Stochastic Simulation: Lecture 15

Prof. Mike Giles

Oxford University Mathematical Institute

Objective

Want to estimate expectations

$$\mathbb{E}[P(X)]$$

where X is a random variable from a distribution for which the p.d.f. $p(x)$ falls into one of the following two categories:

- $p(x) \propto \exp(-V(x))$, but the constant of proportionality is unknown
- $p(x) = \lim_{T\to\infty} p_T(x)$ where $p_T(x)$ is the p.d.f. of solutions to an SDE at time T, subject to initial data X_0 at time $t=0$

Objective

Let's start by considering the autonomous SDE

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t$$

subject to initial data X_0.

The Fokker-Planck (forward Kolmogorov) equation for the p.d.f. $p(x,t)$ is

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}(f_i\, p) + \sum_{i,j} \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}(D_{ij}\, p)$$

where $D = \frac{1}{2} g\, g^T$, and subject to initial data $p(x,0) = \delta(x - X_0)$

Objective

If the drift $f(x)$ is locally Lipschitz and satisfies the dissipative condition

$$\langle x, f(x) \rangle \leq -\alpha \|x||^2 + \beta$$

for some $\alpha, \beta > 0$, and $g(x)$ is bounded, then the system is ergodic and the p.d.f. $p(x,t)$ converges exponentially to an invariant distribution $p_\infty(x)$,

$$\| \, p(x,t) - p_\infty(x) \, \| \;\; \propto \;\; \exp(-\lambda_1 \, t)$$

where λ_1 is the smallest eigenvalue given by

$$-\lambda_1 \, p = -\sum_i \frac{\partial}{\partial x_i} (f_i \, p) + \sum_{i,j} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} (D_{ij} \, p)$$

Objective

Our objective is to estimate $\mathbb{E}[P(X)]$ where the expectation is with respect to the invariant distribution $p_\infty(x)$.

We want to do this in high dimensions, so can't afford to solve the steady state PDE. Instead, we want to estimate

$$\lim_{T\to\infty} \mathbb{E}\left[P(X_T)\right]$$

How best to do this?

MLMC

Answer: use MLMC (of course!)

Initial idea:

- use standard MLMC for SDEs (e.g. $h_\ell \propto 2^{-\ell}$) for fixed T, to estimate $\mathbb{E}[P(X_T)]$
- increase T, see how much the answer changes, estimate error due to T, repeat as needed

Two problems:

- not elegant
- there is a more fundamental problem in some cases

MLMC

The elegant solution: change both the timestep and T with level, for example

$$h_\ell = 2^{-\ell}, \qquad T_\ell = (\ell+1)\log 2 \,/\, \lambda_1$$

Why this choice? Because weak error is

$$O\left(h_L + \exp(-\lambda_1\, T_L)\right) = O(2^{-L})$$

(In practice, can err on the side of underestimating λ_1)

This way we automatically achieve weak convergence as $L \to \infty$, but now we have a new problem: coupling between coarse and fine paths

MLMC

If we start the coarse and fine path calculations at time $t=0$, but run the fine path for longer, then $\widehat{X}^{(f)}_{T_\ell} - \widehat{X}^{(c)}_{T_{\ell-1}}$ will not be small.

The solution to this comes from a paper by Glynn & Rhee on Markov chains, which was adapted to SDEs by Fang & Giles.

Start fine path calculation at time $-T_\ell$, and start coarse path a bit later at time $-T_{\ell-1}$.

Both paths share the same driving Brownian motion W_t for $-T_{\ell-1} < t < 0$, and the "final" values are taken at time 0.

MLMC

This approach works well for contractive SDEs for which

$$\langle x-y, f(x)-f(y), \rangle + \tfrac{1}{2}\|g(x)-g(y)\|^2 \le -\lambda \|x-y\|^2$$

for $\lambda > 0$, in which case for $t > s$

$$\mathbb{E}\left[\, \|X_t - Y_t\|^2 \right] < \exp(-2\,\lambda(t-s))\; \|X_s - Y_s\|^2$$

so the effect of the initial difference $X^{(f)}_{-T_{\ell-1}} - X^{(c)}_{-T_{\ell-1}}$ decays exponentially.

Bonus: under the same contraction condition, we have uniform strong convergence so that we can achieve

$$\mathbb{E}\left[\, \|\widehat{X}^{(f)} - \widehat{X}^{(c)}\|^2 \right] < c\, h_\ell^{2\sigma}$$

where σ is the strong order of convergence

MLMC

This last point hints at the second more fundamental problem: what happens when the SDE is not contractive?

Going back to simulating on a time interval $[0, T]$, standard numerical analysis gives

$$\mathbb{E}\left[\| \widehat{X}_T^{(f)} - \widehat{X}_T^{(c)} \|^2 \right] < c \, \exp(2\mu T) \; h_\ell^{2\sigma}$$

for some $\mu > 0$.

This exponential growth in time really can happen – consider a chaotic ODE to which a little noise is added.

So, increasing T to reduce the weak error can make the MLMC variance much worse, to the point that MLMC is useless.

MLMC

This problem can be fixed by introducing a "spring" between the coarse and fine paths:

$$\mathrm{d}X_t^{(f)} = \left(f(X_t^{(f)}) + \sigma \, (X_t^{(c)} - X_t^{(f)}) \right) \mathrm{d}t + g(X_t^{(f)}) \, \mathrm{d}W_t$$

$$\mathrm{d}X_t^{(c)} = \left(f(X_t^{(c)}) + \sigma \, (X_t^{(f)} - X_t^{(c)}) \right) \mathrm{d}t + g(X_t^{(c)}) \, \mathrm{d}W_t$$

A strong enough spring constant σ ensures the two paths do not diverge exponentially.

The introduction of the spring implies a change of measure – this is corrected for by a Radon-Nikodym derivative so final output is

$$R_\ell \, P(X_\ell) - R_{\ell-1} \, P(X_{\ell-1})$$

This works, both in theory and in practice, even for chaotic SDEs.

MCMC

When g is the identity matrix, if the SDE can be written as

$$\mathrm{d}X_t = -\tfrac{1}{2}\nabla V(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

then the invariant probability distribution is proportional to $\exp(-V(x))$.

An alternative approach to sampling from this distribution is the Markov Chain Monte Method – the simplest example is the Metropolis-Hastings algorithm:

- Generate a candidate X^* (e.g. from Normal distribution centred on X^n)
- With probability $\min(1, \exp(-V(X^*))/\exp(-V(X^n)))$ set $X^{n+1} = X^*$; otherwise set $X^{n+1} = X^n$

MCMC

This produces a sequence of values X^n, whose distribution approaches the target distribution after an initial "burn-in" period.

Note: successive values are clearly strongly correlated, the variance based on N consecutive sample is greater than it would be for independent samples.

The art of MCMC is in the construction of good candidates; too small a step means it takes many steps to sample the whole distribution, but too large a step leads to frequent rejection and hence less movement.

$O(\varepsilon^{-2})$ complexity if samples have $O(1)$ cost, but multilevel ideas have been developed for most costly applications

Key references

P.W. Glynn and C.-H. Rhee. 'Exact estimation for Markov chain equilibrium expectations'. Journal of Applied Probability, 51A:377-389, 2014.

W. Fang, M.B. Giles. 'Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift'. Annals of Applied Probability, 30(2):526-560, 2020.

W. Fang, M.B. Giles. 'Multilevel Monte Carlo method for ergodic SDEs without contractivity'. Journal of Mathematical Analysis and Applications, 476(1):149-176, 2019.

Key references

https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo

V.H. Hoang, Ch. Schwab, A.M. Stuart. 'Complexity analysis of accelerated MCMC methods for Bayesian inversion'. Inverse Problems, 29(8), 2013.

T.J. Dodwell, C. Ketelsen, R. Scheichl, A.L. Teckentrup. 'A hierarchical multilevel Markov Chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow'. SIAM/ASA Journal on Uncertainty Quantification, 3(1):1075-1108, 2015

A. Jasra, K. Kamatani, K.J.H. Law, Y. Zhou. 'A multi-index Markov chain Monte Carlo method'. International Journal for Uncertainty Quantification, 8(1):61-73, 2018.