

Unbiased simulation of stochastic differential equations using parametrix expansions

Patrik Andersson

Uppsala University

Joint work with Arturo Kohatsu-Higa, Ritsumeikan U

International Conference on Monte Carlo techniques
July 5, 2016

Outline

- Stochastic differential equations.
- First unbiased method.
- Parametrix method
- Variance analysis
- Choosing simulation parameters
- Numerical experiments

Stochastic differential equations

- We consider equations of the form

$$X_t = X_0 + \sum_{j=1}^m \int_0^t \sigma_j(X_s) dW_s^j + \int_0^t b(X_s) ds, \quad t \in [0, T].$$

- W is an m -dimensional Wiener process
- $\sigma_j, b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are such that there exists a weak solution.
- The methods in this talk applies to any dimension.
- We only discuss dimension 1.
- The Euler-scheme $\{X_{t_i}^\pi; i = 0, \dots, N + 1\}$ is, given a time partition $\pi : 0 = t_0 < t_1 \dots < t_N < t_{N+1} = T$,

$$\begin{aligned} X_0^\pi &= X_0, \\ X_{t_{i+1}}^\pi &= X_{t_i}^\pi + \sigma_j(X_{t_i}^\pi)(W_{t_{i+1}} - W_{t_i}) + b(X_{t_i})(t_{i+1} - t_i), \quad i = 0, \dots, N. \end{aligned}$$

- We want $E[f(X_T)]$.

A first unbiased method

- Let X_T^n denote an approximation of X_T using an Euler method with uniform time steps of length $2^{-n}T$, $n \geq 0$.
- We may then write

$$\mathbb{E} [f(X_T)] \approx \mathbb{E} [f(X_T^{\bar{n}})] = \mathbb{E} [f(X_T^0)] + \sum_{n=1}^{\bar{n}} \mathbb{E} [f(X_T^n) - f(X_T^{n-1})],$$

- Letting $\bar{n} \rightarrow \infty$ we will get equality.

A first unbiased method

- If N is a random variable with distribution $p_n > 0$, $n \geq 0$,

$$\begin{aligned} E [f(X_T)] &= E [f(X_T^0)] + \sum_{n=1}^{\infty} E [f(X_T^n) - f(X_T^{n-1})] \\ &= \frac{p_0 E [f(X_T^0)]}{p_0} + \sum_{n=1}^{\infty} p_n \frac{E [f(X_T^n) - f(X_T^{n-1})]}{p_n} \\ &= \frac{1}{p_0} E [1(N=0)f(X_T^0)] + E \left[1(N \geq 1) \frac{f(X_T^N) - f(X_T^{N-1})}{p_N} \right], \end{aligned}$$

Variance/Complexity

- The second moment is

$$\mathbb{E} \left[\left(\mathbf{1}(N \geq 1) \frac{f(X_T^N) - f(X_T^{N-1})}{p_N} \right)^2 \right] = \sum_{n=1}^{\infty} \frac{r_n}{p_n},$$

where $r_n := \mathbb{E} \left[\left(f(X_T^n) - f(X_T^{n-1}) \right)^2 \right]$.

- Typically $r_n = O(2^{-n})$.
- If we choose $p_n \sim 2^{-n} n^2$, the second moment will be finite.
- But the complexity $\sum_{n=1}^{\infty} 2^n p_n = \infty$
- We can get an unbiased method by doing a infinite order expansion. The price is high variance or high complexity.
- In Rhee and Glynn (2015) it is shown that by using higher order approximations, finite variance can be achieved.

Parametrix method

- Let $\pi : 0 = t_0 < t_1 < \dots < t_N < t_{N+1} = T$ and define the following discrete time process

X_0^π is random variable with density $\nu(x)$,

$$X_{t_{i+1}}^\pi = X_{t_i}^\pi + \mu(X_{t_i}^\pi)(t_{i+1} - t_i) + \sigma(X_{t_i}^\pi)\sqrt{t_{i+1} - t_i}Z_{i+1}, \quad i = 0, 1, \dots, N,$$

$Z_i, i = 1, \dots, N$ are independent $N(0, 1)$ random variables,

$$\theta_t(x, y) = \frac{1}{2} \sum_{i,j=1}^d \kappa_t^{i,j}(x, y) - \sum_{i=1}^d \rho_t^i(x, y).$$

- The parametrix method is then (Bally and Kohatsu-Higa, 2015)

$$\mathbb{E} [f(X_T)] = \sum_{n=0}^{\infty} \int_{S^n} \mathbb{E} \left[\Phi(X_T^\pi) \prod_{j=0}^{n-1} \theta_{s_{j+1}-s_j}(X_{s_j}^\pi, X_{s_{j+1}}^\pi) \right] ds,$$

$$S^n = \{s = (s_1, \dots, s_n) \in \mathbb{R}^n \mid 0 < s_1 < s_2 < \dots < s_n < T\}.$$

- $\theta_s(x, y)$ is an explicit function, involving derivatives of b and σ .

Forward method

- The forward method is

$$\nu(x) = \delta_{x_0}(x),$$

$$\Phi(x) = f(x),$$

$$\mu(x) = b(x),$$

$$\begin{aligned} \theta_t(x, y) = & \frac{1}{2} \left[a''(y) + 2a'(y)H_{ta(x)}^1(y - x - b(x)t) \right. \\ & \left. + (a(y) - a(x))H_{ta(x)}^2(y - x - b(x)t) \right], \\ & - b'(y) - (b(y) - b(x))H_{ta(x)}^1(y - x - b(x)t). \end{aligned}$$

- With the assumptions

$$a \in C_b^2(\mathbb{R}),$$

$$b \in C_b^1(\mathbb{R}),$$

there exist $\underline{a}, \bar{a} \in \mathbb{R}$ such that $0 < \underline{a} \leq a(x) \leq \bar{a}$,

$$f \in C_c^\infty(\mathbb{R}).$$

Backward method

- The backward method is

$$\nu(x) = f(x),$$

$$\Phi(x) = \delta_{X_0}(x),$$

$$\mu(x) = -b(x),$$

$$\begin{aligned} \theta_t(x, y) &= \frac{1}{2}(a(y) - a(x))H_{ta(x)}^2(y - x + b(x)t) \\ &\quad - (b(x) - b(y))H_{ta(x)}^1(y - x + b(x)t). \end{aligned}$$

- With the assumptions

$$a, b \in C_b^\alpha(\mathbb{R}),$$

there exist $\underline{a}, \bar{a} \in \mathbb{R}$ such that $0 < \underline{a} \leq a(x) \leq \bar{a}$,

$$f \in C_c^\infty(\mathbb{R}).$$

Forward parametrix method

$$\mathbb{E} [f(X_T)] = \sum_{n=0}^{\infty} \int_{S^n} \mathbb{E} \left[f(X_T^\pi) \prod_{j=0}^{n-1} \theta_{s_{j+1}-s_j}(X_{s_j}^\pi, X_{s_{j+1}}^\pi) \right] ds.$$

- In writing the time integral as an expectation, we have the freedom to choose the distribution of the time steps.
- Perhaps the most obvious choice is exponential time steps.
- Let $N(t)$ be a Poisson process with intensity parameter $\lambda > 0$ and define $N \equiv N(T)$. Let τ_1, \dots, τ_N be the event times of the Poisson-process and set $\tau_0 = 0, \tau_{N+1} = T$.
- $\mathbb{P}(N = n, \tau_1 \in dt_1, \dots, \tau_n \in dt_n) = \frac{(\lambda T)^n}{n!} e^{-\lambda T} \frac{n!}{T^n} \mathbb{I}(0 \leq t_1 \leq \dots \leq t_n \leq T)$.
- $\sum_{n=0}^{\infty} \int_{S^n} \lambda^n e^{-\lambda T} ds = 1$.

Exponential time sampling

- We rewrite the time integral as an expectation:

$$\begin{aligned} \mathbb{E} [f(X_T)] &= \sum_{n=0}^{\infty} \int_{S^n} \mathbb{E} \left[f(X_T^\pi) \prod_{j=0}^{n-1} \theta_{s_{j+1}-s_j}(X_{s_j}^\pi, X_{s_{j+1}}^\pi) \right] ds \\ &= \sum_{n=0}^{\infty} \int_{S^n} \mathbb{E} \left[f(X_T^\pi) \prod_{j=0}^{n-1} \theta_{s_{j+1}-s_j}(X_{s_j}^\pi, X_{s_{j+1}}^\pi) \right] \frac{1}{\lambda^n e^{-\lambda T}} \lambda^n e^{-\lambda T} ds \\ &= e^{\lambda T} \mathbb{E} \left[f(X_T^\pi) \prod_{i=0}^{N-1} \lambda^{-1} \theta_{\tau_{i+1}-\tau_i}(X_{\tau_i}^\pi, X_{\tau_{i+1}}^\pi) \right] \end{aligned}$$

- The random variable inside the expectation can be sampled by first generating the time discretization and then doing the Euler scheme as usual.

Variance analysis

Results that can be found for the exponential time sampling are

- If $a(x) \equiv a > 0$ the variance is finite.
- For the backward method, variance is only finite in dimension 1.
- For general $a(x)$, variance is not finite.

Backward method problem

- Need to evaluate $q_{T-\tau_i}^{\text{Euler}}(X_{\tau_i}^\pi, X_0)$.
- As the dimension gets higher, it gets less likely that the path hits X_0 .

Solution

- Change the direction of simulation.
- This will introduce likelihood ratios between the transition densities in the two directions.

General $a(x)$

General $a(x)$ gives infinite variance because for small t

$$\theta_t(x, y)^2 q_t(x, y) \sim \frac{1}{t} \varphi_{4\bar{a}t}(y - x).$$

- Here $q_t(x, y)$ is the transition density from x to y and $\varphi_a(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det a}} \exp \left\{ -\frac{1}{2} x^T a^{-1} x \right\}$.
- Consider the case $N = 1$, $\lambda = 1$, $f(x) \equiv 1$,

$$\begin{aligned} e^{2T} \mathbb{E} [\theta_{\tau_i}(x_0, X_{\tau_i}^\pi)^2] &= e^T \int_0^T \mathbb{E} [\theta_t(x_0, X_t^\pi)^2] dt \\ &= e^T \int_0^T \int \theta_t(x_0, y)^2 q_t(x_0, y) dy dt \sim e^T \int_0^T \int \frac{1}{t} \varphi_{4\bar{a}t}(y - x_0) dy dt \end{aligned}$$

Importance sampling - Toy example

- A simplified version of our problem is to calculate, using MC methods,

$$\int_0^1 t^\rho dt = E [X^\rho], X \sim U(0, 1),$$

which can be calculated using simulation of n i.i.d. copies of X^ρ with $X \sim U(0, 1)$.

- However, if $\rho \in (-1, -1/2]$, the second moment of the random variable X^ρ is

$$E [X^{2\rho}] = \int_0^1 t^{2\rho} dt = \infty,$$

and thus our simulation will have an exploding variance.

Importance sampling - Toy example II

- The solution is to move some of the singularity of t^ρ from the random variable to the density.
- That is, let $\rho > 1$ and Y be a random variable with density function $t^{-\gamma}(1-\gamma)$, for $0 < t < 1$ and $-\frac{\rho\rho+1}{\rho-1} < \gamma < 1$. We then have

$$\int_0^1 t^\rho dt = \int_0^1 \frac{t^{\rho+\gamma}}{(1-\gamma)} \frac{(1-\gamma)}{t^\gamma} dt = \frac{1}{1-\gamma} \mathbb{E} [Y^{\rho+\gamma}].$$

And furthermore, the ρ -moment of the above random variable is always finite as

$$\begin{aligned} \frac{1}{(1-\gamma)^\rho} \mathbb{E} [Y^{\rho(\rho+\gamma)}] &= \frac{1}{(1-\gamma)^\rho} \int_0^1 t^{\rho(\rho+\gamma)} \frac{(1-\gamma)}{t^\gamma} dt \\ &= \frac{1}{(1-\gamma)^{\rho-1} (\rho\rho + (\rho-1)\gamma + 1)}. \end{aligned}$$

Importance sampling of discretization time points

- We should sample where the integrand is large.
- The problem is that θ_t is large for small t , but we are not sampling enough small time steps.
- Instead of using exponential time steps, we use for example, beta, gamma or weibull distributed time steps.
- All of these have density $\sim t^{-\gamma}$, for small t .
- Thus we sample $\{\xi_j; j \in \mathbb{N}\}$ i.i.d. beta (or gamma, weibull).
- Set $\tau_0 = 0$, $\tau_i \equiv \sum_{j=1}^i \xi_j$, $i \geq 1$ and let $N := \inf\{n; \tau_n < T \leq \tau_{n+1}\}$.

Beta sampling

Proposition

Let $\{\xi_j; j \in \mathbb{N}\}$ be i.i.d. r.v. with density $f_\xi(x) = \frac{(1-\gamma)}{x^\gamma \bar{\tau}^{1-\gamma}}$, $0 < x < \bar{\tau}$, $\bar{\tau} > T$, $\gamma \in (0, 1)$. Then, under the usual assumptions the following holds

$$\mathbb{E} [f(X_T)] = \mathbb{E} \left[\frac{f(X_T^\pi)}{\rho_N(\tau_1, \dots, \tau_N)} \prod_{j=0}^{N-1} \theta_{\tau_{j+1} - \tau_j}(X_{\tau_j}^\pi, X_{\tau_{j+1}}^\pi) \right],$$

with

$$\rho_n(s_1, \dots, s_n) = \left(1 - \left(\frac{T - s_n}{\bar{\tau}} \right)^{1-\gamma} \right) \left(\frac{1 - \gamma}{\bar{\tau}^{1-\gamma}} \right)^n \prod_{i=0}^{n-1} \frac{1}{(s_{i+1} - s_i)^\gamma}, n \geq 0.$$

If $0 < \gamma < 1$ then the variance is finite.

Interpretation of time step importance sampling

- The above $\xi_j \stackrel{d}{=} \bar{\tau} B$ where B is distributed as $\text{Beta}(1 - \gamma, 1)$.
- The critical point of the Beta density is that it behaves like $x^{-\gamma}$ close to 0. Other distributions that will give similar results are Gamma and Weibull, and we can get similar results for these.
- Another possibility is to first sample N , from some distribution. Then sample the time steps from a Dirichlet distribution.
- A large value of γ means that the algorithm will take more, smaller time steps on average.

Optimal simulation parameters

- In choosing the sampling distribution of the time steps we introduce simulation parameters that should be chosen in a good way.
- One needs to consider that the choice of parameters will affect the complexity of the method.
- If N is the random number of time steps, let the complexity be $E[N]$

Optimization problem

- The efficiency of the algorithm is

$$\frac{1}{\text{complexity} \times \text{variance}}$$

- Let $V(p)$ denote the variance of a single sample from our simulation method, where p is the simulation parameter to be chosen optimally.
- The optimization problem is

$$\min_p E[N] V(p).$$

- We can find bounds on $V(p)$ and find optimum numerically.
- The main heuristic conclusion is that a non-regular sde requires smaller time steps.

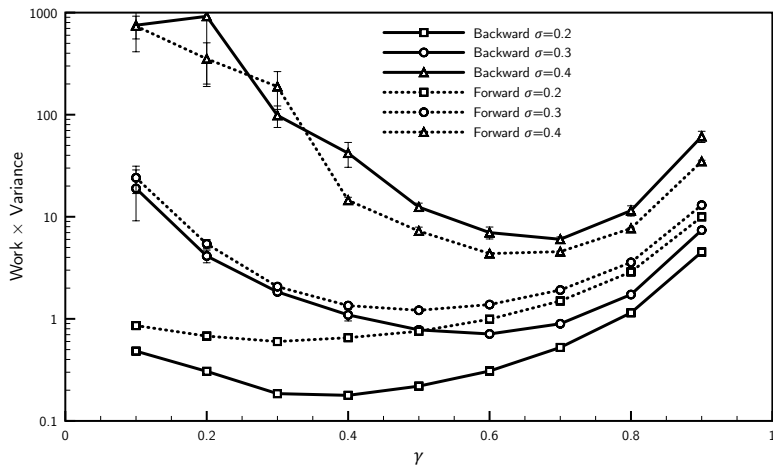
Numerical experiments - Choosing simulation parameters

- We consider

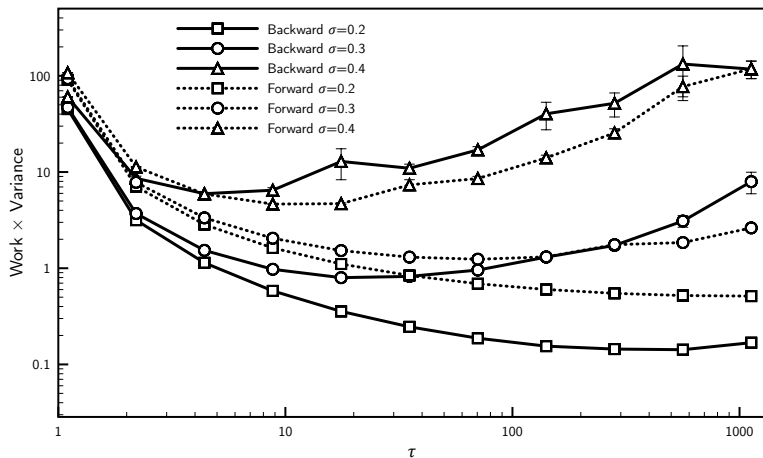
$$dX_t = \sigma (\sin(\omega X_t) + 2) dW_t,$$

- We simulate $P(X_0 - I < X_T < X_0 + I)$, where I is such that the probability is approximately 0.5
- Our metric is standard deviation for a 1 second run time.
- We use three different parameter sets, $\sigma = \omega = 0.2, 0.3$ and 0.4 .

Varying γ , when sampling time steps from Beta

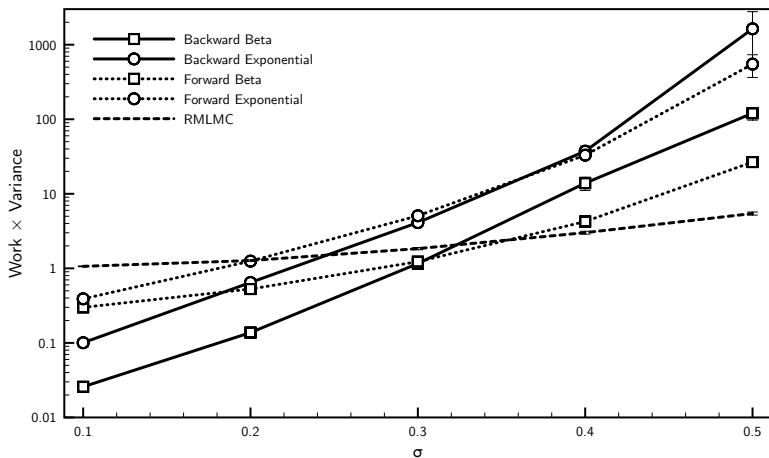


Varying $\bar{\tau}$, when sampling time steps from Beta



Using optimal simulation parameters and varying

$$\sigma = \omega$$



Numerical experiments - Convergence rate

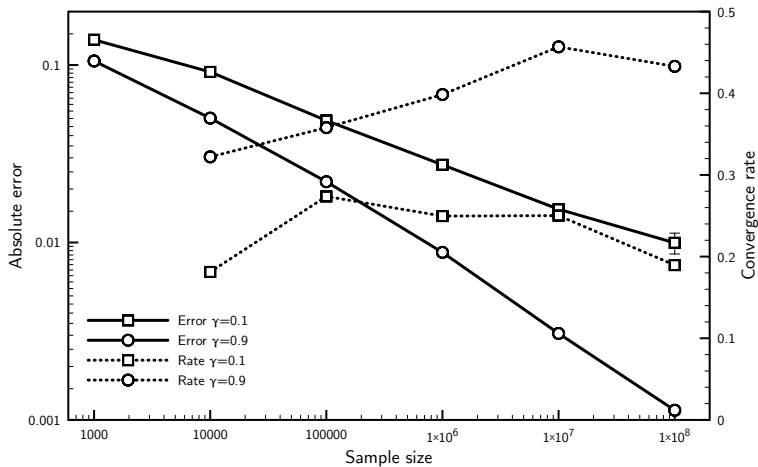
Now, consider

$$dX_t = k(X_0 - X_t)dt + \sigma\sqrt{|X_0 - X_t|^{1/4} + 1}dW_t,$$

where $k = 1.5$, $X_0 = 1$, $\sigma = 0.01$.

- Simulate $P(X_0 - l < X_T < X_0 + l)$.
- Diffusion not differentiable.
- Drift and diffusion not bounded.
- Parameters chosen so that path is close to non differentiable point.
- Choosing γ large enough should give finite variance.

Numerical experiments - Convergence rate



Numerical experiments - Conclusion

- Larger σ gives higher variance and requires smaller time steps.
- The performance gets worse fast as σ increases.
- At least for larger σ , Beta outperforms Exp.
- Choosing γ large enough will give close to optimal convergence rate.

Sampling the space integral

There are various methods for sampling the space integral to reduce the variance.

- Use that if $f(x) \equiv 1$,

$$1 = \mathbb{E} [f(X_T)] = \mathbb{E} \left[\frac{1}{\rho_N(\tau_1, \dots, \tau_N)} \prod_{j=0}^{N-1} \theta_{\tau_{j+1}-\tau_j}(X_{\tau_j}^\pi, X_{\tau_{j+1}}^\pi) \right],$$

which can be used as a control variate.

- Sequential resampling on $\left| \theta_{\tau_{j+1}-\tau_j}(X_{\tau_j}^\pi, X_{\tau_{j+1}}^\pi) \right|$.
- Analysis of $\theta_t(x, y)$ can suggest importance sampling distributions.

Conclusions

- We studied the performance of the parametrix method
- Exponential sampling may work well in certain situations, for example if the diffusion coefficient is constant. In general it may give infinite variance.
- Choosing a sampling distribution that gives smaller time steps can produce a method with finite variance and faster convergence.
- When parameters get large, variance increases fast.
- Simulations seem to agree with general heuristics.

Thanks for listening.

References

- Bally, V. and Kohatsu-Higa, A. (2015). A probabilistic interpretation of the parametrix method. *Ann. Appl. Probab.*, 25(6):3095–3138.
- Rhee, C.-H. and Glynn, P. W. (2015). Unbiased estimation with square root convergence for sde models. *Oper. Res.*, 63(5):1026–1043.