

Two new MLMC applications

Mike Giles

Mathematical Institute, University of Oxford

International Conference on Monte Carlo Techniques

Paris, July 7, 2016

Outline

- multilevel Monte Carlo
- multi-dimensional reflected diffusions
 - ▶ numerical discretisations
 - ▶ adaptive timestepping
 - ▶ numerical analysis and results
- expected value of perfect partial information
 - ▶ formulation
 - ▶ MLMC for nested simulation
 - ▶ numerical analysis and results

Multilevel Monte Carlo

MLMC is based on the telescoping sum

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^L \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$$

where \widehat{P}_ℓ represents an approximation of some output P on level ℓ .

In simple SDE applications with uniform timestep $h_\ell = 2^{-\ell} h_0$,
if the weak convergence is

$$\mathbb{E}[\widehat{P}_\ell - P] = O(2^{-\alpha\ell}),$$

and \widehat{Y}_ℓ is an unbiased estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$, based on N_ℓ samples,
with variance

$$\mathbb{V}[\widehat{Y}_\ell] = O(N_\ell^{-1} 2^{-\beta\ell}),$$

and expected cost

$$\mathbb{E}[C_\ell] = O(N_\ell 2^{\gamma\ell}), \quad \dots$$

Multilevel Monte Carlo

... then the finest level L and the number of samples N_ℓ on each level can be chosen to achieve an RMS error of ε at an expected cost

$$C = \begin{cases} O(\varepsilon^{-2}), & \beta > \gamma, \\ O(\varepsilon^{-2}(\log \varepsilon)^2), & \beta = \gamma, \\ O(\varepsilon^{-2-(\gamma-\beta)/\alpha}), & 0 < \beta < \gamma. \end{cases}$$

I always try to get $\beta > \gamma$, so the main cost comes from the coarsest levels – use of QMC can then give substantial additional benefits.

With $\beta > \gamma$, can also randomise levels to eliminate bias (Rhee & Glynn, Operations Research, 2015).

Multilevel Monte Carlo

The standard estimator for SDE applications is

$$\hat{Y}_\ell = N_\ell^{-1} \sum_{n=0}^{N_\ell} \left(\hat{P}_\ell(W^{(n)}) - \hat{P}_{\ell-1}(W^{(n)}) \right)$$

using the same Brownian motion $W^{(n)}$ for the n^{th} sample on the fine and coarse levels.

However, there is some freedom in how we construct the coupling provided \hat{Y}_ℓ is an unbiased estimator for $\mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1}]$.

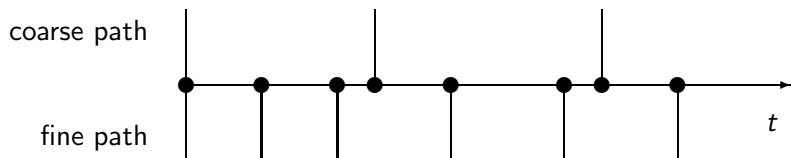
Have exploited this with an antithetic estimator for multi-dimensional SDEs which don't satisfy the commutativity condition.

(G, Szpruch: AAP 2014)

Multilevel Monte Carlo

Also, uniform timestepping is not required – it is fairly straightforward to implement MLMC using non-nested adaptive timestepping.

(G, Lester, Whittle: MCQMC14 proceedings)



This has been exploited for SDEs with a drift which is not globally Lipschitz. (Fang, G: new arXiv paper)

Also, interesting possibilities for applications with discontinuous output functionals.

Reflected diffusions

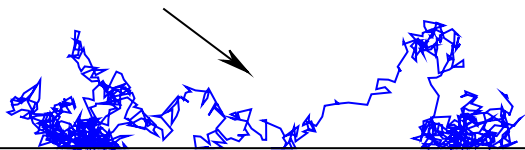
Motivating application comes from modelling of network queues
(Kavita Ramanan, Brown University)

Reflected Brownian diffusion with constant volatility in a domain D has SDE

$$dx_t = a(x_t) dt + b dW_t + \nu(x_t) dL_t,$$

where L_t is a local time which increases when x_t is on the boundary ∂D .

$\nu(x)$ can be normal to the boundary (pointing inwards), but in some cases it is not and reflection from the boundary includes a tangential motion.

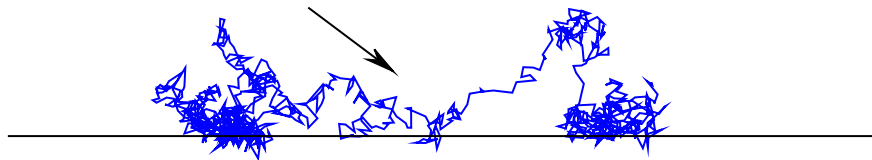


Reflected diffusions

A penalised version is

$$\begin{aligned} dx_t &= a(x_t) dt + b dW_t + \nu(x_t) dL_t, \\ dL_t &= \lambda \max(0, -d(x_t)) dt, \quad \lambda \gg 1 \end{aligned}$$

where $d(x)$ is signed distance to the boundary (negative means outside) and $\nu(x)$ is a smooth extension from the boundary into the exterior.



Reflected diffusions

When D is a polygonal domain, this generalises to

$$dx_t = a(x_t) dt + b dW_t + \sum_{k=1}^K \nu_k(x_t) dL_{k,t},$$

with a different ν_k and local time $L_{k,t}$ for each boundary face.

The corresponding penalised version is

$$\begin{aligned} dx_t &= a(x_t) dt + b dW_t + \sum_{k=1}^K \nu_k(x_t) dL_{k,t}, \\ dL_{k,t} &= \lambda \max(0, -d_k(x_t)) dt, \quad \lambda \gg 1 \end{aligned}$$

where $d_k(x_t)$ is signed distance to the boundary face with a suitable extension.

Numerical approximations

3 different numerical treatments in literature:

- projection (Gobet, Słomiński): predictor step

$$\widehat{X}^{(p)} = \widehat{X}_{t_n} + a(\widehat{X}_{t_n}, t_n) h_n + b \Delta W_n,$$

followed by correction step

$$\widehat{X}_{t_{n+1}} = \widehat{X}^{(p)} + \nu(\widehat{X}^{(p)}) \Delta \widehat{L}_n,$$

with $\Delta \widehat{L}_n > 0$ if needed to put $\widehat{X}_{t_{n+1}}$ on boundary

- reflection (Gobet): similar but with double the value for $\Delta \widehat{L}_n$
– can give improved $O(h)$ weak convergence
- penalised (Słomiński): Euler-Maruyama approximation of penalised SDE with $\lambda = O(h^{-1})$, giving convergence as $h \rightarrow 0$

Numerical approximations

Concern:

- because b is constant, Euler-Maruyama method corresponds to first order Milstein scheme, suggesting an $O(h)$ strong error
- however, all three treatments of boundary reflection lead to a strong error which is $O(h^{1/2})$ – this is based primarily on empirical evidence, with only limited supporting theory
- if the output quantity of interest is Lipschitz with respect to the path then

$$\mathbb{V} \left[\widehat{P} - P \right] \leq \mathbb{E} \left[(\widehat{P} - P)^2 \right] \leq c^2 \mathbb{E} \left[\sup_{[0, T]} (\widehat{X}_t - X_t)^2 \right]$$

so the variance is $O(h)$

- OK, but not great – would like $O(h^\beta)$ with $\beta > 1$ for $O(\varepsilon^{-2})$ MLMC complexity

Adaptive timesteps

Simple idea: use adaptive timestep based on distance from the boundary

- far away, use uniform timestep $h_\ell = 2^{-\ell} h_0$
- near the boundary, use uniform timestep $h_\ell = 2^{-2\ell} h_0$
- in between, define $h_\ell(x)$ to vary smoothly based on distance $d(x)$

What do we hope to achieve?

- strong error $O(2^{-\ell}) \implies$ MLMC variance is $O(2^{-2\ell})$
- computational cost per path $O(2^\ell)$
- $\beta=2, \gamma=1$ in MLMC theorem \implies complexity is $O(\varepsilon^{-2})$

Adaptive timesteps

In intermediate zone, want negligible probability of taking a single step and crossing the boundary.

Stochastic increment in Euler timestep is $b \Delta W$, so define h_ℓ so that

$$(\ell+3) \|b\|_2 \sqrt{h_\ell} = d$$

Final 3-zone max-min definition of h_ℓ is

$$h_\ell = \max \left(2^{-2\ell} h_0, \min \left(2^{-\ell} h_0, (d / ((\ell+3) \|b\|_2)^2) \right) \right)$$

Balancing terms, gives

- boundary zone up to $d = O(2^{-\ell})$
- intermediate zone up to $d = O(2^{-\ell/2})$

Adaptive timesteps

Balancing terms, gives

- boundary zone up to $d \approx O(2^{-\ell})$
- intermediate zone up to $d \approx O(2^{-\ell/2})$

If $\rho(y, t)$, the density of paths at distance y from the boundary at time t , is uniformly bounded then the computational cost per unit time is approximately

$$\int_0^\infty \frac{\rho(y, t) dy}{h_\ell(y)} \sim \underbrace{2^{2\ell} \times 2^{-\ell}}_{\text{boundary}} + \underbrace{\int_{O(2^{-\ell})}^{O(2^{-\ell/2})} \frac{dy}{y^2}}_{\text{intermediate}} + \underbrace{2^\ell \times 1}_{\text{interior}} \approx O(2^\ell)$$

so we get similar cost contributions from all 3 zones.

Numerical analysis

Theorem (Computational cost)

If

- *the density $\rho(y, t)$ for the SDE paths at distance y from the boundary is uniformly bounded*
- *the numerical discretisation with the adaptive timestep has strong convergence $O(2^{-\ell})$*

then *the computational cost is $o(2^{(1+\delta)\ell})$ for any $0 < \delta \ll 1$.*

The second condition is needed to bound the difference between the distributions of the paths and their numerical approximations.

Numerical analysis

Theorem (Strong convergence)

If

- *the drift a is constant*
- *a uniform timestep discretisation has $O(h^{1/2})$ strong convergence*
- *the adaptive timestep h_ℓ is rounded to the nearest multiple of the boundary zone timestep*

then *the strong convergence is $O(2^{-\ell})$*

The proof is based on a comparison with a discretisation using the uniform boundary zone timestep:

- adaptive numerical discretisation is exact when boundary not crossed
- almost zero probability of crossing the boundary unless in the boundary zone using the uniform timestep

Numerical analysis

Future challenges:

- prove that for constant drift a and timestep h , the strong error is $O(h^{1/2})$ for reflected diffusions with oblique reflections, preferably for generalised penalisation method for polygonal boundaries
- extend analysis to include errors in local time
- extend analysis to general drift and adaptive timesteps

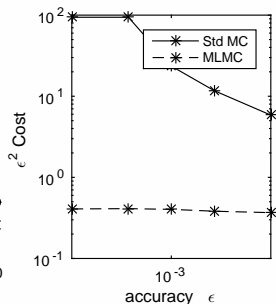
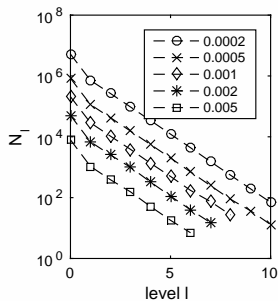
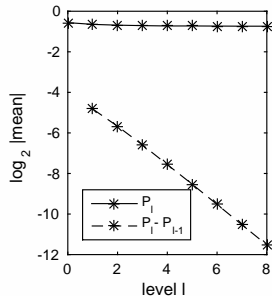
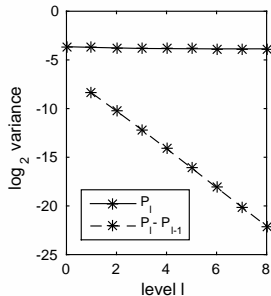
Numerical results

Simple test case:

- 3D Brownian motion in a unit ball
- normal reflection at the boundary
- $x_0 = 0$
- aim is to estimate $\mathbb{E}[\|x\|_2^2]$ at time $t=1$.
- implemented with both projection and penalisation schemes

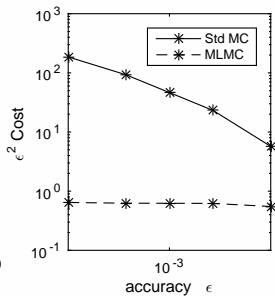
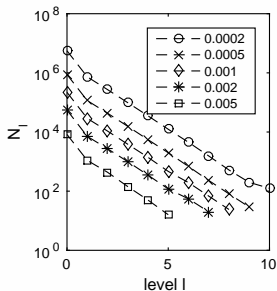
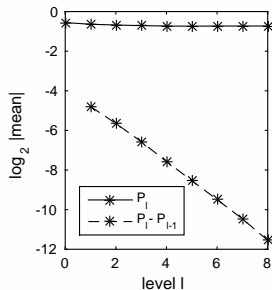
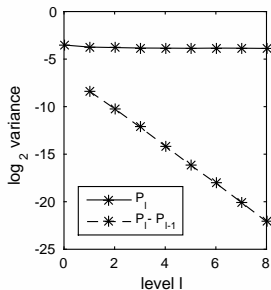
Numerical results

Projection method:



Numerical results

Penalisation method:



Decision making under uncertainty

The motivating application comes from medical decision-making (Howard Thom, Bristol University).

Given no knowledge of independent uncertainties X, Y , best treatment out of some finite set D corresponds to

$$\max_{d \in D} \mathbb{E} [f_d(X, Y)]$$

while with perfect knowledge we have

$$\mathbb{E} \left[\max_{d \in D} f_d(X, Y) \right].$$

However, if X is known but not Y , then best treatment has value

$$\mathbb{E} \left[\max_d \mathbb{E} [f_d(X, Y) | X] \right].$$

EVPI & EVPPI

EVPI, the expected value of perfect information, is the difference

$$\text{EVPI} = \mathbb{E} \left[\max_d f_d(X, Y) \right] - \max_d \mathbb{E}[f_d(X, Y)]$$

which can be estimated with $O(\varepsilon^{-2})$ complexity by standard methods, assuming an $O(1)$ cost per sample $f_d(X, Y)$.

EVPPI, the expected value of partial perfect information, is the difference

$$\text{EVPPI} = \mathbb{E} \left[\max_d \mathbb{E}[f_d(X, Y) | X] \right] - \max_d \mathbb{E}[f_d(X, Y)]$$

which is a nested simulation problem. In practice, we choose to estimate

$$\text{EVPI} - \text{EVPPI} = \mathbb{E} \left[\max_d f_d(X, Y) \right] - \mathbb{E} \left[\max_d \mathbb{E}[f_d(X, Y) | X] \right]$$

MLMC treatment

Based on work by Oxford colleagues (Bujok, Hambly, Reisinger, 2015) Takashi Goda (arXiv, April 2016) has proposed an efficient MLMC estimator using 2^ℓ samples on level ℓ for conditional expectation.

For given sample X , define

$$Z_\ell = \frac{1}{2} \left(\max_d \overline{f}_d^{(a)} + \max_d \overline{f}_d^{(b)} \right) - \max_d \overline{f}_d$$

where

- $\overline{f}_d^{(a)}$ is an average of $f_d(X, Y)$ over $2^{\ell-1}$ independent samples for Y ;
- $\overline{f}_d^{(b)}$ is an average over a second independent set of $2^{\ell-1}$ samples;
- \overline{f}_d is an average over the combined set of 2^ℓ inner samples.

MLMC treatment

The expected value of this estimator is

$$\mathbb{E}[Z_\ell] = \mathbb{E}[\max_d \bar{f}_{d,2^{\ell-1}}] - \mathbb{E}[\max_d \bar{f}_{d,2^\ell}]$$

where $\bar{f}_{d,2^\ell}$ is an average of 2^ℓ inner samples, and hence

$$\begin{aligned} \sum_{\ell=1}^L \mathbb{E}[Z_\ell] &= \mathbb{E}[\max_d f] - \mathbb{E}[\max_d \bar{f}_{d,2^L}] \\ &\rightarrow \mathbb{E}[\max_d f] - \mathbb{E}[\max_d \mathbb{E}[f(X, Y) | X]] \end{aligned}$$

as $L \rightarrow \infty$, giving us the desired estimate.

MLMC treatment

How good is the estimator? $\gamma=1$, but what are α and β ?

Define

$$F_d(X) = \mathbb{E}[f_d(X, Y) | X], \quad d_{opt}(X) = \arg \max_d F_d(X)$$

so $d_{opt}(x)$ is piecewise constant, with a lower-dimensional manifold K on which it is not uniquely-defined.

Note that $\frac{1}{2}(\overline{f_d^{(a)}} + \overline{f_d^{(b)}}) - \overline{f_d} = 0$, so $Z_\ell = 0$ if the same d maximises each term in Z_ℓ .

Numerical analysis

Heuristic analysis:

- $\overline{f_d}^{(a)} - \overline{f_d}^{(b)} = O(2^{-\ell/2})$, due to CLT
- $O(2^{-\ell/2})$ probability of both being near K
- under this condition, $Z_\ell = O(2^{-\ell/2})$
- Hence $\mathbb{E}[Z_\ell] = O(2^{-\ell})$ and $\mathbb{E}[Z_\ell^2] = O(2^{-3\ell/2})$, so $\alpha=1, \beta=3/2$.

It is possible to make this rigorous given some assumptions.

Numerical analysis

Assumptions

- $\mathbb{E} [|f_d(X, Y)|^p]$ is finite for all $p \geq 2$.

Comment: helps to bound the difference between \overline{f}_d and $F_d(X)$.

- There exists a constant $c_0 > 0$ such that for all $0 < \epsilon < 1$

$$\mathbb{P}(\min_{y \in K} \|X - y\| \leq \epsilon) \leq c_0 \epsilon.$$

Comment: bounds the probability of X being close to K .

- There exist constants $c_1, c_2 > 0$ such that if $X \notin K$, then

$$\max_d F_d(X) - \max_{d \neq d_{\text{opt}}(X)} F_d(X) > \min(c_1, c_2 \min_{y \in K} \|X - y\|).$$

Comment: ensure linear separation of the optimal F_d away from K .

Numerical analysis

Building on the heuristic analysis, and other past analyses, we obtain the following theorem:

Theorem

If the Assumptions are satisfied, and $\overline{f}_d^{(a)}$, $\overline{f}_d^{(b)}$, \overline{f}_d are as defined previously for level ℓ , with 2^ℓ inner samples being used for \overline{f}_d , then for any $\delta > 0$

$$\mathbb{E} \left[\frac{1}{2} (\max_d \overline{f}_d^{(a)} + \max_d \overline{f}_d^{(b)}) - \max_d \overline{f}_d \right] = o(2^{-(1-\delta)\ell}).$$

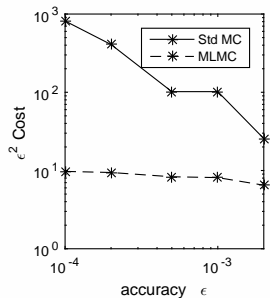
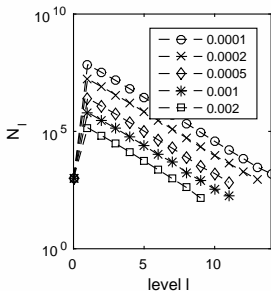
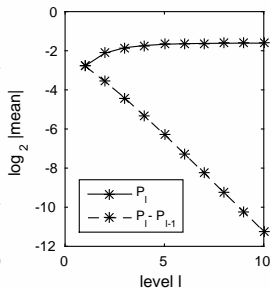
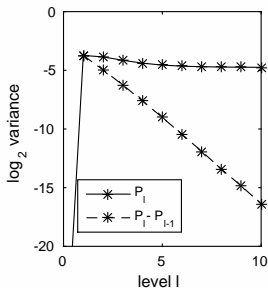
and

$$\mathbb{E} \left[\left(\frac{1}{2} (\max_d \overline{f}_d^{(a)} + \max_d \overline{f}_d^{(b)}) - \max_d \overline{f}_d \right)^2 \right] = o(2^{-(3/2-\delta)\ell}).$$

Numerical results

Goda test case:

- $X \sim N(0, 2)$
 $Y \sim N(0, 3)$
- $f_1(X, Y) = X + Y$
 $f_2(X, Y) = 0$
- As expected
 $\alpha \approx 1$
 $\beta \approx 3/2$
- $O(\varepsilon^{-2})$ and $O(\varepsilon^{-3})$
complexity for MLMC
and MC, respectively.



Nested simulation

There are lots of other applications involving nested simulation.

One example is a McKean-Vlasov equation of the form

$$dX_t = a(X_t, \mathbb{E}[X_t]) dt + b(X_t, \mathbb{E}[X_t]) dW$$

which can be simulated with P paths

$$\hat{X}_{p,n+1} = \hat{X}_{p,n} + a \left(\hat{X}_{p,n}, \frac{1}{P} \sum_{q=1}^P \hat{X}_{q,n} \right) h + b \left(\hat{X}_{p,n}, \frac{1}{P} \sum_{q=1}^P \hat{X}_{q,n} \right) \Delta W_{p,n}$$

Different levels can use different values for P and/or different timesteps – a good application for multi-index Monte Carlo?

Nested simulation

In other applications in finance and big data, each outer “sample” requires a summation over a large set of data:

$$S = \sum_1^N x(n)$$

This can be represented as

$$S = N \mathbb{E}[x(n)]$$

where n is uniformly distributed over $\{1, 2, 3, \dots, N\}$. It can then be approximated by

$$\frac{N}{P} \sum_1^P x(n_p)$$

and different levels can again use different values for P .

Conclusions

- still lots of new directions for MLMC research
- adaptive timestepping can be helpful, and is easy to implement
- nested simulation may be a significant new direction – splitting a large sample into multiple sub-samples is key to a good coupling
- I think multi-index Monte Carlo (MIMC) will be important for nested simulation and other areas

Webpages:

<http://people.maths.ox.ac.uk/gilesm/>

http://people.maths.ox.ac.uk/gilesm/mlmc_community.html