Unbiased Monte Carlo estimate of stochastic differential equations expectations

Mahamadou Doumbia Nadia Oudjane * Xavier Warin *

July 5, 2017

Abstract

We propose an unbiased Monte Carlo method to compute $\mathbb{E}(g(X_T))$ where g is a Lipschitz function and X an Ito process. This approach extends the method proposed in [**?**] to the case where X is solution of a multidimensional stochastic differential equation with varying drift and diffusion coefficients. A variance reduction method relying on interacting particle systems is also developed.

Key words: unbiased estimate, linear parabolic PDEs, interacting particle systems
MSC2010: Primary 65C05, 60J60; secondary 60J85, 35K10

1 Introduction

Let $d \geq 1$ and W be a $d-$dimensional Brownian motion. We introduce the process X defined as the unique strong solution of the multi-dimensional Stochastic differential Equation (SDE) with coefficients satisfying the usual Lipschitz conditions :

$$\left\{ \begin{array}{lcl} dX_t^{0,x_0} & = & b(t, X_t^{0,x_0})dt + \sigma(t, X_t^{0,x_0})dW_t, \\ X_0^{0,x_0} & = & x_0, \end{array} \right. \tag{1.1}$$

where $b : [0,T] \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is the drift and $\sigma : [0,T] \times \mathbb{R}^d \longrightarrow \mathcal{S}^d$ is the diffusion of the process, $\mathcal{S}^d$ being the set of $d \times d$ dimensional matrices.

In this paper, we are interested in a Monte Carlo approach to compute an expectation of the form

$$u(t,x) := \mathbb{E}[g(X_T^{t,x})] \ . \tag{1.2}$$

When no explicit solution is available, the classical method to solve equation (1.2) consists in using a discretization scheme of (1.1) (for example the Euler scheme [**?**], the Milstein scheme [**?**], or the Burrage scheme [**?**]) and the error can be decomposed as a sum of an error due to the discretization time step δt and a statistical error of order $N^{-1/2}$ due to the Monte Carlo method for a number N of simulations.
In principle, this bias/variance tradeoff should carefully be adjusted in order to optimize the rate of convergence. This type of analysis has been conducted in [**?**] showing that, for instance with the the simple *Euler Monte Carlo* method, (using the Euler scheme to discretize

*EDF R&D & FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org)

the time), the best choice of time step δt as a function of the sample size N would lead to a rate of order $c_N^{-\frac{1}{3}}$, where $c_N = N/\delta t$ measures the computing time. Hence the combination of the bias and variance error deteriorates the standard rate $c_N^{-1/2}$, due to the statistical error, when X is easily simulatable and $c_N = N$. Moreover, in practice it is difficult to evaluate properly the bias error so that the optimal tradeoff is rarely practicable.
The *Multilevel Monte Carlo* (MLMC) method introduced in [**?**] is a way to improve the bias/variance tradeoff and to reduce the variance by combining several *Euler-Monte Carlo* estimates, associated with different time discretization steps. The idea is then to adjust judiciously the size of the sample simulated for each discretization *level*, in order to achieve a better rate of convergence.
This approach has been extended in [**?**] allowing for an infinite number of levels so that the bias vanishes. The estimate is then expressed as an infinite sum (over the levels), which is randomized by introducing a probability distribution driving the levels. However, when the order of the time discretization scheme is not sufficiently high, this method results in an infinite variance estimate. More precisely, as soon as the time discretization scheme implies a strong error greater than or equal to the order $\sqrt{\delta t}$, either the variance or the computing time blows up. Unfortunately this situation includes the case of the Euler scheme, which is so far the most widely usable discretization scheme in multidimensional cases.
This approach has been improved in [**?**], where the authors rely on the parametrix expansion presented in [**?**] to propose a finite variance estimate. More specifically, the parametrix method provides a precise expansion of the expected difference considered at two successive levels in terms of a difference between the infinitesimal generator, $\mathcal{L}$, associated with (1.1) and the one associated with the same SDE with frozen coefficients at a given point, as defined hereafter by (2.3). Finally, importance sampling is used to change the levels distribution in order to control the variance. These developments lead to the *backward simulation method* or the *forward simulation method*, depending on whether $\mathcal{L}$ or its adjoint is used to represent the expectation. The backward method consists in generating some independent and identically distributed (i.i.d.) Euler type discretizations of a process, at random discrete times, from time T to time t. The payoff function, g is used as initial distribution at time T and the estimator results from a weighted average over the trials that hit the initial point, x, at time t. Therefore, this approach requires the payoff g to be integrable and is limited to small dimensions (for which the probability of reaching a given point can efficiently be computed). The forward method consists in generating some i.i.d. Euler type discretizations of (1.1) at random discrete times from time t to T and then computing a weighted average of the payoff function evaluated at the final points. In both methods the weights depend on the drift and diffusion coefficients b and σ evaluated along the simulated path. However, the forward approach relies on a stronger regularity assumption on the SDE coefficients. In particular, the related weights involve the first derivatives of the drift and the first and second derivatives of the volatility function.
Another approach called *Exact simulation* was initialized in [**?**]. The idea relies on Lamperti transform to come down to a unit diffusion process. It has been extended to more general SDEs in for instance [**?**, **?**]. However, the Lamperti transform is limited to the one dimensional case and extensions to the multidimensional are still limited to some specific cases.

In this paper, we propose to extend a method originally developed in [**?**]. The main idea

developed in this seminal paper is to start by simulating exactly a SDE :

$$\left\{ \begin{array}{rcl} dY_t^{0,x_0} & = & \hat{b}(t, Y_t^{0,x_0})dt + \hat{\sigma}(t, Y_t^{0,x_0})dW_t \\ Y_0^{0,x_0} & = & x_0 \ , \end{array} \right.$$

where the coefficients $\hat{b}$ and $\hat{\sigma}$ are updated at independent exponential switching times. Then the change in coefficients in SDE (1.1) is taken into account in an expectation representation via weights derived from the automatic differentiation technique developed in [**?**]. By carefully choosing the coefficients $\hat{b}$, $\hat{\sigma}$, the authors were able to provide a finite variance method in the case where the diffusion coefficient is constant or with a general diffusion term but without drift and in dimension one. However, the variance of the resulting estimator is proved to be infinite in the most general case. One interest of this approach which is very similar to the forward parametrix representation [**?**, **?**] is that the weights do not involve any derivatives of the coefficients b or σ so that no differentiability assumptions on those coefficients is required. Besides, one major motivation of this type of approach goes beyond the scope of the present paper. The idea is to generalize the branching diffusion representation of nonlinear Partial Differential Equations (PDEs) considered in [**?**, **?**] to a more general class of nonlineariries. One step in that direction has already been done in [**?**] with an extension of the branching diffusion representation to a class of semilinear PDEs.
To bypass the infinite variance obstacle faced in [**?**], the idea developed in the present paper consists in extending the original framework to more general switching times and exploit the switching time distribution to control the estimator variance. Notice that the same idea has been independently investigated in [**?**] to control the variance of the parametrix representation proposed in [**?**]. We prove that under suitable assumptions on the switching times distribution, we can provide a finite variance estimate of the solution of (1.2) in the most general case with drift and diffusion coefficients both varying. For instance, the gamma distribution is proved to verify those assumptions as soon as the shape parameter κ satisfies $\kappa \leq \alpha \wedge \frac{1}{2}$, when the coefficients b and $a := \sigma\sigma^\top$ are supposed to be uniformly α-Hölder continuous w.r.t the time variable. Another contribution consists in proposing an original interacting particle scheme that helps to stabilize even more the estimator. This approach results in a new estimator combining both branching and interacting particle techniques. The new estimator is proved to be unbiased with finite variance. Finally, numerical tests confirm the interest of our new algorithm showing significant variance reduction in various examples.

2 Notations

Let $C_b^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ denote the set of continuously differentiable bounded functions with bounded derivatives of order 1 for the time variable and bounded derivatives up to order 2 for the space variable. Let $\mathcal{L}$ denote the infinitesimal generator associated with (1.1) such that for any sufficiently regular function $\varphi : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}$ in the domain of $\mathcal{L}$, $\mathcal{L}\varphi$ is given as the real valued function such that

$$(\mathcal{L}\varphi)(t,x) = b(t,x).D\varphi(t,x) + \frac{1}{2}a(t,x) : D^2\varphi(t,x) \ , \quad \text{for all } (t,x) \in [0,T] \times \mathbb{R}^d \ , \qquad (2.1)$$

where $a(t,x) := \sigma(t,x)\sigma(t,x)^\top$, $A : B := tr(AB^\top)$ and D (resp. D^2) denotes the differential operator of order 1 (resp. of order 2) w.r.t. the space variable x. Let us consider a real

valued Lipschitz continuous function g defined on $\mathbb{R}^d$. By the Feynman-Kac formula it is well-known that if there exists $v^* \in C_b^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ solution of the linear Partial Differential Equation (PDE)

$$\begin{cases} \partial_t v + \mathcal{L}v = 0 \\ v(T,x) = g(x)\ , \end{cases} \tag{2.2}$$

then this PDE has a unique classical solution $v^*(t,x) = u(t,x) = \mathbb{E}[g(X_T^{t,x})]$. In the sequel $\|x\|$ stands for the L_∞ norm of a vector or a matrix x.
First we introduce an intermediary assumption that will be relaxed for our main results:

Assumption 1. *The linear PDE* (2.2) *admits a unique classical solution* $v^* \in C_b^{1,2}$.

All along this paper, the following assumption will be in force.

Assumption 2.

1. *The diffusion* $\sigma(t,x)$ *is non-degenerated such that for some constant* $\epsilon_0 > 0$*:*
$$a(t,x) \geq \epsilon_0 \mathbb{I}, \quad \forall (t,x) \in [0,T]\times\mathbb{R}^d.$$

2. b *and* a *are uniformly Lipschitz w.r.t. the space variable i.e. there exists a finite constant* L *such that for any* $(t,x,x') \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d$
$$\|b(t,x) - b(t,x')\| + \|a(t,x) - a(t,x')\| \leq L\|x - x'\|\ .$$

3. *There exists* $\alpha \in (0,1]$ *such that* b *and* a *are uniformly* α*-Hölder continuous w.r.t. variable* t *i.e. there exists a finite constant* H *such that for any* $(t,t',x) \in [0,T]\times[0,T]\times\mathbb{R}^d$
$$\|b(t,x) - b(t',x)\| + \|a(t,x) - a(t',x)\| \leq H|t - t'|^\alpha\ .$$

For a fixed point $(\tilde{t},\tilde{x}) \in [0,T]\times\mathbb{R}^d$, we introduce some operators and processes that will be useful in the sequel

- $\mathcal{L}^{\tilde{t},\tilde{x}}$ the differential operator similar to $\mathcal{L}$ with the drift and diffusion frozen at $(\tilde{t},\tilde{x})$ such that for any regular function φ in the domain of $\mathcal{L}^{\tilde{t},\tilde{x}}$
$$\mathcal{L}^{\tilde{t},\tilde{x}}\varphi(t,x) = b(\tilde{t},\tilde{x}).D\varphi(t,x) + \frac{1}{2}a(\tilde{t},\tilde{x}) : D^2\varphi(t,x)\ , \quad \text{for all } (t,x) \in [0,T]\times\mathbb{R}^d\ , \tag{2.3}$$

- $(\tilde{X}_t^{\tilde{t},\tilde{x},t_0,x_0})_{t\geq t_0}$ the Gaussian process with infinitesimal operator $\mathcal{L}^{\tilde{t},\tilde{x}}$ defined by
$$\tilde{X}_t^{\tilde{t},\tilde{x},t_0,x_0} = x_0 + b(\tilde{t},\tilde{x})(t - t_0) + \sigma(\tilde{t},\tilde{x})(W_t - W_{t_0})\ . \tag{2.4}$$
for a given initial condition $(t_0,x_0) \in [0,T]\times\mathbb{R}^d$.

- $h^{*,\tilde{t},\tilde{x}} : [0,T]\times\mathbb{R}^d \mapsto \mathbb{R}$ involving the unique solution v^* of (2.2) is defined by
$$h^{*,\tilde{t},\tilde{x}}(t,x) := (b(t,x) - b(\tilde{t},\tilde{x})).Dv^*(t,x) + \frac{1}{2}(a(t,x) - a(\tilde{t},\tilde{x})) : D^2v^*(t,x)\ . \tag{2.5}$$
Notice that $h^{*,\tilde{t},\tilde{x}}$ is a well defined continuous function since $v^* \in C_b^{1,2}$ and in particular
$$h^{*,\tilde{t},\tilde{x}}(t,x) = \mathcal{L}v^*(t,x) - \mathcal{L}^{\tilde{t},\tilde{x}}v^*(t,x) \quad \text{for all } (t,x) \in [0,T]\times\mathbb{R}^d\ . \tag{2.6}$$

3 Probabilistic representation using a regime switching process

Recalling [**?**], the following representation holds

Lemma 3.1. *Suppose that Assumptions 1 and 2 hold and $\tilde{X}^{\tilde{t},\tilde{x}}$ is the Gaussian process defined in (2.4), then u defined by (1.2) and its (bounded and continuous) derivatives Du and D^2u are solutions of the system*

$$\begin{array}{rcl} u(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x}) + \int_t^T H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}))\, ds] \\ Du(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathcal{M}_{t,T}^{\tilde{t},\tilde{x}} + \\ & & \int_t^T H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}))\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}}\, ds] \\ D^2u(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{t,x,t,x})\mathcal{V}_{t,T}^{t,x} + \\ & & \int_t^T H^{t,x}(s,\tilde{X}_s^{t,x,t,x}, Du(s,\tilde{X}_s^{t,x,t,x}), D^2u(s,\tilde{X}_s^{t,x,t,x}))\mathcal{V}_{t,s}^{t,x}\, ds]\ , \end{array} \tag{3.1}$$

where for any $(\tilde{t},\tilde{x}) \in [0,T]\times\mathbb{R}^d$ the function $H^{\tilde{t},\tilde{x}} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathcal{S}^d \mapsto \mathbb{R}$ is such that

$$H^{\tilde{t},\tilde{x}}(t,x,y,z) = (b(t,x) - b(\tilde{t},\tilde{x})).y + \frac{1}{2}(a(t,x) - a(\tilde{t},\tilde{x})) : z\ , \tag{3.2}$$

$\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}}$ and $\mathcal{V}_{t,s}^{\tilde{t},\tilde{x}}$ are respectively the first and second order Malliavin weights associated with the process $\tilde{X}^{\tilde{t},\tilde{x}}$ that is using $\delta_{t,s}W = W_s - W_t$

$$\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}} := (\sigma(\tilde{t},\tilde{x})^{-1})^\top \frac{\delta_{t,s}W}{s-t}\ , \quad \textit{and} \quad \mathcal{V}_{t,s}^{\tilde{t},\tilde{x}} := (\sigma(\tilde{t},\tilde{x})^{-1})^\top \frac{\delta_{t,s}W\delta_{t,s}W^\top - (s-t)\mathbb{I}}{(s-t)^2}\sigma(\tilde{t},\tilde{x})^{-1}\ . \tag{3.3}$$

Proof. The proof relies on the uniqueness property of classical solutions of PDEs satisfying the Feynman-Kac representation. Notice that under Assumptions 1 2, u is the unique classical solution of (2.2). Of course, thanks to equation (2.6), for any $(\tilde{t},\tilde{x}) \in [0,T]\times\mathbb{R}^d$ u is also a $C_b^{1,2}$ solution of the following linear PDE

$$\partial_t u + \mathcal{L}^{\tilde{t},\tilde{x}}u + h^{*,\tilde{t},\tilde{x}} = 0\ .$$

Then one can use again Feynman-Kac formula to represent the unique solution u of the above PDE as

$$u(t,x) = \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x}) + \int_t^T h^{*,\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x})\, ds]\ . \tag{3.4}$$

Finally observe that

$$h^{*,\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}) = H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x}}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}))\ . \tag{3.5}$$

The equations relative to Dv and D^2v are obtained by applying Elworthy's formula [**?**] (which simply results here in the Likelihood ratio of Broadie and Glasserman [**?**]) in (3.4) and by using some technical estimates placed in the Appendix **??** to be able to differentiate under the time integral.

□

Let τ be a random time independent of the Brownian W following the density f supposed to be strictly positive on $[0,\infty]$ and $\mathcal{P}[\tau > T] > 0$. One can rewrite representation (3.1) by using a change of measure to replace the time integral by an expectation according to the random time τ.

$$\begin{aligned}
u(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathbf{1}_{\tau\geq T-t}]}{1-F(T-t)} \\
&\quad +\mathbb{E}[\frac{H^{\tilde{t},\tilde{x}}(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x},Du(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}),D^2u(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}))}{f(\tau)}\,\mathbf{1}_{\tau<T-t}] \\
Du(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathcal{M}_{t,T}^{\tilde{t},\tilde{x}}\mathbf{1}_{\tau\geq T-t}]}{1-F(T-t)} \\
&\quad +\mathbb{E}[\frac{H^{\tilde{t},\tilde{x}}(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x},Du(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}),D^2u(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}))}{f(\tau)}\mathcal{M}_{t,t+\tau}^{\tilde{t},\tilde{x}}\,\mathbf{1}_{\tau<T-t}] \\
D^2u(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{t,x,t,x})\mathcal{V}_{t,T}^{t,x}\mathbf{1}_{\tau\geq T-t}]}{1-F(T-t)} \\
&\quad +\mathbb{E}[\frac{H^{t,x}(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x},Du(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x}),D^2u(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x}))}{f(\tau)}\mathcal{V}_{t,t+\tau}^{t,x}\,\mathbf{1}_{\tau<T-t}]
\end{aligned} \tag{3.6}$$

where F is the cumulative distribution of f. We will now apply recursively this representation (3.6) by considering a sequence of i.i.d. random times (τ_k).

Let us introduce a non regular (stochastic) mesh of the interval $[t_0, T]$,

$$\pi := (T_0 := t_0 < T_1 < \cdots < T_k \cdots < T_{N_T} < T_{N_T+1} = T) \ , \tag{3.7}$$

characterized by the Markov chain (T_k) defined by

$$\left\{\begin{array}{lcl} T_0 &=& t_0 \\ T_{k+1} &=& T_k + \delta T_{k+1} \text{ , for } k \in \mathbb{N} \text{ where} \\ \delta T_{k+1} &=& \tau_{k+1} \wedge (T - (T_k + \tau_{k+1}))^+ \ , \end{array}\right. \tag{3.8}$$

where (τ_k) is an i.i.d. sequence of random times distributed according the common probability density f. Notice that (T_k) defines a Markov chain with an absorbing state, T. (T_k) will define the so-called *switching time*.
The random integer N_T is defined as the following stopping time

$$N_T = \inf\{n \mid T_{n+1} \geq T\} \ . \tag{3.9}$$

Now notice by the law of large numbers that $\frac{1}{n}\sum_{k=1}^n \tau_k \longrightarrow \mathbb{E}[\tau] > 0$ so $\sum_{k=1}^n \tau_k \longrightarrow \infty$ almost surely so N_T is almost surely finite. In the sequel, we will consider an i.i.d. sequence (τ_k) of gamma variables with parameters $(\kappa > 0, \theta > 0)$ recalling that the gamma density with parameter $(\kappa > 0, \theta > 0)$ is given by

$$f_\Gamma^{\kappa,\theta}(s) = \frac{s^{\kappa-1}e^{-s/\theta}}{\Gamma(\kappa)\theta^\kappa} \ , \quad \text{for all } s > 0 \ , \tag{3.10}$$

where Γ is the gamma Euler function.

For a given mesh π (defined as in (3.7) (3.8)), we consider the following sequence (defining a Markov chain conditionally to the mesh π)

$$\left\{ \begin{array}{l} \bar{X}_0 = X_{T_0}^{t_0,x_0} = x_0 \\ \bar{X}_{k+1} = \bar{X}_k + b(T_k, \bar{X}_k)\delta T_{k+1} + \sigma(T_k, \bar{X}_k)\delta W_{k+1} \ , \end{array} \right. \tag{3.11}$$

where $\delta W_{k+1} := W_{T_{k+1}} - W_{T_k}$. For the sake of simplicity, we will often note b_k or σ_k instead of $b(T_k, \bar{X}_k)$ or $\sigma(T_k, \bar{X}_k)$.

Using representation (3.6) with $(\tilde{t}, \tilde{x}) = (T_k, \bar{X}_k)$ and $\tau = \tau_{k+1}$, conditioning with respect to $(\tilde{t}, \tilde{x})$ one gets for any integer $k \geq 0$

$$u(T_k, \bar{X}_k) \ = \ \frac{\mathbb{E}[g(\tilde{X}_T^{T_k,\bar{X}_k})\mathbf{1}_{T_{k+1}=T}]}{1 - F(T - T_k)} + \mathbb{E}[H_{k+1}\,\mathbf{1}_{T_{k+1}<T}] \tag{3.12}$$

with $\tilde{X}_s^{T_k,X_k} := \tilde{X}_s^{T_k,X_k,T_k,X_k}$ for $s \geq T_k$ and

$$H_{k+1} := \frac{H^{T_k,\bar{X}_k}(T_{k+1}, \bar{X}_{k+1}, Du(T_{k+1}, \bar{X}_{k+1}), D^2u(T_{k+1}, \bar{X}_{k+1}))}{f(\delta T_{k+1})} \ .$$

The derivatives Du and D^2u in H_{k+1} are given by applying the representation (3.6) with $(\tilde{t}, \tilde{x}) = (T_{k+1}, \bar{X}_{k+1})$ and $\tau = \tau_{k+2}$, conditioning with respect to $(\tilde{t}, \tilde{x})$ one gets for any integer $k \geq 0$

$$\begin{array}{rcl} Du(T_{k+1}, \bar{X}_{k+1}) & = & \dfrac{\mathbb{E}[g(\tilde{X}_T^{T_{k+1},\bar{X}_{k+1}})\mathcal{M}_{T_{k+1},T}^{T_{k+1},\bar{X}_{k+1}}\mathbf{1}_{T_{k+2}=T}]}{1 - F(T - T_{k+1})} + \mathbb{E}[H_{k+2}\mathcal{M}_{T_{k+1},T_{k+2}}^{T_{k+1},\bar{X}_{k+1}}\,\mathbf{1}_{T_{k+2}<T}] \\ Du(T_{k+1}, \bar{X}_{k+1}) & = & \dfrac{\mathbb{E}[g(\tilde{X}_T^{T_{k+1},\bar{X}_{k+1}})\mathcal{V}_{T_{k+1},T}^{T_{k+1},\bar{X}_{k+1}}\mathbf{1}_{T_{k+2}=T}]}{1 - F(T - T_{k+1})} + \mathbb{E}[H_{k+2}\mathcal{V}_{T_{k+1},T_{k+2}}^{T_{k+1},\bar{X}_{k+1}}\,\mathbf{1}_{T_{k+2}<T}] \ . \end{array}$$

Let us introduce the sequence of weights $(P_k)_{k\geq 1}$ such that for $k = 1, \cdots, N_T$

$$\left\{ \begin{array}{lcl} P_{k+1} & = & \frac{M_{k+1} + \frac{1}{2}V_{k+1}}{f(\delta T_k)} \ , \\ M_{k+1} & = & \delta b_k.(\sigma_k^{-1})^\top \frac{\delta W_{k+1}}{\delta T_{k+1}}, \quad \text{with} \quad \delta b_k := b_k - b_{k-1} \\ V_{k+1} & = & \delta a_k : (\sigma_k^{-1})^\top \frac{\delta W_{k+1}\delta W_{k+1}^\top - \delta T_{k+1}\mathbb{I}}{(\delta T_{k+1})^2}\sigma_k^{-1} \ , \quad \text{with} \quad \delta a_k := a_k - a_{k-1} \ . \end{array} \right. \tag{3.13}$$

Following the same lines as the proof of Theorem 2.2 in [**?**], one can derive by recurrence a representation formula for u as the expectation of an exactly simulatable variable. Before one has to introduce some new assumptions.

Assumption 3. *The coefficients b and a are uniformly bounded i.e. there exists a finite constant M such that for any $(t, x) \in [0, T] \times \mathbb{R}^d$*

$$\|b_t(x)\| \leq M \ , \quad \|a_t(x)\| \leq M \ .$$

Assumption 4. *The function g is Lipschitz.*

Proposition 3.2. *Under assumptions 2, 3 and 4 , the following representation holds*

$$u(t_0,x_0) \quad := \quad \mathbb{E}[g(X_T^{t_0,x_0})] = \mathbb{E}[\tfrac{g(\bar{X}_{N_T+1})}{1-F(\delta T_{N_T+1})}\prod_{k=2}^{N_T+1} P_k] \ , \tag{3.14}$$

with the convention $\prod_{k\in\varnothing} = 1$.

Remark 3.1.

1. *Proposition 3.2 proves that any v satisfying the equation* (3.1) *is given by the above explicit equation* (3.14)*: this a posteriori proves the uniqueness of the solution of* (3.1)*.*
2. *Using an exponential distribution for* f*, one recovers the representation given in [?].*
3. *Representation* (3.14) *is very similar to the forward representation developed in [?] and used in [?] for the same purpose of unbiased simulation of SDEs expectations. The main interest of* (3.14) *being that the coefficients b and a have no need to be differentiable.*

Proof. We will only give the sketch of the proof since it mimics step by step the proof of Theorem 2.2 in [?], which proceeds into two steps.

1. First suppose that Assumption 1 is satisfied, then the representation (3.14) holds. Indeed, Lemma 3.1 applies and following the recurrence arguments developed in [?] implies the representation (3.14).

2. For the clarity of the paper we recall here the arguments developed in [?] to extend, by smooth approximations, the representation proved at item 1. outside of Assumption 1. Since according to assumptions 2 and 4, (b,σ,g) are Lipschtiz we can find a sequence of bounded smooth functions $(b^\epsilon,\sigma^\epsilon,g^\epsilon)$ converging locally uniformly to (b,σ,g) as $\epsilon \longrightarrow 0$ such that Assumption 1 is verified when replacing $\mathcal{L}$ by $\mathcal{L}^\epsilon$ (the infinitesimal generator associated to $(b^\epsilon,\sigma^\epsilon)$) in the PDE (2.2). Let X^ϵ denote the solution of

$$dX_t^\epsilon = b^\epsilon(t,X_t^\epsilon)dt + \sigma^\epsilon(t,X_t^\epsilon)dW_t$$

and set $u^\epsilon(t_0,x_0) := \mathbb{E}[g^\epsilon(X_T^\epsilon)]$. By item 1. The following representation holds

$$u^\epsilon(t_0,x_0) = \psi^\epsilon := \mathbb{E}[\frac{g^\epsilon(\bar{X}^\epsilon_{N_T+1})}{1-F(\delta T_{N_T+1})}\prod_{k=2}^{N_T+1} P_k^\epsilon] \ ,$$

where $(\bar{X}_k^\epsilon)$ (and respectively the weights (P_k^ϵ)) are given by (3.11) (resp. the recursion (3.13)), where X is replaced by X^ϵ. By stability of SDEs, and dominated convergence theorem, $u^\epsilon(t_0,x_0) \xrightarrow[\epsilon\to 0]{} u(t_0,x_0)$. Similarly one can prove that $\psi^\epsilon \xrightarrow[\epsilon\to 0]{} \mathbb{E}[\frac{g(\bar{X}_{N_T+1})}{1-F(\delta T_{N_T+1})}\prod_{k=2}^{N_T+1} P_k]$, which ends the proof.

□

We next define a second representation that will be interesting in order to get some finite variance estimator for some given switching distribution, f. Following [?], one can

introduce antithetic variables to control the variance induced by the last time step. Let $\mathcal{G}_T := \sigma(N_T, T_k, \Delta W_k \mathbf{1}_{k\leq N_T}, k \geq 1)$. Observe that

$$\mathbb{E}[M_{N_T+1} \,|\mathcal{G}_T] = \mathbb{E}[V_{N_T+1} \,|\mathcal{G}_T] = \mathbb{E}[P_{N_T+1} \,|\mathcal{G}_T] = 0 \ .$$

Hence replacing $g(\bar{X}_{N_T+1})$ by $g(\bar{X}_{N_T+1}) - g(\bar{X}_{N_T})$ in (3.14) does not change the expectation since due to the tower property:

$$\mathbb{E}[\frac{g(\bar{X}_{N_T})}{1-F(\delta T_{N_T+1})} \prod_{k=2}^{N_T+1} P_k] = \mathbb{E}[\frac{g(\bar{X}_{N_T})}{1-F(\delta T_{N_T+1})} P_{N_T+1} \prod_{k=2}^{N_T} P_k] = 0 \ .$$

Notice that the following decomposition holds whenever $N_T \geq 1$

$$g(\bar{X}_{N_T+1}) \prod_{k=2}^{N_T+1} P_k = g(\bar{X}_{N_T+1}) \frac{M_{N_T+1}}{f(\delta T_{N_T})} \prod_{k=2}^{N_T} P_k + \frac{1}{2} g(\bar{X}_{N_T+1}) \frac{V_{N_T+1}}{f(\delta T_{N_T})} \prod_{k=2}^{N_T} P_k \ .$$

Then using antithetic variables for the second term in the r.h.s. of the above equality yields the following estimator.

Proposition 3.3. *Under assumptions 2, 3 and 4 , the following representation holds*

$$u(t_0, x_0) = \mathbb{E}[\beta \prod_{k=2}^{N_T} P_k \mathbf{1}_{N_T \geq 1}] + \mathbb{E}[\frac{g(\bar{X}_1)}{1-F(\delta T_1)} \mathbf{1}_{N_T=0}] \ , \tag{3.15}$$

where $\beta := \frac{1}{2}(\beta_1 + \beta_2)$ *with*

$$\begin{cases} \beta_1 & := & \frac{g(\bar{X}_{N_T+1}) - g(\bar{X}_{N_T})}{1-F(\delta T_{N_T+1})} \frac{M_{N_T+1} + \frac{1}{2} V_{N_T+1}}{f(\delta T_{N_T})}, \\ \beta_2 & := & \frac{g(\hat{X}_{N_T+1}) - g(\bar{X}_{N_T})}{1-F(\delta T_{N_T+1})} \frac{-M_{N_T+1} + \frac{1}{2} V_{N_T+1}}{f(\delta T_{N_T})} \end{cases} \tag{3.16}$$

and for any $n \in \mathbb{N}$, $\hat{X}_{n+1} = \bar{X}_n + b_n \delta T_{n+1} - \sigma_n \delta W_{n+1}$.

4 Variance Analysis in the case of Gamma distribution

The previous representation given by Proposition 3.3 is general but the variance associated to the estimator is generally infinite as it is the case when f is an exponential density. From now on, we will suppose that the density $f = f_\Gamma^{\kappa,\theta}$ is the Gamma density (3.10) with parameters (κ, θ) with cumulative distribution $F = F_\Gamma^{\kappa,\theta}$.
First, we will introduce the following assumptions.

Assumption 5. *The following assertions hold*

1. *g is Lipschitz and $g \in C^2$.*
2. *$\kappa \leq \alpha \wedge \frac{1}{2}$.*

Now, we can state the following proposition.

Proposition 4.1. *Under Assumption 2, 3 and 5, the estimator defined by* (3.15) *in Proposition 3.3 has finite variance.*

Proof. Let $\bar{\mathcal{F}}_k$ denote the sigma-field generated by the Brownian up to the random time T_k and the random times up to the random time T_{k+1} i.e. $\bar{\mathcal{F}}_k := \sigma(T_1, ..., T_{k+1}, (W_s)_{s \leq T \wedge T_k})$. Let us consider the second term on the r.h.s of (3.15). Notice that $\mathbb{E}[(g(\bar{X}_1))^2]$ can easily be bounded by the boundness assumptions on b and σ and the Lipschitz property of g. Let us consider the first term on the r.h.s. of (3.15).

$$\mathbb{E}[\Big(\beta \mathbf{1}_{N_T \geq 1} \prod_{k=2}^{N_T} P_k\Big)^2] \;=\; \sum_{n=1}^{\infty} \mathbb{E}[\Big(\beta \prod_{k=2}^{n} P_k\Big)^2 | N_T = n] \mathbb{P}(N_T = n) \tag{4.1}$$

The proof will be decomposed into several steps. We will first try to bound the general term of the above series $\mathbb{E}[\Big(\beta \prod_{k=2}^{n} P_k\Big)^2 | N_T = n]$, then we will consider the sum.

1. Bounding $\mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n, N_T = n]$
   First considering M_{k+1} and V_{k+1} one easily obtains

$$\begin{aligned} \mathbb{E}[M_{k+1}^4 \, | \bar{\mathcal{F}}_k, N_T = n] &\leq C \frac{(\delta b_k)^4}{(\delta T_{k+1})^2} \, , \\ \mathbb{E}[V_{k+1}^4 \, | \bar{\mathcal{F}}_k, N_T = n] &\leq C \frac{(\delta a_k)^4}{(\delta T_{k+1})^4} \, . \end{aligned} \tag{4.2}$$

   Notice that in the sequel, C will denote finite constants that may change from line to line that do not depend on k or n but only on the characteristics of the problem (T, the bounds or Lipschitz constants related to g, b, σ, a). Then consider the general term of the sum (4.1).

$$\mathbb{E}[\beta^2 \prod_{k=2}^{n} P_k^2 \, | N_T = n] \;=\; \mathbb{E}\Big[\mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n, N_T = n] \Big(\prod_{k=2}^{n} P_k \Big)^2 | N_T = n \Big] \, .$$

   We get

$$\begin{aligned} &\mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n, N_T = n] \\ &\leq \frac{C}{(1 - F_\Gamma^{\kappa,\theta}(T))^2} \mathbb{E}[(g(\bar{X}_{n+1}) - g(\hat{X}_{n+1}))^2 \frac{M_{n+1}^2}{f_\Gamma^{\kappa,\theta}(\delta T_n)^2} | \bar{\mathcal{F}}_n, N_T = n] + \\ &\frac{C}{(1 - F_\Gamma^{\kappa,\theta}(T))^2} \mathbb{E}[(g(\bar{X}_{n+1}) + g(\hat{X}_{n+1}) - 2g(\bar{X}_n))^2 \frac{V_{n+1}^2}{f_\Gamma^{\kappa,\theta}(\delta T_n)^2} | \bar{\mathcal{F}}_n, N_T = n] \end{aligned} \tag{4.3}$$

   Consider the first term on the r.h.s. of inequality (4.3), by the Lipschitz property of g, the boundness of b,σ and using the fact that σ is uniformly bounded away from zero, we obtain

$$\begin{aligned} &\mathbb{E}[|g(\bar{X}_{n+1}) - g(\hat{X}_{n+1})|^2 \frac{M_{n+1}^2}{(f_\Gamma^{\kappa,\theta}(\delta T_n))^2} | \bar{\mathcal{F}}_n, N_T = n] \\ &\qquad \leq C \frac{\|\delta b_n\|^2}{(f_\Gamma^{\kappa,\theta}(\delta T_n))^2} \leq C \|\delta b_n\|^2 (\delta T_n)^{2(1-\kappa)} \end{aligned} \tag{4.4}$$

Consider the second term of (4.3). By Assumption 5.2 ($g \in C^2$) one can apply Ito and obtain

$$|g(\bar{X}_{n+1}) + g(\bar{X}_n + b_n\delta T_{n+1} - \sigma_n \delta W_{n+1}) - 2g(\bar{X}_n)| \quad \leq \quad C\delta T_{n+1} \ .$$

This implies still using the boundness of b, σ and using the fact that σ is uniformly bounded away from zero :

$$\begin{aligned}\mathbb{E}[(g(\bar{X}_{n+1}) + g(\hat{X}_{n+1}) - 2g(\bar{X}_n))^2 \frac{V_{n+1}^2}{f_\Gamma^{\kappa,\theta}(\delta T_n)^2} |\bar{\mathcal{F}}_n, N_T = n] \\ \leq C(\delta T_{n+1})^2 \frac{\|\delta a_n\|^2}{\delta T_{n+1}^2 f_\Gamma^{\kappa,\theta}(\delta T_n)^2} \leq C(\delta T_n)^{2(1-\kappa)} \|\delta a_n\|^2 \ . \quad (4.5)\end{aligned}$$

Injecting (4.4) and (4.5) into (4.3) finally yields

$$\mathbb{E}[\beta^2 \, |\bar{\mathcal{F}}_n, N_T = n] \leq C(\delta T_n)^{2(1-\kappa)} \Big(\|\delta b_n\|^2 + \|\delta a_n\|^2 \Big) \ . \quad (4.6)$$

2. Bounding $\mathbb{E}[C_k^2 \, |\bar{\mathcal{F}}_{k-1}, N_T = n]$, where the r. v. C_k is defined by

$$C_k := \|\delta b_k\|^2 + \|\delta a_k\|^2 \ . \quad (4.7)$$

Consider the term $\|\delta b_k\|$,

$$\begin{aligned}\mathbb{E}[\|\delta b_k\|^4 \, |\bar{\mathcal{F}}_{k-1}, N_T = n] \\ &= \mathbb{E}[\|b(T_k, \bar{X}_k) - b(T_{k-1}, \bar{X}_{k-1})\|^4 \, |\bar{\mathcal{F}}_{k-1}, N_T = n] \\ &\leq 8\mathbb{E}\big[\|b(T_k, \bar{X}_k) - b(T_k, \bar{X}_{k-1})\|^4 + \\ &\qquad \|b(T_k, \bar{X}_{k-1}) - b(T_{k-1}, \bar{X}_{k-1})\|^4 \, |\bar{\mathcal{F}}_{k-1}, N_T = n\big] \\ &\leq C(1 + (\delta T_k)^2)(\delta T_k)^2 + C(\delta T_k)^{4\alpha} \leq C(\delta T_k)^{4(\alpha \wedge \frac{1}{2})}\end{aligned}$$

using the fact that b is Lipschitz w.r.t. the space variable and α-Hölder continuous w.r.t. the time variable. With the same development on δa_k one finally gets

$$\mathbb{E}[C_k^2 \, |\bar{\mathcal{F}}_{k-1}, N_T = n] \leq C(\delta T_k)^{4(\alpha \wedge \frac{1}{2})} \ , \quad (4.8)$$

3. Bounding $\mathbb{E}[P_{k+1}^4 \, |\bar{\mathcal{F}}_k, N_T = n]$
Using (4.2), we obtain

$$\begin{aligned}\mathbb{E}[P_{k+1}^4 \, |\bar{\mathcal{F}}_k, N_T = n] \\ &= \mathbb{E}[(M_{k+1} + \frac{1}{2}V_{k+1})^4 (\delta T_k)^{4(1-\kappa)} \theta^{4\kappa} \Gamma^4(\kappa) e^{4\delta T_k/\theta} \Big| \bar{\mathcal{F}}_k, N_T = n] \\ &\leq C\Big(\|\delta b_k\|^4 + \frac{\|\delta a_k\|^4}{(\delta T_{k+1})^2} \Big) \frac{(\delta T_k)^{4(1-\kappa)}}{(\delta T_{k+1})^2} \\ &\leq C C_k^2 \frac{(\delta T_k)^{4(1-\kappa)}}{(\delta T_{k+1})^4} \ , \quad (4.9)\end{aligned}$$

observing that $\delta T_{k+1} \leq T \leq C$ and recalling that C_k is defined by (4.7). Using the tower property of expectation and bound (4.6) yields

$$\begin{aligned}
&\mathbb{E}[\beta^2 \prod_{k=1}^{n-1} P_{k+1}^2 \,|N_T = n] \\
&\qquad = \mathbb{E}\Big[E[\beta^2 \,|\bar{\mathcal{F}}_n, N_T = n] \prod_{k=1}^{n-1} P_{k+1}^2 \,|N_T = n\Big] \\
&\qquad \leq C\mathbb{E}\Big[(\delta T_n)^{2(1-\kappa)} \Big(\|\delta a_n\|^2 + \|\delta b_n\|^2\Big) \prod_{k=1}^{n-1} P_{k+1}^2 \,|N_T = n\Big] \\
&\qquad \leq C\mathbb{E}\Big[\mathbb{E}[(\delta T_n)^{2(1-\kappa)} C_n P_n^2 \,|\bar{\mathcal{F}}_{n-1}, N_T = n] \prod_{k=1}^{n-2} P_{k+1}^2 \,|N_T = n\Big] \ .
\end{aligned}$$

By Cauchy-Schwarz, for any k and using (4.8), we have

$$\begin{aligned}
&\mathbb{E}[C_k P_k^2 |\bar{\mathcal{F}}_{k-1}, N_T = n] \\
&\qquad \leq \Big(\mathbb{E}[C_k^2 |\bar{\mathcal{F}}_{k-1}, N_T = n]\Big)^{1/2} \Big(\mathbb{E}[P_k^4 |\bar{\mathcal{F}}_{k-1}, N_T = n]\Big)^{1/2} \\
&\qquad \leq C(\delta T_k)^{2(\alpha \wedge \frac{1}{2})} C_{k-1} \frac{(\delta T_{k-1})^{2(1-\kappa)}}{(\delta T_k)^2} \\
&\qquad \leq C C_{k-1} \frac{(\delta T_{k-1})^{2(1-\kappa)}}{(\delta T_k)^{2((1-\alpha)\vee \frac{1}{2})}} \ . \qquad (4.10)
\end{aligned}$$

Hence, we obtain by recursion

$$\mathbb{E}[\beta^2 \prod_{k=1}^{n-1} P_{k+1}^2 \,|N_T = n] \leq C^{n+1} \mathbb{E}[(\delta T_n)^{2(1-\kappa)} \prod_{k=1}^{n-1} \frac{(\delta T_k)^{2(1-\kappa)}}{(\delta T_{k+1})^{2((1-\alpha)\vee \frac{1}{2})}} \,|N_T = n] \ , \quad (4.11)$$

observing that $\mathbb{E}[C_1 |\bar{\mathcal{F}}_0, N_T = n] \leq C(\delta T_1)^{2(\alpha \wedge \frac{1}{2})} \leq CT$.

Then recalling that $\kappa \leq \alpha \wedge \frac{1}{2}$ implies $(\delta T_k)^{2((\frac{1}{2}-\kappa)\wedge(\alpha-\kappa))} \leq T^{2((\frac{1}{2}-\kappa)\wedge(\alpha-\kappa))}$ finally yields

$$\mathbb{E}[\beta^2 \prod_{k=1}^{n-1} P_{k+1}^2 \,|N_T = n] \leq CC^n \ . \qquad (4.12)$$

4. Convergence of the sum $\sum_{n=1}^{\infty} C^n \mathbb{P}(N_T = n)$. Let us introduce $S_n = \sum_{k=1}^{n} \tau_k$, notice that $S_n \sim \Gamma(n\kappa, \theta)$ with cumulative distribution

$$F_{S_n}(s) = \int_0^s \frac{r^{n\kappa-1} e^{-r/\theta}}{\Gamma(n\kappa)\theta^{n\kappa}} dr \ .$$

Hence one can bound $\mathbb{P}(N_T = n)$ as follows

$$\begin{aligned}
\mathbb{P}(N_T = n) &\leq \mathbb{P}(N_T = n) \\
&\leq \mathbb{P}(S_n \leq T) \\
&\leq \int_0^T \frac{r^{n\kappa-1}}{\Gamma(n\kappa)\theta^{n\kappa}} dr = \frac{T^{n\kappa}}{n\kappa\Gamma(n\kappa)\theta^{n\kappa}}
\end{aligned}$$

This implies that

$$\sum_{n=1}^{\infty} C^n \mathbb{P}(N_T = n) \leq \sum_{n=1}^{\infty} \frac{\hat{C}^n}{n\kappa\Gamma(n\kappa)} \ ,$$

with $\hat{C} = C\frac{T}{\theta}$. Using the generalization of the Stirling formula $\Gamma(z) \sim z^{z-1/2}e^{-z}\sqrt{2\pi}$ one proves that $\frac{\hat{C}^n}{n\kappa\Gamma(n\kappa)} \sim \frac{\hat{C}^{\frac{1}{2\kappa}}e^{-\frac{1}{2}}}{\sqrt{2\pi}}(\frac{\hat{C}^{\frac{1}{\kappa}}e}{n\kappa})^{n\kappa+\frac{1}{2}}$ which is the general term of a convergent sum.

□

Remark 4.1. *The convergence of the series* (4.1) *relies on two facts :*

- *The general term of the series* (4.1) *has to be finite:* $\mathbb{E}[\Big(\beta\prod_{k=2}^{n} P_k\Big)^2 |N_T = n] < \infty$, *for any fixed number of switching times* $N_T = n$. *However, one can observe that our bound on the r.h.s. of* (4.11) *can possibly blow up to infinity when* $\kappa > \alpha \wedge 1/2$. *In particular, in the case of an exponential density, corresponding to* $\kappa = 1$, *it is well-known that the conditional distribution,* $\mathcal{L}(\delta T_k|N_T = n)$, *is the uniform distribution on* $[0, T]$, *hence the expectation on the r.h.s. of* (4.11) *is infinite. When* $\kappa \leq \alpha \wedge 1/2$, *we observe that our bound is finite whatever the conditional distribution,* $\mathcal{L}(\delta T_k|N_T = n)$. *Notice that using gamma switching times increases the occurrence of small jumps w.r.t. the exponential case and hence the occurence of high numbers of time steps is also increased. To better adjust the complexity and variance tradeoff, one could consider other switching times densities with a smaller intensity of small jumps and rely on the conditional law* $\mathcal{L}(\delta T_k|N_T = n)$ *to ensure that the expectation on the r.h.s. of* (4.11) *is bounded.*

- *The sum* $\sum_{k=1}^{\infty} C^n \mathbb{P}(N_T = n)$ *has to converge. By increasing the intensity of small jumps as explained at point 1., we expect that* $\mathbb{P}(N_T = n)$ *will decrease more slowly with* n. *This results in a tradeoff one has to achieve: increasing small jumps intensity to be able to bound each term of the series but not too strongly to ensure the convergence of* $\sum_{k=1}^{\infty} C^n \mathbb{P}(N_T = n)$.

Consequently, the representation (3.15) provides a Monte Carlo approach to compute $\mathbb{E}[g(X_T)]$, by simulating the regime switching process (3.11) instead of the SDE (1.1) which would potentially require to implement a stochastic Euler discretization scheme. However, even though our estimator is proved to have finite variance, one can observe in practice huge variances due to the product of a random number of terms P_k that could potentially take values greater that one. This expectation of products is *by nature* not a good candidate for Monte Carlo estimation. Hence, we propose to use a resampling procedure to change this expectation of products in a product of expectations which is known to be much more stable for estimation.

5 Resampling method for regime switching processes

In this section, we propose to introduce an interacting particle system (in the same vein as those thoroughly discussed in the reference books [**?**] and [**?**]) to approximate $u(t_0, x_0)$. We will prove that the resulting estimator has finite variance under the same assumptions

required to bound the variance of estimator (3.15). However, in practice, the new estimator relying on interacting particle systems will show better performances providing smaller variances in many examples, as illustrated in Section 6.

5.1 A Feynman-Kac measure representation

First we have to express $u(t_0, x_0)$ as an integral according to a Feynman-Kac measure. Let us consider the Markov chain consisting of the sequence of random variables $\check{X}_k := (T_k, \bar{X}_k)$, where (T_k) and $(\bar{X}_k)$ are given respectively by the dynamics (3.8) and (3.11). In the sequel, we note $\check{X}_{0:k} := (\check{X}_0, \cdots, \check{X}_k)$ the path valued Markov chain. Let us introduce, for any integer $k \geq 0$, the real valued function $\check{G}_k$ depending on the path $\check{x}_{0:k} \in E_k := (\mathbb{R}_+ \times \mathbb{R}^d)^{k+1}$ with the notations $\check{x}_{0:k} := (\check{x}_0, \cdots, \check{x}_k)$ and $\check{x}_p := (t_p, x_p) \in \mathbb{R}_+ \times \mathbb{R}^d$ such that

$$\check{G}_k(\check{x}_{0:k}) := \begin{cases} 1 & \text{if } k = 0 \text{ or } k = 1 \\ \frac{\check{M}_k(\check{x}_{0:k}) + \frac{1}{2}\check{V}_k(\check{x}_{0:k})}{f_\Gamma^{\kappa,\theta}(\delta t_{k-1})} & \text{if } k \geq 2 \text{ and } \delta t_{k-1}\delta t_k > 0 \\ 1 & \text{elsewhere .} \end{cases} \tag{5.1}$$

with $\delta t_{k+1} := t_{k+1} - t_k$ and where the real valued functions $\check{M}_{k+1}$, $\check{V}_{k+1}$ and $\delta\check{W}_{k+1}$ are such that for any $\check{x}_{0:k+1} \in E_{k+1}$

$$\begin{aligned} \check{M}_{k+1}(\check{x}_{0:k+1}) &:= \begin{cases} (b(t_k, x_k) - b(t_{k-1}, x_{k-1})).(\sigma(t_k, x_k)^{-1})^\top \frac{\delta\check{W}_{k+1}(\check{x}_{0:k+1})}{\delta t_{k+1}} & \text{if } \delta t_{k+1} > 0 \\ 1 \quad \text{elsewhere} \end{cases} \\ \check{V}_{k+1}(\check{x}_{0:k+1}) &:= \begin{cases} (a(t_k, x_k) - a(t_{k-1}, x_{k-1})) : \frac{B_{k+1}(\check{x}_{0:k+1})}{(\delta t_{k+1})^2} & \text{if } \delta t_{k+1} > 0 \\ 1 \quad \text{elsewhere ,} \end{cases} \end{aligned} \tag{5.2}$$

with

$$\begin{aligned} B_{k+1}(\check{x}_{0:k+1}) &:= (\sigma(t_k, x_k)^{-1})^\top \Big(\delta\check{W}_{k+1}(\check{x}_{0:k+1})\delta\check{W}_{k+1}(\check{x}_{0:k+1})^\top - \delta t_{k+1}\mathbb{I}\Big)\sigma(t_k, x_k)^{-1} \\ \delta\check{W}_{k+1}(\check{x}_{0:k+1}) &:= \sigma(t_k, x_k)^{-1}(x_{k+1} - x_k - b(t_k, x_k)\delta t_{k+1}) \ . \end{aligned}$$

Observe that $\check{G}_{k+1}$ does not really depend on the whole path $\check{x}_{0:k+1}$, but only on $(\check{x}_{k-1}, \check{x}_k, \check{x}_{k+1})$, for $k > 0$. Recalling (3.13), notice that the following identity holds

$$\check{G}_k(\check{X}_{0:k}) = P_k \ , \quad \mathbb{P} \ a.s. \quad \text{for all} \quad k = 2, \cdots, N_T \ .$$

In the sequel, it will appear to be crucial to consider positive *potential functions* with uniformly bounded conditional variances, more specifically such that $\sup_{\check{x}_{0:k} \in E_k} \mathbb{E}[G_{k+1}^2(\check{X}_{0:k+1}) \,|\check{X}_{0:k} = \check{x}_{0:k}] < \infty$, thus we define the potential functions $(G_k)_{k\geq 0}$ (depending implicitly on T) such that for any $k \geq 0$ and for any $\check{x}_{0:k} \in E_k$,

$$G_k(\check{x}_{0:k}) := \begin{cases} 1 & \text{if } k = 0 \\ |\check{G}_1(\check{x}_{0:1})|(\delta t_1)^{1-\kappa}\sqrt{c_1(\check{x}_{0:1})} & \text{if } k = 1, \ \delta t_1 > 0 \\ |\check{G}_k(\check{x}_{0:k})|\sqrt{\frac{c_k(\check{x}_{0:k})}{c_{k-1}(\check{x}_{0:k-1})}}\Big(\frac{\delta t_k}{\delta t_{k-1}}\Big)^{1-\kappa} & \text{if } k \geq 2 \ , \ \delta t_{k-1}\delta t_k > 0 \ , \\ 1 & \text{elsewhere ,} \end{cases} \tag{5.3}$$

where the real valued function c_k is defined on E_k, for $k \geq 1$, by

$$c_k(\check{x}_{0:k}) = |\delta t_k| + \|b(t_k, x_k) - b(t_{k-1}, x_{k-1})\|^2 + \|a(t_k, x_k) - a(t_{k-1}, x_{k-1})\|^2 \tag{5.4}$$

Notice that this definition of c_k is such that $c_k(\check{X}_{0:k}) = C_k + \delta T_k$ where C_k was defined in (4.7), hence

$$G_k^2(\check{X}_{0:k}) = \frac{C_k + \delta T_k}{C_{k-1} + \delta T_{k-1}} \Big(\frac{\delta T_k}{\delta T_{k-1}} \Big)^{2(1-\kappa)} P_k^2 \ , \quad \mathbb{P}\ a.s. \quad \text{for all} \quad k = 2, \cdots, N_T \ .$$

Then observe that one can prove an inequality similar as (4.10) with C_k replaced by $c_k(\check{X}_{0:k})$

$$\begin{aligned} &\mathbb{E}[c_k(\check{X}_{0:k}) P_k^2 | \bar{\mathcal{F}}_{k-1}, N_T = n] \\ &\qquad \leq \Big(\mathbb{E}[c_k^2(\check{X}_{0:k}) | \bar{\mathcal{F}}_{k-1}, N_T = n] \Big)^{1/2} \Big(\mathbb{E}[P_k^4 | \bar{\mathcal{F}}_{k-1}, N_T = n] \Big)^{1/2} \\ &\qquad \leq C c_{k-1}(\check{X}_{0:k-1}) \frac{(\delta T_{k-1})^{2(1-\kappa)}}{(\delta T_k)^{2((1-\alpha)\vee \frac{1}{2})}} \ , \end{aligned} \tag{5.5}$$

which yields as announced, that for any $\kappa \leq \alpha \wedge \frac{1}{2}$ and $\check{x}_{0:k-1} \in E_{k-1}$

$$\mathbb{E}[G_k^2(\check{X}_{0:k}) \, | \check{X}_{0:k-1} = \check{x}_{0:k-1}] \leq C < \infty \ . \tag{5.6}$$

Notice that $\prod_{k=2}^{N_T} P_k = H_{N_T+1}(\check{X}_{0:N_T+1}) \prod_{k=0}^{N_T} G_k(\check{X}_{0:k}) S_k(\check{X}_{0:k})$, $\quad \mathbb{P}\ a.s.$ where for any $k \geq 0$ and for any $\check{x}_{0:k} \in E_k$,

$$S_k(\check{x}_{0:k}) := Sign(\check{G}_k(\check{x}_{0:k})) \ , \tag{5.7}$$

and

$$H_{k+1}(\check{x}_{0:k+1}) := \begin{cases} \frac{1}{(\delta t_k)^{1-\kappa} \sqrt{c_k(\check{x}_{0:k})}} & \text{if } k \geq 1 \text{ and } \delta t_k > 0 \\ 1 & \text{elsewhere .} \end{cases} \tag{5.8}$$

Let us introduce $\beta_{n+1} := \frac{1}{2}\beta_{1,n+1} + \frac{1}{2}\beta_{2,n+1}$ defined on E_{n+1} such that $\beta_{1,1}(\check{x}_{0:1}) = \beta_{2,1}(x_{0:n+1}) = \frac{1}{(1-F_\Gamma^{\kappa,\theta}(\delta t_1))} g(x_1)$ and for any $n \geq 1$

$$\begin{cases} \beta_{1,n+1}(\check{x}_{0:n+1}) &:= \frac{g(x_{n+1}) - g(x_n)}{1 - F_\Gamma^{\kappa,\theta}(\delta t_{n+1})} \frac{M_{n+1}(\check{x}_{0:n+1}) + \frac{1}{2}\check{V}_{n+1}(\check{x}_{0:n+1})}{f_\Gamma^{\kappa,\theta}(\delta t_n)} \\ \beta_{2,n+1}(\check{x}_{0:n+1}) &:= \frac{g(\hat{x}_{n+1}) - g(x_n)}{1 - F_\Gamma^{\kappa,\theta}(\delta t_{n+1})} \frac{-M_{n+1}(\check{x}_{0:n+1}) + \frac{1}{2}\check{V}_{n+1}(\check{x}_{0:n+1})}{f_\Gamma^{\kappa,\theta}(\delta t_n)} \ , \end{cases} \tag{5.9}$$

with $\hat{x}_{n+1} = x_n + b(t_n, x_n)\delta t_{n+1} - \sigma(t_n, x_n)\delta \check{W}_{n+1}(\check{x}_{0:n+1})$.
Recalling (3.15), observe that

$$\begin{aligned} u(t_0, x_0) &= \mathbb{E}[\beta \prod_{k=2}^{N_T} P_k \mathbf{1}_{N_T \geq 1}] + \mathbb{E}[\frac{g(\bar{X}_1)}{1 - F_\Gamma^{\kappa,\theta}(\delta T_1)} \mathbf{1}_{N_T = 0}] \\ &= \mathbb{E}[(\beta_{N_T+1} H_{N_T+1})(\check{X}_{0:N_T+1})(S_{0:N_T} G_{0:N_T})(\check{X}_{0:N_T})] \ , \end{aligned} \tag{5.10}$$

where to simplify the notation $G_{p:q}$ (resp. $S_{p:q}$) denotes the product $\prod_{k=p}^q G_k$ (resp. $\prod_{k=p}^q S_k$), with in particular $G_{p,q} = \mathbf{1}$ when $p > q$, where $\mathbf{1}$ denotes the function which takes the unique value 1. Now, we can define the sequence of non negative measures $(\gamma_k)_{k \geq 0}$ such that for any real valued bounded test function φ defined on $E_n := (\mathbb{R}_+ \times \mathbb{R}^d)^{n+1}$, we have

$$\gamma_k(\varphi) := \mathbb{E}[\varphi(\check{X}_{0:k}) \prod_{p=0}^{k-1} G_p(\check{X}_{0:p})] = \mathbb{E}[\varphi(\check{X}_{0:k}) G_{0:k-1}(\check{X}_{0:k-1})] \ , \quad \text{for } k \geq 1 \ . \tag{5.11}$$

We set by convention $\gamma_0 := \mu_0$ where μ_0 denotes the probability distribution, $\mathcal{L}(\check{X}_0)$, of the initial condition $\check{X}_0 = (t_0, x_0)$ i.e. $\mu_0 := \mathcal{L}(\check{X}_0) = \delta_{(t_0,x_0)}$. Gathering (5.10) together with the above definition one readily obtains the following proposition expressing $u(t_0, x_0)$ as an integral w.r.t. the non-negative measures γ_n.

Remark 5.1. *The weights used in equation* (5.3) *can be generalized with* $\rho \in [\frac{1}{2}, 1-\kappa]$ *as*

$$G_k(\check{x}_{0:k}) := \begin{cases} 1 & \text{if } k = 0 \\ |\check{G}_1(\check{x}_{0:1})|(\delta t_1)^\rho \sqrt{c_1(\check{x}_{0:1})} & \text{if } k = 1,\ \delta t_1 > 0 \\ |\check{G}_k(\check{x}_{0:k})| \sqrt{\frac{c_k(\check{x}_{0:k})}{c_{k-1}(\check{x}_{0:k-1})}} \left(\frac{\delta t_k}{\delta t_{k-1}}\right)^\rho & \text{if } k \geq 2\ ,\ \delta t_{k-1}\delta t_k > 0\ , \\ 1 & \text{elsewhere}\ , \end{cases} \tag{5.12}$$

Proposition 5.1. *Under Assumptions 2 and 3, the following identity holds for any* $n \geq 1$

$$u(t_0, x_0) = \gamma_n(\varphi_n)\ , \tag{5.13}$$

where $(\varphi_n)_{n\geq 1}$ *is a sequence of real valued functions such that for any* $n \geq 1$ *and* $\check{x}_{0:n} \in E_n := (\mathbb{R}_+ \times \mathbb{R}^d)^{n+1}$

$$\varphi_n(\check{x}_{0:n}) := \mathbb{E}[(\beta_{N_T+1} H_{N_T+1})(\check{X}_{0:N_T+1})(S_{1:N_T} G_{n:N_T})(\check{X}_{0:N_T})\,|\check{X}_{0:n} = \check{x}_{0:n}]\ . \tag{5.14}$$

Remark 5.2. *Observe that for a given* $n \geq 1$, φ_n *is defined by* (5.14) *as a conditional expectation of a terminal payoff delivered at a future random time* $N_T + 1$, *knowing the state of the Markov chain from time* 0 *to* n. *Hence, evaluating* $\varphi_n(\check{x}_{0:n})$ *is not trivial, for a given* $\check{x}_{0:n}$, *since it requires to compute a conditional expectation. However whenever* $\check{x}_n = (t_n, x_n)$ *is such that* $t_n \geq T$, *then the knowledge of* $\check{X}_{0:n} = \check{x}_{0:n}$ *determines completely both* $N_T = q < n$ *and* $\check{X}_{0:N_T+1}$, *which implies*

$$\begin{aligned} \varphi_n(\check{x}_{0:n}) &:= \mathbb{E}[(\beta_{N_T+1} H_{N_T+1})(\check{X}_{0:N_T+1})(S_{1:N_T} G_{n:N_T})(\check{X}_{0:N_T})\,|\check{X}_{0:n} = \check{x}_{0:n}] \\ &= (\beta_{q+1} H_{q+1})(\check{x}_{0:q+1}) S_{1:q}(\check{x}_{0:q})\ . \end{aligned}$$

Now, let us introduce the sequence of probability measures (η_k) defined by normalization of $(\gamma_k)_{k\geq 1}$

$$\eta_k(\varphi) := \frac{\gamma_k(\varphi)}{\gamma_k(\mathbf{1})} = \frac{\mathbb{E}[\varphi(\check{X}_{0:k}) G_{0:k-1}(\check{X}_{0:k-1})]}{\mathbb{E}[G_{0:k-1}(\check{X}_{0:k-1})]}\ , \quad \text{for any } k \geq 0\ , \tag{5.15}$$

where $\mathbf{1}$ denotes the function which takes the unique value 1. Observing that for $k \geq 1$, $\gamma_k(\mathbf{1}) = \gamma_{k-1}(G_{k-1})$, we obtain by recurrence

$$\begin{aligned} \gamma_k(\varphi) &= \eta_k(\varphi)\gamma_k(\mathbf{1}) \\ &= \eta_k(\varphi)\gamma_{k-1}(G_{k-1}) \\ &= \eta_k(\varphi)\eta_{k-1}(G_{k-1}) \cdots \eta_0(G_0)\ . \end{aligned} \tag{5.16}$$

As announced, we have replaced the expectation of a product of functions by the product of expectations of functions, since for any $n \geq 1$

$$u(t_0, x_0) = \gamma_n(\varphi_n) = \mathbb{E}[\varphi_n(\check{X}_{0:n})] = \eta_n(\varphi_n)\eta_{n-1}(G_{n-1}) \cdots \eta_0(G_0)\ .$$

Our objective is now to approximate the sequence of probability measures $(\eta_k)_{k\geq 0}$ by a sequence of empirical measures $(\eta_k^N)_{k\geq 0}$ based on a system of N particles to finally end up with an approximation of the type

$$u(t_0, x_0) \approx \eta_n^N(\varphi_n)\eta_{n-1}^N(G_{n-1}) \cdots \eta_0^N(G_0)\ .$$

5.2 The particle approximation scheme

The sequence of approximating measures $(\eta_k^N)_{k\geq 0}$ will be defined by mimicking the dynamics of $(\eta_k)_{k\geq 0}$. Hence, we begin by describing this recursive dynamics.
First let K_k denote the transition kernel of the path valued Markov chain $(X'_k := \check{X}_{0:k})$ from $k-1$ to k for any integer $k \geq 1$. Recall that K_k can be considered both as an integral operator on the space of measurable functions defined on E_k and on the space of finite measures, $\mathcal{M}(E_{k-1})$, such that

- for any measurable test function f_k defined on E_k, $K_k(f_k)$ is a measurable function defined on E_{k-1} such that for any $x'_{k-1} \in E_{k-1}$

$$K_k(f_k)(x'_{k-1}) = \mathbb{E}[f_k(X'_k)\,|X'_{k-1} = x'_{k-1}] = \int_{y'_k \in E_k} K_k(x'_{k-1}, dy'_k) f_k(y'_k) \ ,$$

- for any finite measure m_{k-1} on E_{k-1}, $m_{k-1}K_k$ is a finite measure on E_k such that for any $x'_k \in E_k$

$$(m_{k-1}K_k)(dx'_k) = \int_{y'_{k-1} \in E_{k-1}} m_{k-1}(dy'_{k-1}) K_k(y'_{k-1}, dx'_k) \ .$$

  In particular, let μ_k denote the probability law underlying the random variable $X'_k := \check{X}_{0:k}$ (we will often write $\mu_k = \mathcal{L}(X'_k)$), for any $k \geq 0$. Then observe that $\mu_k K_{k+1} = \mu_{k+1}$ the probability law of $X'_{k+1} := \check{X}_{0:k+1}$. Besides, notice that if $\check{K}_k$ denotes the transition kernel of the Markov chain $(\check{X}_k)$ from $k-1$ to k, then the transition kernel K_k is obtained as the following cartesian product, for any $(y'_{k-1}, x'_k) := (y_{0:k-1}, dx_{0:k}) \in E_{k-1} \times E_k$

$$K_k(y'_{k-1}, dx'_k) = K_k(y_{0:k-1}, dx_{0:k}) = \delta_{y_{0:k-1}}(dx_{0:k-1}) \times \check{K}_k(y_{k-1}, dx_k) \ .$$

Now we can describe the dynamics of $(\eta_k)_{k\geq 0}$ with k. For any real valued test function f_k defined on E_k, the following identities holds

$$\begin{aligned}
\eta_k(f_k) &:= \frac{\gamma_k(f_k)}{\gamma_k(\mathbf{1})} \\
&= \frac{\mu_k(f_k G_{1:k-1})}{\mu_k(G_{1:k-1})} \ , \quad \text{where } \mu_k := \mathcal{L}(X'_k) = \mathcal{L}(\check{X}_{0:k}) \text{ , and } G_{1:k} := \prod_{p=1}^{k} G_p \\
&= \frac{\mu_{k-1}(K_k(f_k) G_{1:k-1})}{\mu_{k-1}(G_{1:k-1})} \quad \text{by the tower property of conditional expectation} \\
&= \frac{\gamma_{k-1}(K_k(f_k) G_{k-1})}{\gamma_{k-1}(G_{k-1})} \quad \text{by definition (5.11) of } \gamma_{k-1} \\
&= \frac{\eta_{k-1}(K_k(f_k) G_{k-1})}{\eta_{k-1}(G_{k-1})} \quad \text{by dividing the numerator and denominator by } \gamma_{k-1}(\mathbf{1}) \\
&= ((G_{k-1} \cdot \eta_{k-1}) K_k)(f_k) \ ,
\end{aligned}$$

where the $\cdot$ sign denotes the projective product between a non-negative function G defined on E and a non-negative measure $\mu \in \mathcal{M}^+(E)$ returning the probability measure $G \cdot \mu$ such that

$$(G \cdot \mu)(dx) := G(x)\mu(dx)/\mu(G) \ . \tag{5.17}$$

Hence, one can describe the evolution from η_{k-1} to η_k into two steps

$$\eta_{k-1} \xrightarrow{\text{Correction}} \hat{\eta}_{k-1} := G_{k-1} \cdot \eta_{k-1} \xrightarrow{\text{Evolution}} \eta_k := \hat{\eta}_{k-1} K_k \ , \tag{5.18}$$

In other words, the sequence of probability measures (η_k) satisfies the following recursion

$$\begin{cases} \eta_0 = \mu_0 \ , \quad \text{where } \mu_0 := \mathcal{L}(X_0') = \mathcal{L}(\check{X}_0) \\ \\ \hat{\eta}_k := G_k \cdot \eta_k \ , \quad \text{for all} \quad 1 \leq k \leq n \ , \\ \\ \eta_{k+1} = \hat{\eta}_k K_k \ , \quad \text{for all} \quad 1 \leq k \leq n \ . \end{cases} \tag{5.19}$$

An Interacting Particle System will be used to approximate the sequence of probability measures $(\eta_k)_{0 \leq k \leq n}$ by a sequence of empirical probability measures $(\eta_k^N)_{0 \leq k \leq n}$, such that for all $1 \leq k \leq n$, η_k^N is associated with an N-samples $(\xi_k^{1,N}, \cdots, \xi_k^{N,N})$ approximately distributed according to η_k. To simplify the notation, we will often drop the exponent N and write $(\xi_k^i)_{i=1,\cdots N}$ instead of $(\xi_k^{i,N})_{i=1,\cdots N}$. The recursive evolution described by (5.19) is approximated by the following dynamics:

$$\begin{cases} \eta_0^N = \mu_0 \\ \\ \hat{\eta}_k^N = G_k \cdot \eta_k^N \ , \quad \text{for all} \quad 1 \leq k \leq n \\ \\ \eta_{k+1}^N = S^N(\hat{\eta}_k^N K_k) \ , \quad \text{for all} \quad 1 \leq k \leq n \ , \end{cases} \tag{5.20}$$

where $S^N(\mu)$ denotes the empirical measure associated to an N-sample $(\xi^1, \cdots, \xi^N)$ i.i.d. according to μ, that is

$$S^N(\mu) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i} \ , \quad \text{where } (\xi^1, \cdots, \xi^N) \text{ i.i.d. } \sim \mu \ .$$

Hence, the algorithm proceeds as follows. Recalling that $G_0 = \mathbf{1}$, we initiate the algorithm by generating N i.i.d. random variables $(\xi_1^1, \cdots, \xi_1^N)$ according to μ_0, then we set

$$\eta_1^N = S^N(G_0 \cdot \mu_0) = S^N(\mu_0) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_1^i} \ . \tag{5.21}$$

The evolution of the discrete measures, $(\eta_k^N)_{0 \leq k \leq n}$, (where N denotes the size of the particle system) between two iterations k and $k+1$, consists into three steps:

1. **Weighting**: each particle is weighted according to the value of the current potential function G_k. For all $i \in \{1, \cdots, N\}$, we compute $\omega_k^i = \dfrac{G_k(\xi_k^i)}{\sum_{j=1}^N G_k(\xi_k^j)}$ and we set $\hat{\eta}_k^N = \displaystyle\sum_{i=1}^N \omega_{k+1}^i \, \delta_{\xi_{k+1}^i}$.

2. **Selection**: N i.i.d. random variables $(\hat{\xi}_k^1, \cdots, \hat{\xi}_k^N)$ are generated according to the weighted discrete probability distribution $\hat{\eta}_k^N = \sum_{i=1}^N \omega_k^i \, \delta_{\xi_k^i}$. More specifically, for all $i \in \{1, \cdots, N\}$, an index $I \in \{1, \cdots, N\}$ is generated independently with probability $\mathbb{P}(I = j) = \omega_k^j$ and we set $\hat{\xi}_k^i = \xi_k^I$.

3. **Mutation**: Each selected particle evolves independently according to the dynamics K_{k+1}. This produces a new particle system $(\xi_{k+1}^1, \cdots, \xi_{k+1}^N)$. More specifically, for all $i \in \{1, \cdots, N\}$, we generate independently ξ_{k+1}^i according to the conditional distribution $\mathcal{L}(X'_{k+1} | X'_k = \hat{\xi}_k^i)$, then we set

$$\eta_{k+1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1}^i} \ . \tag{5.22}$$

For all $k \geq 1$, let us introduce γ_k^N, the particle approximation of γ_k based on η_k^N defined by recursion (5.20) and such that for any real valued measurable test function f_k defined on E_k,

$$\gamma_k^N(f_k) = \eta_k^N(f_k) \prod_{0 \leq p \leq k-1} \eta_p^N(G_p) \ . \tag{5.23}$$

We begin by stating a Lemma that will be crucial to prove the convergence of our new estimator.

Lemma 5.2. *Let $(X'_n)_{n \geq 0}$ be a Markov chain (with initial distribution μ_0 and transition kernel K_k) defined on a sequence of measurable spaces $(E_n, \mathcal{E}_n)_{n \geq 0}$ and $(G_n)_{n \geq 0}$ be a sequence of positive measurable functions defined on $(E_n, \mathcal{E}_n)_{n \geq 0}$ such that there exists a finite constant $A \geq 2$ such that*

$$\sup_{x'_0 \in E_0} G_0(x'_0) \leq A \ , \quad \text{and} \quad \sup_{x'_{p-1} \in E_{p-1}} \mathbb{E}[G_p^2(X'_p) | X'_{p-1} = x'_{p-1}] \leq A \ , \quad \text{for any } p \geq 1 \ . \tag{5.24}$$

We consider the sequence of Feynman-Kac measures (γ_n) such that for any measurable real valued function f_n defined on E_n,

$$\gamma_n(f_n) := \mathbb{E}[f_n(X'_n) \prod_{k=0}^{n-1} G_k(X'_k)] \ . \tag{5.25}$$

Let (γ_n^N) be a sequence of particle approximation measures of (γ_n) defined similarly as in (5.23), with $(\eta_p^N)_{0 \leq p}$ defined by (5.20). For a given $n \geq 1$, let us consider a real valued measurable function f_n defined on E_n such that there exists a finite positive constant B such that

$$\sup_{x'_{p-1} \in E_{p-1}} |\mathbb{E}[f_n^2(X'_n) G_{p:n-1}^2(X'_{p:n}) | X'_{p-1} = x'_{p-1}] \leq B \quad \text{for any } p = 1, \cdots n \ . \tag{5.26}$$

Then the particle approximation $\gamma_n^N(f_n)$ is unbiased with finite variance, more precisely

$$\mathbb{E}[\gamma_n^N(f_n)] = \gamma_n(f_n) \ , \ \text{and} \quad \mathbb{E}[\big(\gamma_n^N(f_n) - \gamma_n(f_n)\big)^2] \leq 2B \frac{A^{n+2}}{N} \quad \text{for } N \geq A^{n+1} \ . \tag{5.27}$$

The proof of this Lemma relies on the formalism developed in the reference books [**?**, **?**]. However, we had to carry out an original proof to take into account our specific framework where the potential functions G_k are unbounded which is not considered to our knowledge in the existing literature. The proof is placed in the Appendix **??**.

We are now in a position to state the main result of this section.

Theorem 5.3. *Suppose that Assumptions 2, 3 and 5 are satisfied. For any $n \geq 1$, the resampling estimator $\gamma_n^N(\varphi_n)$ defined by (5.23) is unbiased with finite variance. More precisely,*

$$\mathbb{E}[\gamma_n^N(\varphi_n)] = u(t_0, x_0) \ , \quad \textit{and} \quad \mathbb{E}[\big(\gamma_n^N(\varphi_n) - u(t_0, x_0)\big)^2] \leq \frac{C^{n+2}}{N} \quad \textit{for } N \geq C^{n+1} \ , \tag{5.28}$$

where $(\varphi_n)_{n\geq 1}$ is a sequence of real valued functions defined on E_n by (5.14) and C is a constant depending only on the characteristics of the problem (T, the bounds or Lipschitz constants related to g, b, σ, a).

Remark 5.3. *1. Computing $\gamma_n^N(\varphi_n)$ reduces to compute the following product of empirical means*

$$\begin{aligned} \gamma_n^N(\varphi_n) &= \eta_n^N(\varphi_n)\eta_{n-1}^N(G_{n-1}) \cdots \eta_0^N(G_0) \\ &= \left(\frac{1}{N}\sum_{i=1}^N \varphi_n(\xi_n^i)\right)\left(\frac{1}{N}\sum_{i=1}^N G_{n-1}(\xi_{n-1}^i)\right)\cdots\left(\frac{1}{N}\sum_{i=1}^N G_0(\xi_0^i)\right) \ , \end{aligned}$$

where $(\xi_k^i)_{1\leq i\leq N}$ is the particle system at the kth iteration of the algorithm as stated by (5.22). This in particular requires to compute $\varphi_n(\xi_n^i)$ for each particle of the final particle system $(\xi_n^i)_{i=1,\cdots N}$. Recalling Remark 5.2, this may require to compute a conditional expectation. In practice, one chooses n large enough such that most of particles have already reached time T after n iterations implying that for most particles $\varphi_n(\xi_n^i)$ can be computed explicitly. In the rare cases of particles ξ_n^i that have not reached yet time T, the computation of $\varphi_n(\xi_n^i)$ that should normally require to compute a conditional expectation is approximated by one simulation according to

$$\mathcal{L}((\beta_{N_T+1}H_{N_T+1}S_{0:N_T+1})(\check{X}_{0:N_T+1})G_{n:N_T}(\check{X}_{0:N_T})\,|\check{X}_{0:n} = \xi_n^i) \ .$$

Notice that it would be interesting to consider the estimator

$$\gamma_{n_N}^N(\varphi_{n_N}) \ , \quad \textit{with} \quad n_N = \inf\{n \,|\, \xi_n^i \textit{ has reached } T \textit{ for all } i = 1, \cdots N\} \ .$$

This will be left for future work.

2. Another approach to avoid this problem would consists in doing the resampling procedure only on the space variables. First simulate a sequence of random switching times $(T_1, \cdots, T_{N_T})$ and conditionally to this time mesh run an interacting particle system on the Markov chain $\bar{X}$ (3.11). The estimator would then be given as an empirical mean of the resampling estimates over i.i.d. time meshes.

Proof. Theorem 5.3 is a direct consequence of Proposition 5.1 stating that $\gamma_n(\varphi_n) = u(t_0, x_0)$ and of Lemma 5.2 after having verified that there exists a finite positive constant C for which

the bounds (5.24) and (5.26) are verified. Observe that (5.24) is automatically implied by (5.6). Let us consider (5.26), similarly to the proof of Proposition 4.1 one obtains

$$\begin{aligned}&\mathbb{E}[\varphi_n^2(\check{X}_{0:n})G_{p:n-1}^2(\check{X}_{p:n})|\check{X}_{0:p-1}=\check{x}_{0:p-1}]\\&=\sum_{q=0}^{\infty}\mathbb{E}[\varphi_n^2(\check{X}_{0:n})G_{p:n-1}^2(\check{X}_{p:n})|\check{X}_{0:p-1}=\check{x}_{0:p-1},N_T=q]\mathbb{P}(N_T=q)\ .\end{aligned}\tag{5.29}$$

Now considering the general term of this sum for $q\geq p\geq 2$

$$\begin{aligned}&\mathbb{E}[(\beta_{N_T+1}^2H_{N_T+1}^2)(\check{X}_{0:N_T+1})\prod_{k=p}^{N_T}G_k^2(\check{X}_{0:k})\,|\check{X}_{0:p-1}=\check{x}_{0:p-1},N_T=q]\\&=\mathbb{E}[\beta^2\frac{1}{c_{p-1}(\check{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}\prod_{k=p-1}^{q-1}P_{k+1}^2\,|\check{X}_{0:p-1}=\check{x}_{0:p-1},N_T=q]\\&\leq C\mathbb{E}[(\delta T_q)^{2(1-\kappa)}\frac{1}{c_{p-1}(\check{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}c_q(\check{X}_{0:q})P_q^2\prod_{k=p-1}^{q-2}P_{k+1}^2\,|\check{X}_{0:p-1}=\check{x}_{0:p-1},N_T=q]\ ,\end{aligned}$$

where C is a constant that may change from line to line. Recalling (5.5) finally gives

$$\begin{aligned}&\mathbb{E}[(\beta_{N_T+1}^2H_{N_T+1}^2)(\check{X}_{0:N_T+1})\prod_{k=p}^{N_T}G_k^2(\check{X}_{0:k})\,|\check{X}_{0:p-1}=\check{x}_{0:p-1},N_T=q]\\&\leq C^{q-p+1}\mathbb{E}[(\delta T_q)^{2(1-\kappa)}\frac{c_{p-1}(\check{X}_{0:p-1})}{c_{p-1}(\check{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}\prod_{k=p-1}^{q-1}\frac{(\delta T_k)^{2(1-\kappa)}}{(\delta T_{k+1})^{2((1-\alpha)\vee\frac{1}{2})}}\,|N_T=q]\\&\leq C^{q-p+1}\ .\end{aligned}$$

We proceed similarly when $p=1$. We conclude by observing that the sum (5.29) is finite by the same argument as in the proof of Proposition 4.1. □

6 Numerical simulations

In this section, we begin by an empirical analysis of complexity then we analyse and compare the performances of the three approaches described previously

1. Switching Monte Carlo method with exponential switching times;
2. Switching Monte Carlo method with gamma switching times (with parameter $\kappa\leq 1/2$);
3. Resampling and Switching Monte Carlo method with gamma switching times (with parameter $\kappa\leq 1/2$).

On one test case, we compare numerically the Switching Monte Carlo method with gamma switching times with the Euler Monte Carlo method.
First, we consider a simple example for which all assumptions of Proposition 4.1 are satisfied. Then we consider simulations involving a more standard payoff function g occuring in finance (corresponding to the call option) that does not fulfill Assumption 4.1. However, this offers the opportunity to check the robustness of our approach out of theoritical assumptions.
In all cases, we consider

- a drift coefficient $b(t,x)=-10\vee(1-x)\wedge 10$,
- an initial condition $x_0=1$,
- a terminal time $T=1$.