Neural variance reduction for stochastic differential equations

University of Nottingham
UK | CHINA | MALAYSIA

Piers D. Hinds [1] Michael V. Tretyakov [1]

[1]School of Mathematical Sciences, University of Nottingham

Objective

Compute

$$u(t,x) = \mathbb{E} f(X_{t,x}(T)), \tag{1}$$

where $X(s) = X_{t,x}(s)$ is the solution to

$$\mathrm{d}X = b(s,X)\mathrm{d}s + \sigma(s,X)\mathrm{d}W, \quad X(t) = x, \tag{2}$$

and $f : \mathbb{R}^d \to \mathbb{R}$.

Motivation

- Option pricing: In financial mathematics, $X(s)$ represents the price of an asset at time s, and f represents the payoff of an option. Computing $\mathbb{E}f(X(T))$ amounts to finding the fair price of an option.

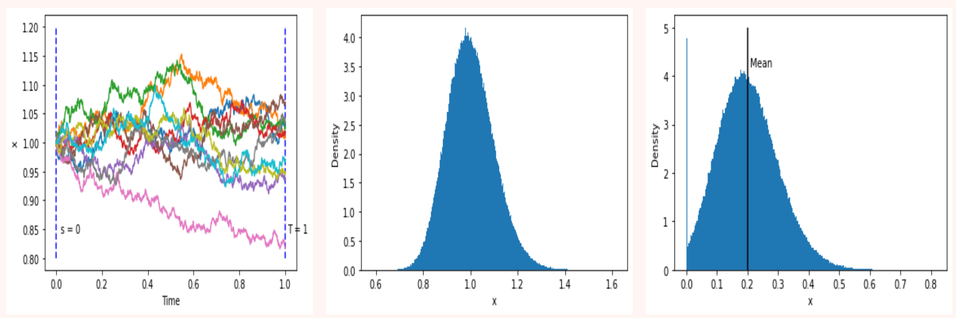

Figure 1. $\mathrm{d}X = \sigma X \mathrm{d}W$ and $f(x) = (x-K)_+$ for $\sigma = 0.1$, $K = 0.8$

- Cauchy problems: Any linear parabolic PDE has a probabilistic representation in the form of (2). The function u solves the Cauchy problem:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^\top)_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_i b_i \frac{\partial u}{\partial x_i} = 0, \tag{3}$$

$$u(T,x) = f(x). \tag{4}$$

This is a particularly practical approach when:

1. the dimension of the domain of the PDE is large; and/or
2. the PDE only needs to be solved at a small number of points.

The Monte Carlo method

In general there is no analytical solution to (1), so we solve the problem via the Monte Carlo (MC) method:

$$\mathbb{E}f(X(T)) \approx M^{-1}\sum_{m=1}^{M} f(\bar{X}^{(m)}(T)) =: \hat{u}, \tag{5}$$

where the $\bar{X}^{(m)}(T)$'s are independent approximations (in the weak sense) of $X(T)$.

Crucially, the Monte Carlo error depends on the variance of $f(X(T))$. The Central Limit Theorem gives us confidence intervals, for example:

$$\hat{u} \pm 1.96\sqrt{\frac{\mathrm{Var} f(X(T))}{M}}. \tag{6}$$

The idea of variance reduction is to replace $\Gamma := f(X(T))$ with a different random variable, say $\tilde{\Gamma}$, which has smaller variance than Γ but the same expectation, so that the confidence intervals in (6) are tighter.

Variance reduction

Find $\tilde{\Gamma}$ with

$$\mathbb{E}\tilde{\Gamma} = \mathbb{E}\Gamma, \ \ \mathrm{Var}\tilde{\Gamma} << \mathrm{Var}\Gamma, \tag{7}$$

then sample from $\tilde{\Gamma}$ instead.

Variance reduction for SDEs

For our problem, we have the well-known method of control variates (e.g. [3]).

Take

$$\tilde{\Gamma} := f(X(T)) + Z(T), \tag{8}$$

where

$$\mathrm{d}Z = G(s,X)^\top \mathrm{d}W, \quad Z(t) = 0, \tag{9}$$

for an arbitrary function G. $Z(T)$ acts as a control variate and Z is called the control variate process. Then:

- No matter the choice of G, $\mathbb{E}[f(X(T)) + Z(T)] = \mathbb{E}f(X(T))$;
- For the correct choice of G, we can reduce the variance of $f(X(T)) + Z(T)$ to zero (see the next theorem).

Theorem (Milstein & Schoenmakers [2])

If

$$G(s,x) = -\sigma(s,x)^\top \nabla u(s,x), \tag{10}$$

for all $s \in [t,T]$, $x \in \mathbb{R}^d$ then

$$\mathrm{Var}[f(X(T)) + Z(T)] = 0. \tag{11}$$

The above theorem cannot be used practically to construct G since the solution to the related PDE, $u(s,x)$ is unknown. However, if we think of finding the optimal G as minimizing the variance of $\tilde{\Gamma}$, it shows that there is an optimal solution to this problem. This motivates the following method.

Preprint | Poster | GitHub

Method

We follow the idea of [4, Algorithm 4]:

1. Parameterize G with a neural network G_θ, $\theta \in \mathbb{R}^p$, to obtain the system:

$$\mathrm{d}X = b(s,X)\mathrm{d}s + \sigma(s,X)\mathrm{d}W, \tag{12}$$

$$\mathrm{d}Z_\theta = G_\theta(s,X)^\top \mathrm{d}W. \tag{13}$$

2. Learn optimal θ^* by minimizing the empirical variance of simulated realizations:

$$\theta^* = \arg\min_\theta \mathrm{Var}_{M_r}\tilde{\Gamma}_\theta, \tag{14}$$

where $\tilde{\Gamma}_\theta = f(X(T)) + Z_\theta(T)$ and $\mathrm{Var}_{M_r}(\cdot)$ denotes the sample variance over M_r samples.

3. Use the trained neural network, G_{θ^*}, to simulate low variance realizations of $\tilde{\Gamma}_{\theta^*}$ in order to compute the expectation:

$$\mathbb{E}f(X(T)) \approx \hat{u} = M^{-1}\sum_{m=1}^{M}\left[f(X^{(m)}(T)) + Z_{\theta^*}^{(m)}(T)\right]. \tag{15}$$

SDEs driven by Lévy processes

We extend the theoretical result and the numerical method to the case that the SDE is driven by Lévy processes. As an example, consider the case that our SDE is driven by both a Wiener process and a Poisson process. Let $N(s)$ be a Poisson process with rate λ, $Y(s)$ denotes a jump size at time s according to some distribution ρ. Consider the SDEs

$$\mathrm{d}X = b(s,X)\mathrm{d}s + \sigma(s,X)\mathrm{d}W + F(s,X)Y(s)\mathrm{d}N, \tag{16}$$

$$\mathrm{d}Z_\theta = G_{W,\theta}(s,X)^\top \mathrm{d}W + G_{N,\theta}(s,X,Y)[\mathrm{d}N - \lambda \mathrm{d}s]. \tag{17}$$

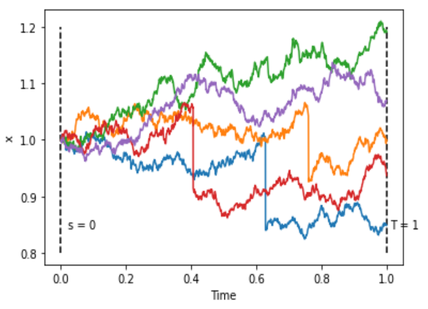

Figure 3. Merton model trajectories

Theorem (Optimality for jump-diffusions [1])

If

$$G_W(s,x) = -\sigma(s,x)^\top \nabla u(s,x), \tag{18}$$

$$G_N(s,x,z) = u(s,x) - u\big(s, x + F(s,x)z\big), \tag{19}$$

for all $(s,x,z) \in [t,T] \times \mathbb{R}^d \times \mathbb{R}^q$, then $\mathrm{Var}\tilde{\Gamma} = 0$

In [1], we consider the more general setting of the SDE being driven by a Lévy process with infinite activity.

Results

We consider two examples from option pricing. For each example we compare our method with vanilla MC (by which we mean MC without any variance reduction methods). The proposed control variate method is up to 42 times faster than vanilla MC.

Table 1. European Call option, $f(x) = (x-K)_+$, under the Heston model with $v = 0.15$, $r = 0.02$, $\kappa = 0.25$, $\theta = 0.5$, $\sigma = 0.3$, $\rho = -0.3$, $T = 3$: MC approximations (and a 95% confidence interval) with and without a control variate.

K	$u(0,1)$	Vanilla MC			Control Variate MC		
		$\hat{u}(0,1)$	Time (s)	M	$\hat{u}(0,1)$	Time (s)	M
0.8	0.42623	0.42625 ± 0.00011	1120	$2.57{\cdot}10^8$	0.42636 ± 0.00010	30.9	$1.65{\cdot}10^6$
0.9	0.38271	0.38274 ± 0.00010	1180	$2.72{\cdot}10^8$	0.38267 ± 0.00011	32.0	$1.73{\cdot}10^6$
1	0.34406	0.34401 ± 0.00010	1070	$2.47{\cdot}10^8$	0.34407 ± 0.00009	54.6	$3.17{\cdot}10^6$
1.1	0.30977	0.30977 ± 0.00010	1170	$2.70{\cdot}10^8$	0.30971 ± 0.00010	27.4	$1.43{\cdot}10^6$
1.2	0.27934	0.27927 ± 0.00009	1250	$2.88{\cdot}10^8$	0.27927 ± 0.00011	39.5	$2.20{\cdot}10^6$

Table 2. European Call option, $f(x) = (x-K)_+$, under the Merton model with $r = 0.02$, $\sigma = 0.2$, $\lambda = 1$, $\alpha = -0.05$, $\gamma = 0.3$ and $T = 3$: MC approximations (and a 95% confidence interval) with and without a control variate.

K	$u(0,1)$	Vanilla MC			Control Variate MC		
		$\hat{u}(0,1)$	Time (s)	M	$\hat{u}(0,1)$	Time (s)	M
0.8	0.35593	0.35575 ± 0.00009	2010	$1.74{\cdot}10^8$	0.35592 ± 0.00011	85.5	$2.05{\cdot}10^6$
0.9	0.30592	0.30575 ± 0.00011	1480	$1.28{\cdot}10^8$	0.30604 ± 0.00010	104	$2.49{\cdot}10^6$
1	0.26298	0.26278 ± 0.00010	1460	$1.27{\cdot}10^8$	0.26302 ± 0.00010	127	$3.09{\cdot}10^6$
1.1	0.22634	0.22624 ± 0.00011	1200	$1.05{\cdot}10^8$	0.22633 ± 0.00010	140	$3.47{\cdot}10^6$
1.2	0.19519	0.19502 ± 0.00011	1090	$9.52{\cdot}10^7$	0.19532 ± 0.00010	151	$3.79{\cdot}10^6$

Summary

- To improve the efficiency of the MC method, we want to sample from a random variable with very small variance
- We propose a method to find the optimal control variate $Z_\theta(T)$ and then we use it to sample from $\tilde{\Gamma} = f(X(T)) + Z_\theta(T)$
- We extend the theoretical result and numerical method to the case SDEs driven by Lévy processes.
- Numerical experiments demonstrate the efficiency of our method: it is up to 42 times faster than vanilla MC

References

[1] Hinds, P. D. and Tretyakov, M. V. "Neural variance reduction for stochastic differential equations". arXiv:2209.12885 (submitted) (2022).

[2] Milstein, G. N. and Schoenmakers, J. G. M. "Monte Carlo construction of hedging strategies against multi-asset European claims". Stoch. Stoch. Reports 73(1-2) (2002), pp. 125–157.

[3] Milstein, G. N. and Tretyakov, M. V. Stochastic numerics for mathematical physics. 2nd Edition. Berlin: Springer, 2021.

[4] Vidales, M. S., Siska, D., and Szpruch, L. "Unbiased deep solvers for linear parametric PDEs". Applied Mathematical Finance 28(4) (2021), pp. 299–329.