T\int_{s_1}^T\cdots\int_{s_{k-1}}^T\prod_{l=1}^{k-1}\Theta_{s_l,s_{l+1}}^{(l)}\Theta_{s_k,T}^{(k)}g(\tilde{X}_T^{(k)})\mathrm{d}s_{1:k}\right], \quad k\geq 1, \text{ and,} \qquad (7)\\ \tilde{u}^{(n+1)}(t,x) &:= E\left[\int_t^T\int_{s_1}^T\cdots\int_{s_{n-1}}^T\int_{s_n}^T\prod_{l=1}^n\Theta_{s_l,s_{l+1}}^{(l)}H^{(n)}(s_{n+1},\tilde{X}_{s_{n+1}}^{(n)})\mathrm{D}^2u(s_{n+1},\tilde{X}_{s_{n+1}}^{(n)})\mathrm{d}s_{1:(n+1)}\right]. \qquad (8) \end{aligned}$$

In the above, we have employed the following notation:

$$\begin{aligned} s_0 := t, \qquad \tilde{X}_{s_0}^{(-1)} &:= x, \qquad \tilde{X}_{s_0}^{(0)} := x, \\ \tilde{X}_s^{(l-1)} &:= \tilde{X}_{s_{l-1}}^{(l-2)}+\sigma(s_{l-1},\tilde{X}_{s_{l-1}}^{(l-2)})(W_s-W_{s_{l-1}}), && 1\leq l\leq k,\, s\geq s_{l-1}, \\ H^{(l)}(s_{l+1},\tilde{X}_{s_{l+1}}^{(l)}) &:= a(s_{l+1},\tilde{X}_{s_{l+1}}^{(l)})-a(s_l,\tilde{X}_{s_l}^{(l-1)}), && 0\leq l\leq k, \\ \mathscr{V}_{s_l,s}^{(l)} &:= \sigma(s_l,\tilde{X}_{s_l}^{(l-1)})^{-2}\frac{(\delta W_{s_l,s})^2-(s-s_l)}{(s-s_l)^2}, && 1\leq l\leq k,\, s\geq s_l, \\ \Theta_{s_l,s}^{(l)} &:= H^{(l-1)}(s_l,\tilde{X}_{s_l}^{(l-1)})\cdot\mathscr{V}_{s_l,s_{l+1}}^{(l)}, && 1\leq l\leq k,\, s\geq s_l. \end{aligned}$$

The formula in (6) is an infinite series where $\sum_{k=0}^{n} u^{(k)}(t,x)$ and $\tilde{u}^{(n+1)}(t,x)$ are the partial sums. We obtain a representation of $u(t,x)$ in terms of a convergent infinite series by showing that $\tilde{u}^{(n+1)}(t,x)$ is absolutely convergent (see Rudin (1964), Theorem 3.22). We have due to the uniform bound $C_2 := |\mathrm{D}^2 u|_\infty$ for any $(t,x) \in [0,T]\times\mathbb{R}$,

$$\begin{aligned}
|\tilde{u}^{(n+1)}(t,x)| &\leq E\left[\int_t^T \int_{s_1}^T \cdots \int_{s_{n-1}}^T \int_{s_n}^T \prod_{l=1}^{n} \left|\Theta^{(l)}_{s_l,s_{l+1}}\right| \left|H^{(n)}(s_{n+1},\tilde{X}^{(n)}_{s_{n+1}})\right| \left|\mathrm{D}^2 u(s_{n+1},\tilde{X}^{(n)}_{s_{n+1}})\right| \mathrm{d}s_{1:(n+1)}\right] \\
&\leq C_2 L^{n+1} C^n \int_t^T \int_{s_1}^T \cdots \int_{s_{n-1}}^T \int_{s_n}^T \sqrt{s_1-t} \prod_{l=2}^{n} \frac{1}{\sqrt{s_l-s_{l-1}}} \frac{1}{\sqrt{s_{n+1}-s_n}} \mathrm{d}s_{1:(n+1)} \\
&= C_2 L^{n+1} C^n B(1/2,1) \int_t^T \int_{s_1}^T \cdots \int_{s_{n-1}}^T \sqrt{s_1-t} \prod_{l=2}^{n-1} \frac{1}{\sqrt{s_l-s_{l-1}}} \frac{\sqrt{T-s_n}}{\sqrt{s_n-s_{n-1}}} \mathrm{d}s_{1:n} \\
&= C_2 L^{n+1} C^n B(1/2,1) B(1/2,3/2) \int_t^T \int_{s_1}^T \cdots \int_{s_{n-2}}^T \sqrt{s_1-t} \prod_{l=2}^{n-2} \frac{1}{\sqrt{s_l-s_{l-1}}} \frac{T-s_{n-1}}{\sqrt{s_{n-1}-s_{n-2}}} \mathrm{d}s_{1:n-1} \\
&= (T-t)^{3/2+n/2} C_2 L^{n+1} C^n B(3/2,n/2+1) \prod_{l=2}^{n+1} B(1/2,1+(n+1-l)/2) \\
&= (T-t)^{3/2+n/2} C_2 L^{n+1} C^n B(3/2,n/2+1) \frac{\Gamma(1/2)^n}{\Gamma(n)} \to 0 \text{ as } n\to\infty.
\end{aligned}$$

Thus, the infinite series converges absolutely and uniformly for $(t,x) \in [0,T]\times\mathbb{R}$ and we have

$$u(t,x) = \sum_{k=0}^{\infty} u^{(k)}(t,x). \tag{9}$$

2.2 Finite Variance Representation for $g \in C^2$

Due to the presence of Malliavin weights $\left(\mathscr{V}^{(l)}_{s_l,s_{l+1}}\right)_{1\leq l\leq k}$ (5), it can be seen that each of the integrand in representation (7) explodes as $s_{l+1}\to s_l$ since each term is of the form $(s_{l+1}-s_l)^{-\varepsilon}$ for some $0<\varepsilon<1$. As a result, it is not straightforward to obtain estimators with finite variance unless we impose strong conditions on b,σ and g as done in Rhee and Glynn (2015). Most of the recent research has been focused to obtain schemes with finite variance without imposing severe regularity conditions. As discussed earlier, two new ideas have been recently introduced in (Bally and Kohatsu-Higa 2015) and (Henry-Labordere et al. 2017) which are further improved in (Andersson and Kohatsu-Higa 2017) and (Doumbia et al. 2017) respectively. The new modifications can be equivalently seen as an importance sampling change of measure approach. In this paper, we rather use the technique of change of variables for singular integrands of Helluy, Maire, and Ravel (1998) to evaluate the integrals present in (9). We propose the change of variable from s_k to r_k through the following transformation

$$P_{s_{k-1},T}(r_k) = s_k,\ 1\leq k\leq n, \tag{10}$$

where $P_{s_{k-1},T} : [s_{k-1},T] \to [s_{k-1},T]$ is a monotone and surjective map. In the following, we will also interchangeably use a short-hand notation $\rho_k(r_k)$ for transformation $P_{s_{k-1},T}(r_k)$, $1\leq k\leq n$. We perform the change of variables from $(s_k)_{1\leq k\leq n}$ to $(r_k)_{1\leq k\leq n}$ and get for $k\geq 1$,

$$u^{(k)}(t,x) = \mathbb{E}\left[\int_t^T \int_{\rho_1(r_1)}^T \cdots \int_{\rho_{k-1}(r_{k-1})}^T \prod_{l=1}^{k-1} \hat{\Theta}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})} \hat{\Theta}^{(k)}_{\rho_k(r_k),T} \prod_{l=1}^{k} P'_{\rho_{l-1}(r_{l-1}),T}(r_l) g(\hat{\mathrm{X}}^{(k)}_T) \mathrm{d}r_{1:k}\right]. \tag{11}$$

The new notations denote

$$
\begin{aligned}
& r_0 := t, \qquad \rho_0(r_0) := t, \qquad \hat{\mathrm{X}}^{(-1)}_{\rho_0(r_0)} := x, \qquad \hat{\mathrm{X}}^{(0)}_{\rho_0(r_0)} := x, \\
\hat{\mathrm{X}}^{(l-1)}_{\rho_l(r_l)} &:= \hat{\mathrm{X}}^{(l-2)}_{\rho_{l-1}(r_{l-1})} + \sigma(\rho_{l-1}(r_{l-1}), \hat{\mathrm{X}}^{(l-2)}_{\rho_{l-1}(r_{l-1})})(W_{\rho_l(r_l)} - W_{\rho_{l-1}(r_{l-1})}), \quad 1 \le l \le k, \\
\hat{H}^{(l)}(\rho_{l+1}(r_{l+1}), \hat{\mathrm{X}}^{(l)}_{\rho_{l+1}(r_{l+1})}) &:= a(\rho_{l+1}(r_{l+1}), \hat{\mathrm{X}}^{(l)}_{\rho_{l+1}(r_{l+1})}) - a(\rho_l(r_l), \hat{\mathrm{X}}^{(l-1)}_{\rho_l(r_l)}), \quad 0 \le l \le k, \\
\hat{\mathscr{V}}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})} &:= \sigma(\rho_l(r_l), \hat{\mathrm{X}}^{(l-1)}_{\rho_l(r_l)})^{-2} \\
& \quad \times \frac{(\delta W_{\rho_l(r_l),\rho_{l+1}(r_{l+1})})^2 - (\rho_{l+1}(r_{l+1}) - \rho_l(r_l))}{(\rho_{l+1}(r_{l+1}) - \rho_l(r_l))^2}, \quad 1 \le l \le k, \\
\hat{\Theta}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})} &:= \hat{H}^{(l-1)}(\rho_l(r_l), \hat{\mathrm{X}}^{(l-1)}_{\rho_l(r_l)}) \cdot \hat{\mathscr{V}}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})}, \quad 1 \le l \le k.
\end{aligned}
$$

Next, we denote by $\mathbf{R} = (R_1, R_2, \ldots, R_k)$ a k-dimensional random vector with the correct distribution such that it allows us to write the following formula:

$$
u^{(k)}(t,x) = c_k \mathbb{E}\left[\prod_{l=1}^{k-1} \hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})} \prod_{l=1}^{k} P'_{\rho_{l-1}(\mathbf{R}),1}(R_l) \hat{\Theta}^{(k)}_{\rho_k(R_k),T} g(\hat{\mathrm{X}}^{(k)}_T)\right], \quad k \ge 1, \tag{12}
$$

where $\rho_l(\mathbf{R})$ is also a shorthand notation for $\rho_l(R_l)$. In (12), the normalization constant c_k is given as

$$
c_k := \int_t^T \int_{\rho_1(r_1)}^T \cdots \int_{\rho_{k-1}(r_{k-1})}^T \mathrm{d}r_{1:k}. \tag{13}
$$

For an integer valued independent random variable N_T with probability mass function f which denotes the number of arrivals in the interval $[t,T]$, we further write

$$
u^{(k)}(t,x) = \mathbb{E}\left[c_{N_T} \prod_{l=1}^{N_T-1} \hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})} \prod_{l=1}^{N_T} P'_{\rho_{l-1}(\mathbf{R}),T}(R_l) \hat{\Theta}^{(N_T)}_{\rho_{N_T}(R_{N_T}),T} g(\hat{\mathrm{X}}^{(N_T)}_T) \middle| N_T = k\right], \quad k \ge 1. \tag{14}
$$

Then, for the following choice of probability mass function for $\lambda > 1$,

$$
f(n) = \mathbb{P}(N_T = n) = \frac{c_f}{n^{1+\lambda}}, n \ge 1, \text{ with } c_f^{-1} = \sum_{n \ge 1} n^{-1-\lambda} = \zeta(1+\lambda) \text{ (Riemann zeta function)}, \tag{15}
$$

we have the following result:

Theorem 1 Suppose Assumption 1, 2 and 3 hold. For an appropriate choice of transformation $(P'_{\rho_l(r_l),T})_{l \ge 0}$ and f as in (15), we have

$$
u(t,x) = \mathbb{E}\left[\frac{c_{N_T}}{f(N_T)} \prod_{l=1}^{N_T-1} \hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})} \prod_{l=1}^{N_T} P'_{\rho_{l-1}(\mathbf{R}),T}(R_l) \hat{\Theta}^{(N_T)}_{\rho_{N_T}(R_{N_T}),T} \Xi^{(N_T)}_T\right], \text{ with } \hat{\Theta}^{(0)}_{\rho_0(R_0),T} := 1, \tag{16}
$$

and

$$
\Xi^{(0)}_T := g(\hat{\mathrm{X}}^{(0)}_T), \quad \Xi^{(k)}_T := g(\hat{\mathrm{X}}^{(k)}_T) + g\big(\hat{\mathrm{X}}^{(k)}_T - 2(\hat{\mathrm{X}}^{(k)}_T - \hat{\mathrm{X}}^{(k-1)}_{R_k})\big) - 2g(\hat{\mathrm{X}}^{(k-1)}_{R_k}), \quad k \ge 1.
$$

Furthermore, the estimator has finite variance and finite expected computational cost.

The probabilistic representation (16) follows from the representation result in (9). The choice of appropriate transformation to achieve finite variance is provided in Remark 1. The proof for finite variance follows similar arguments as the proof of Theorem 2 and we thus omit it to avoid repetition. The expected computational cost is (up to a constant) given by $\sum_{n \ge 1} n \mathbb{P}(N_T = n)$ which remains finite for the choice of f in (15). Furthermore, we discuss about the feasible choices of f in Section 3.

2.3 Finite Variance Representation for Lipschitz continuous g

To obtain the probabilistic representation in this case, we first go back to the term $\tilde{u}^{(n+1)}(t,x)$ in the infinite series representation (6). For Lipschitz continuous g, we have the upper bound on the second derivative of the solution u as $|\mathrm{D}^2 u(t,x)| \le \frac{K\mathrm{e}^{|x|}}{\sqrt{T-t}}$ (see Gobet and Temam (2001)). Then, we get

$$\begin{aligned}|\tilde{u}^{(n+1)}(t,x)| &\le KL^{n+1}C^n \int_t^T \int_{s_1}^T \cdots \int_{s_{n-1}}^T \int_{s_n}^T \sqrt{s_1-t}\prod_{l=2}^{n} \frac{1}{\sqrt{s_l-s_{l-1}}} \frac{1}{\sqrt{s_{n+1}-s_n}} \frac{1}{\sqrt{T-s_{n+1}}} \mathrm{d}s_{1:(n+1)} \\ &= (T-t)^{1+n/2} KL^{n+1}C^n B(3/2,(n+1)/2) \prod_{l=2}^{n+1} B(1/2, 1/2+(n+1-l)/2) \to 0 \text{ as } n\to\infty.\end{aligned}$$

Thus, the series representation (9) still holds in the case of Lipschitz continuous g. For the discretized SDE sample path with Euler scheme $(\hat{\mathrm{X}}^{(l-1)}_{\rho_l(r_l)})_{1\le l\le n+1}$ for some $(r_0,r_1,\ldots,r_n,r_{n+1})$ where $r_0 := t$, $r_{n+1} := T$ and $\rho_{n+1}(r_{n+1}) := T$, we denote $\hat{\mathrm{Y}}$ as the *ghost path* which is based on the same Euler discretized Brownian motion except for the last time step component which is generated separately from the original discretized Brownian motion. Next, we use the *ghost path* and repeat the arguments preceding formula (16) to show that we have the following formula with the choice of f as in (15)

$$u(t,x) = \mathbb{E}\left[\frac{c_{N_T}}{f(N_T)} \prod_{l=1}^{N_T-1} \hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})} \prod_{l=1}^{N_T} P'_{\rho_{l-1}(\mathbf{R}),T}(R_l) \hat{\Theta}^{(N_T)}_{\rho_{N_T}(R_{N_T}),T} \Xi^{(N_T)}_T\right], \text{ with } \hat{\Theta}^{(0)}_{\rho_0(R_0),T} := 1, \tag{17}$$

and

$$\Xi^{(0)}_T := g(\hat{\mathrm{X}}^{(0)}_T), \quad \Xi^{(k)}_T := g(\hat{\mathrm{X}}^{(k)}_T) - g(\hat{\mathrm{Y}}^{(k)}_T), k\ge 1.$$

Now, we address the choice of appropriate change of variable transformation $\left(P_{\rho_l(r_l),T}\right)_{l\ge 0}$ to obtain a finite variance estimator. Recall that we need to change the variable for taking care of the singularity arising at the left boundary and right boundary of the time interval due to the Malliavin weights $\hat{\mathscr{V}}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})}$ in the terms $\hat{\Theta}^{(l)}_{\rho_l(r_l),\rho_{l+1}(r_{l+1})}$. Then, from formula (11), it is intuitively clear that we need to select $P'_{s_{l-1},T}$ such that

$$P'_{s_{l-1},T}(\rho_k^{-1}(s_l)) \sim (s_l - s_{l-1})^{\eta}(T-s_l)^{\nu}, \quad \eta,\nu > 0.$$

We also need the property that $P_{s_{l-1},T} : [s_{l-1},T] \to [s_{l-1},T]$ is a monotone surjective map. For any invertible function $P(x)$, we have

$$\frac{\mathrm{d}P^{-1}(x)}{\mathrm{d}x} = \frac{1}{P'(P^{-1}(x))}.$$

This gives us for the invertible map $P_{s_{l-1},T}$,

$$\rho_k^{-1}(s_l) \sim \int_{s_{l-1}}^{s_l} (y-s_{l-1})^{(1-\eta)-1}(T-y)^{(1-\nu)-1}\mathrm{d}y.$$

Thus, from the above heuristics, we define the transformation as

$$\begin{aligned} r_l = \rho_l^{-1}(s_l) &= (T-s_{l-1})\mathrm{I}(s_l; s_{l-1}, 1-\eta, 1-\nu) + s_{l-1}, \\ \left(P'_{s_{l-1},T}(\rho_l^{-1}(s_l))\right)^{-1} = \frac{\mathrm{d}\rho_l^{-1}(s_l)}{s_l} &= (T-s_{l-1})\frac{(s_l-s_{l-1})^{(1-\eta)-1}(T-s_l)^{(1-\nu)-1}}{(T-s_{l-1})^{1-(\eta+\nu)}B(1-\eta,1-\nu)} \\ &= (T-s_{l-1})^{\eta+\nu}\frac{(s_l-s_{l-1})^{(1-\eta)-1}(T-s_l)^{(1-\nu)-1}}{B(1-\eta,1-\nu)}, \end{aligned} \tag{18}$$

where

$$\begin{aligned}
\mathrm{I}(b;a,1-\eta,1-\nu) &:= \frac{B(b;a,1-\eta,1-\nu)}{B(T;a,1-\eta,1-\nu)},\\
B(b;a,1-\eta,1-\nu) &:= \int_a^b (y-a)^{(1-\eta)-1}(T-y)^{(1-\nu)-1}\mathrm{d}y,\\
B(T;a,1-\eta,1-\nu) &= (T-a)^{1-(\eta+\nu)}B(1-\eta,1-\nu),\\
B(x,y) &:= \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.
\end{aligned}$$

In the probabilistic representation of (12), this transform necessitates to compute the normalizing constant (13) for the distribution of k-dimensional random vector $\mathbf{R}$.

Lemma 1 For change of variable transformation $P_{s_{k-1},T}$ (18), the normalizing constant for distribution of k-dimensional random vector $\mathbf{R}$ in (12) is

$$c_n = (T-t)^n \frac{\prod_{k=1}^n B(1-\eta,k-\nu)}{\big(B(1-\eta,1-\nu)\big)^n}.$$

Proof. We prove the result by method of induction. From the definition of normalizing constant c_k in (13) and transformation $P_{s_{k-1},T}$ (18), we have for $n=1$,

$$\begin{aligned}
c_1 = \int_t^T \Big(P'_{t,T}(\rho_1^{-1}(s_1))\Big)^{-1}\mathrm{d}s_1 &= \int_t^T (T-t)^{\eta+\nu}\frac{(s_1-t)^{-\eta}(T-s_1)^{-\nu}}{B(1-\eta,1-\nu)}\mathrm{d}s_1\\
&= \frac{(T-t)}{B(1-\eta,1-\nu)}\int_0^1 z^{-\eta}(1-z)^{-\nu}\mathrm{d}z = T-t.
\end{aligned}$$

Next, we suppose that the induction hypothesis holds for some n. Then, we have

$$\begin{aligned}
c_{n+1} &= \int_t^T\int_{s_1}^T\cdots\int_{s_{n-1}}^T\int_{s_n}^T \prod_{l=1}^{n+1}\Big(P'_{s_{l-1},T}(\rho_l^{-1}(s_l))\Big)^{-1}\mathrm{d}s_{1:n+1}\\
&= \int_t^T (T-s_1)^n\frac{\prod_{k=1}^n B(1-\eta,k-\nu)}{\big(B(1-\eta,1-\nu)\big)^n}\times\frac{(T-t)^{\eta+\nu}(s_1-t)^{-\eta}(T-s_1)^{-\nu}}{B(1-\eta,1-\nu)}\mathrm{d}s_1\\
&= \frac{\prod_{k=1}^n B(1-\eta,k-\nu)}{\big(B(1-\eta,1-\nu)\big)^{n+1}}(T-t)^{n+1}\int_0^1 z^{-\eta}(1-z)^{n-\nu}\mathrm{d}z
\end{aligned}$$

which shows that the result holds for any n by the principle of mathematical induction. $\square$

3 MAIN RESULT

Theorem 2 Suppose Assumption 1, 2 and 4 hold. Then, if we use the appropriate transformation defined in (18) and f as in (15), the estimator in (17) has finite variance and finite expected computational cost.

Proof. Notice that if we have $u(t,x)=\mathbb{E}[\prod_{l=1}^{N_T}Z_l]$ for a random variable N_T which is independent of random variables $(Z_l)_{l\geq 1}$, we have that the variance $\mathrm{Var}(\prod_{l=1}^{N_T}Z_l)=\mathbb{E}[\mathrm{Var}(\prod_{l=1}^{N_T}Z_l|N_T)]+\mathrm{Var}(\mathbb{E}[\prod_{l=1}^{N_T}Z_l|N_T])$. Thus, to prove the claim, we are required to look at the following two sums:

$$\sum_{n=1}^{\infty}\mathbb{E}\left[\Big(\frac{c_n}{f(n)}\Big)^2\prod_{l=1}^{n-1}\Big(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\Big)^2\prod_{l=1}^{n}\Big(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\Big)^2\Big(\hat{\Theta}^{(n)}_{\rho_n(R_n),T}\Big)^2(\Xi_T^{(n)})^2\Big|N_T=n\right]\mathbb{P}(N_T=n), \tag{19}$$

$$\sum_{n=1}^{\infty}\left(\mathbb{E}\left[\frac{c_n}{f(n)}\prod_{l=1}^{n-1}\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\prod_{l=1}^{n}P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\hat{\Theta}^{(n)}_{\rho_n(R_n),T}\Xi^{(n)}_T\Big|N_T=n\right]\right)^2\mathbb{P}(N_T=n), \tag{20}$$

and show that each is finite. We focus on the term in (19) and the term in (20) follows similarly. For $n\geq 1$, consider the following term which appears in (19)

$$\mathbb{E}\left[\prod_{l=1}^{n-1}\left(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\right)^2\prod_{l=1}^{n}\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2\left(\hat{\Theta}^{(n)}_{\rho_n(R_n),T}\right)^2(\Xi^{(n)}_T)^2\Big|N_T=n\right]. \tag{21}$$

We further condition on the independently generated intermediate times $\mathbf{R}=(R_l)_{1\leq l\leq n}$ and use the shorthand notation $\mathbb{E}_{N_T,\mathbf{R}}[\cdot]=\mathbb{E}[\cdot|N_T,\mathbf{R}]$ to write

$$\begin{aligned}&\mathbb{E}_{N_T,\mathbf{R}}\left[\prod_{l=1}^{n-1}\left(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\right)^2\prod_{l=1}^{n}\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2\left(\hat{\Theta}^{(n)}_{\rho_n(R_n),T)}\right)^2(\Xi^{(n)}_T)^2\right]\\&=\mathbb{E}_{N_T,\mathbf{R}}\left[\prod_{l=1}^{n-1}\left(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\right)^2\prod_{l=1}^{n}\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2\mathbb{E}\left[\left(\hat{\Theta}^{(n)}_{\rho_n(R_n),T}\right)^2(\Xi^{(n)}_T)^2\Big|\mathscr{F}_{\rho_n(R_n)}\right]\right].\end{aligned}$$

We apply Cauchy-Schwarz inequality and then use Assumption 1 and 4 to obtain the following:

$$\mathbb{E}_{N_T,\mathbf{R}}\left[\left(\hat{\Theta}^{(n)}_{\rho_n(R_n),T}\right)^2(\Xi^{(n)}_T)^2\Big|\mathscr{F}_{\rho_n(R_n)}\right]\leq L\big(\rho_n(R_n)-\rho_{n-1}(R_{n-1})\big)\times\frac{C_g}{\big(T-\rho_n(R_n)\big)}$$

where C_g is Lipschitz constant of g. Next, we continue to iteratively take conditional expectation and similarly obtain upper bounds for the terms in the product (21) such as

$$\mathbb{E}_{N_T,\mathbf{R}}\left[\left(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\right)^2\Big|\mathscr{F}_{\rho_l(R_l)}\right]\leq L\big(\rho_l(R_l)-\rho_{l-1}(R_{l-1})\big)\times\frac{1}{\big(\rho_{l+1}(R_{l+1})-\rho_l(R_l)\big)^2}.$$

Then, we get

$$\begin{aligned}&\mathbb{E}_{N_T,\mathbf{R}}\left[\prod_{l=1}^{n-1}\left(\hat{\Theta}^{(l)}_{\rho_l(R_l),\rho_{l+1}(R_{l+1})}\right)^2\prod_{l=1}^{n}\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2\left(\hat{\Theta}^{(n)}_{\rho_n(R_n),T)}\right)^2(\Xi^{(n)}_T)^2\right]\\&\leq L^nC_g\prod_{l=1}^{n-1}\frac{\big(\rho_l(R_l)-\rho_{l-1}(R_{l-1})\big)}{\big(\rho_{l+1}(R_{l+1})-\rho_l(R_l)\big)^2}\prod_{l=1}^{n}\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2\times\frac{\big(\rho_n(R_n)-\rho_{n-1}(R_{n-1})\big)}{\big(T-\rho_n(R_n)\big)}.\end{aligned} \tag{22}$$

Thus by taking expectation over $\mathbf{R}$, we get the following upper bound on the term in (21),

$$\begin{aligned}&\mathbb{E}_{N_T}\left[L^nC_g\frac{(\rho_1(R_1)-t)\prod_{l=1}^n\left(P'_{\rho_{l-1}(\mathbf{R}),T}(R_l)\right)^2}{\prod_{l=2}^n\big(\rho_l(R_l)-\rho_{l-1}(R_{l-1})\big)}\big(T-\rho_n(R_n)\big)^{-1}\right]\\&=\frac{L^nC_g}{c_n}\int_t^T\dots\int_{\rho_{n-1}(r_{n-1})}^T\frac{(\rho_1(r_1)-t)\prod_{l=1}^n\left(P'_{\rho_{l-1}(\mathbf{r}),T}(r_l)\right)^2}{\prod_{l=2}^n\big(\rho_l(r_l)-\rho_{l-1}(r_{l-1})\big)}\big(T-\rho_n(r_n)\big)^{-1}\mathrm{d}r_{1:n}\\&=\frac{L^nC_g}{c_n}\int_t^T\dots\int_{s_{n-1}}^T\frac{(s_1-t)\prod_{l=1}^nP'_{s_{l-1},T}(\rho_l^{-1}(s_l))}{\prod_{l=2}^n(s_l-s_{l-1})}(T-s_n)^{-1}\mathrm{d}s_{1:n}.\end{aligned} \tag{23}$$

We collect the results in (23) for each $n \geq 1$ to see that the sum in (19) is upper bounded by

$$\sum_n \frac{L^n C_g}{f(n)} c_n I_n \tag{24}$$

where

$$I_n := \int_t^T \dots \int_{s_{n-1}}^T \frac{(s_1 - t)\prod_{l=1}^n P'_{s_{l-1},T}(\rho_l^{-1}(s_l))}{\prod_{l=2}^n (s_l - s_{l-1})} (T - s_n)^{-1} \mathrm{d}s_{1:n}. \tag{25}$$

For the multiple integral in (25), we obtain a formula in Lemma 2. Going back to the upper bound in (24), we have from the result in Lemma 1 and 2 that it equals

$$\begin{aligned}
&(T-t)\sum_n (T-t)^n \frac{L^n C_g}{f(n)} \frac{\prod_{l=1}^n B(1-\eta, l-\nu)}{\big(B(1-\eta,1-\nu)\big)^n} B(1-\eta,1-\nu)^n B^{n-1}(\eta,\nu) B(2+\eta,\nu) \\
&= (T-t)\sum_n \frac{\tilde{L}^n C_g}{f(n)} \prod_{k=1}^n B(1-\eta,k-\nu) B^{n-1}(\eta,\nu) B(2+\eta,\nu),
\end{aligned}$$

where we denote $\tilde{L} := (T-t)L$. For fixed x and large y, we know that Stirling's approximation for Beta function $B(x,y)$ is given as $B(x,y) \sim \Gamma(x) y^{-x}$. Thus, for large n, $\prod_{k=1}^n B(1-\eta,k-\nu) \sim \left(\frac{1}{(n-\nu)!}\right)^{(1-\eta)}$. Further, denote $C_{\eta,\nu} := B(\eta,\nu)$, $\tilde{C}_{\eta,\nu} := B(2+\eta,\nu)$. Then, the upper bound is given as

$$(T-t)\sum_n \frac{\tilde{L}^n C_{\eta,\nu}^{n-1} \tilde{C}_{\eta,\nu} C_g}{f(n)} \Big(\frac{1}{(n-\nu)!}\Big)^{(1-\eta)}. \tag{26}$$

For the choice of f as suggested in (15) (any f which makes the sum in (26) finite is acceptable), it follows from Stirling's approximation $n! \sim \sqrt{2\pi n}(n/\mathrm{e})^n$, that the upper bound is indeed finite as

$$(T-t)\sum_n \frac{\tilde{L}^n C_{\eta,\nu}^{n-1} n^{1+\lambda} \tilde{C}_{\eta,\nu} C_g}{\lambda} \Big(\frac{1}{(n-\nu)!}\Big)^{(1-\eta)} < \infty.$$

This concludes the proof. □

Lemma 2 For the change of variable transformation $P_{s_{l-1},1}$ as in (18), the multiple integral in (25) simplifies to

$$I_n = (T-t)\big(B(1-\eta,1-\nu)\big)^n B^{n-1}(\eta,\nu) B(2+\eta,\nu).$$

Proof. We have from the transformation in (18),

$$\begin{aligned}
I_n = (B(1-\eta,1-\nu))^n \int_t^T \dots \int_{s_{n-2}}^T \frac{(s_1-t)}{\prod_{l=2}^{n-1}(s_l - s_{l-1})} \prod_{l=1}^{n-1} \frac{(s_l - s_{l-1})^\eta (T-s_l)^\nu}{(T-s_{l-1})^{\eta+\nu}} \\
\times \frac{1}{(T-s_{n-1})^{\eta+\nu}} \mathrm{d}s_{1:n-1} \int_{s_{n-1}}^T (s_n - s_{n-1})^{-1+\eta} (T-s_n)^{-1+\nu} \mathrm{d}s_n.
\end{aligned}$$

Then, we first consider

$$\begin{aligned}
\int_{s_{n-1}}^T \frac{(s_n - s_{n-1})^\eta}{(s_n - s_{n-1})} \frac{(T-s_n)^\nu}{(T-s_n)} \mathrm{d}s_n &= \int_{s_{n-1}}^T (s_n - s_{n-1})^{\eta-1} (T-s_n)^{\nu-1} \mathrm{d}s_n \\
&= (T-s_{n-1})^{-1+(\eta+\nu)} B(\eta,\nu).
\end{aligned}$$

This gives us

$$I_n = \big(B(1-\eta,1-\nu)\big)^n B(\eta,\nu)\int_t^T \cdots \int_{s_{n-3}}^T \frac{(s_1-t)}{\prod_{l=2}^{n-2}(s_l-s_{l-1})}\times\prod_{l=1}^{n-2}\frac{(s_l-s_{l-1})^{\eta}(T-s_l)^{\nu}}{(T-s_{l-1})^{\eta+\nu}}$$
$$\times\frac{1}{(T-s_{n-2})^{\eta+\nu}}\mathrm{d}s_{1:n-2}\times\int_{s_{n-2}}^T\frac{(s_{n-1}-s_{n-2})^{\eta}}{(s_{n-1}-s_{n-2})}\frac{(T-s_{n-1})^{\nu}}{(T-s_{n-1})}\mathrm{d}s_{n-1}.$$

We continue to iterate to get

$$\begin{aligned} I_n &= B(1-\eta,1-\nu)^n B^{n-1}(\eta,\nu)\int_t^T (s_1-t)\frac{(s_1-t)^{\eta}(T-s_1)^{\nu}}{(T-t)^{\eta+\nu}}\times(T-s_1)^{-1}\mathrm{d}s_1 \\ &= (T-t)B(1-\eta,1-\nu)^n B^{n-1}(\eta,\nu)B(2+\eta,\nu). \end{aligned}$$

□

Remark 1 The choice of transformation to achieve finite variance in Theorem 1 for $0<\varepsilon<1$ is given as follows

$$\begin{aligned} s_k = P_{s_{k-1},T}(r_k) &= \left(\frac{r_k-s_{k-1}}{(T-s_{k-1})^{\varepsilon}}\right)^{\frac{1}{(1-\varepsilon)}}+s_{k-1}, \\ r_k &= (T-s_{k-1})^{\varepsilon}(s_k-s_{k-1})^{(1-\varepsilon)}+s_{k-1} \\ P'_{s_{k-1},T}(y) &= \frac{1}{(1-\varepsilon)}\left(\frac{y-s_{k-1}}{T-s_{k-1}}\right)^{\frac{\varepsilon}{(1-\varepsilon)}} \\ P'_{s_{k-1},T}(\rho_k^{-1}(s_k)) &= \frac{1}{(1-\varepsilon)}\left(\frac{s_k-s_{k-1}}{T-s_{k-1}}\right)^{\varepsilon}. \end{aligned}$$

The normalization constant $c_n=(T-t)^n(1-\varepsilon)^n\prod_{k=1}^n B(1-\varepsilon,k)$.

4 CONCLUSION

We provided a new finite variance unbiased estimation method with finite expected computational cost to calculate the expectation of a function of the solution of a stochastic differential equation (SDE). Our work extended the current results in the literature to the case of Lipschitz continuous SDE coefficients and Lipschitz continuous objective function. We first carefully handled the singular integrands in our probabilistic representation by introducing a variable transformation which cancels out the singularity in the transformed integrands. Finally, to show that our unbiased estimator has a finite variance, we generate an appropriate number of random intermediate time steps where the probability mass function of the number of steps is chosen in such a way that the expectation of the corresponding error contribution remains finite.

REFERENCES

Andersson, P., and A. Kohatsu-Higa. 2017. “Unbiased Simulation of Stochastic Differential Equations using Parametrix Expansions”. *Bernoulli* 23 (3): 2028–2057.

Bally, V., and A. Kohatsu-Higa. 2015. “A Probabilistic Interpretation of the Parametrix Method”. *The Annals of Applied Probability* 25 (6): 3095–3138.

Beskos, A., and G. O. Roberts. 2005. “Exact Simulation of Diffusions”. *The Annals of Applied Probability* 15 (4): 2422–2444.

Bompis, R., and E. Gobet. 2014. “Stochastic Approximation Finite Element method: Analytical formulas for Multidimensional Diffusion Process”. *SIAM Journal on Numerical Analysis* 52 (6): 3140–3164.

Broadie, M., and P. Glasserman. 1996. “Estimating Security Price Derivatives using Simulation”. *Management Science* 42 (2): 269–285.

Chen, N., and Z. Huang. 2013. "Localization and Exact Simulation of Brownian Motion-driven Stochastic Differential Equations". *Mathematics of Operations Research* 38 (3): 591–616.

Doumbia, M., N. Oudjane, and X. Warin. 2017. "Unbiased Monte Carlo Estimate of Stochastic Differential Equations Expectations". *ESAIM: PS* 21:56–87.

Fournié, E., J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. 1999. "Applications of Malliavin Calculus to Monte Carlo Methods in Finance". *Finance and Stochastics* 3 (4): 391–412.

Giles, M. B. 2008. "Multilevel Monte Carlo Path Simulation". *Operations Research* 56 (3): 607–617.

Gobet, E., and R. Munos. 2005. "Sensitivity Analysis using Itô-Malliavin Calculus and Martingales. Application to Stochastic Control Problem.". *SIAM Journal of Control and Optimization* 43 (5):1676–1713.

Gobet, E., and E. Temam. 2001. "Discrete Time Hedging Errors for Options with Irregular Payoffs.". *Finance and Stochastics* 5 (3):357–367.

Helluy, P., S. Maire, and P. Ravel. 1998. "Intégration Numérique d'Ordre Élevé de Fonctions Régulières ou Singulières sur un Intervalle". *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics* 327 (9): 843–848.

Henry-Labordere, P., X. Tan, and N. Touzi. 2017. "Unbiased Simulation of Stochastic Differential Equations". *To appear in the Annals of Applied Probability. arXiv preprint arXiv:1504.06107*.

Milstein, G. 1975. "Approximate Integration of Stochastic Differential Equations". *Theory of Probability & Its Applications* 19 (3): 557–562.

Platen, E., and P. E. Kloeden. 1992. " Numerical Solution of Stochastic Differential Equations." *Springer-Verlag Berlin Heidelber* 23.

Rhee, C.-H., and P. W. Glynn. 2015. "Unbiased Estimation with Square Root Convergence for SDE models". *Operations Research* 63 (5): 1026–1043.

Rudin, W. 1964. "Principles of Mathematical Analysis". *McGraw-Hill New York* 3.

AUTHOR BIOGRAPHIES

ANKUSH AGARWAL is a Post-doctoral Researcher at Centre of Applied Mathematics at École Polytechnique, Universtié Paris-Saclay (France). His research interests include Monte Carlo methods, Markov Chain Monte Carlo methods, stochastic analysis and mathematical finance. His research is part of the Chair *Financial Risks* of the *Risk Foundation*. His e-mail address is ankush.agarwal@polytechnique.edu.

EMMANUEL GOBET is Professor of Applied Mathematics at École Polytechnique, Universtié Paris-Saclay (France), and director of the post graduate Master degree in Probability and Finance. His research interests are in the fields of probabilistic algorithms and stochastic approximations, financial mathematics, Malliavin calculus and stochastic analysis, Monte Carlo simulations, statistical learning and statistics for stochastic processes. His research is part of the Chair *Financial Risks* of the Risk Foundation and the *Finance for Energy Market Research Centre*. His email address is emmanuel.gobet@polytechnique.edu.