Using Multilevel Monte Carlo to estimate Expected Value of Partial Perfect Information

Mike Giles

Mathematical Institute, University of Oxford

University of Bristol seminar

January 20, 2017

Outline

- introduction to MLMC ideas
- MLMC for nested simulation
- MLMC for EVPPI
- numerical analysis and simple testcase
- extension to MCMC inner sampling

Monte Carlo method

In stochastic models, we often have

ω	$\longrightarrow$	S	$\longrightarrow$	P
random input		intermediate variables		scalar output

The Monte Carlo estimate for $\mathbb{E}[P]$ is an average of N independent samples $\omega^{(n)}$:

$$Y = N^{-1} \sum_{n=1}^{N} P(\omega^{(n)}).$$

This is unbiased, $\mathbb{E}[Y]=\mathbb{E}[P]$, and the Central Limit Theorem proves that as $N \to \infty$ the error becomes Normally distributed with variance $N^{-1}\mathbb{V}[P]$.

Monte Carlo method

In many cases, this is modified to

ω	$\longrightarrow$	$\widehat{S}$	$\longrightarrow$	$\widehat{P}$
random input		intermediate variables		scalar output

where $\widehat{S}, \widehat{P}$ are approximations to S, P, in which case the MC estimate

$$\widehat{Y} = N^{-1} \sum_{n=1}^{N} \widehat{P}(\omega^{(n)})$$

is biased, and the Mean Square Error is

$$\mathbb{E}[\,(\widehat{Y} - \mathbb{E}[P])^2] \;=\; N^{-1}\, \mathbb{V}[\widehat{P}] + \left(\mathbb{E}[\widehat{P}] - \mathbb{E}[P]\right)^2$$

Greater accuracy requires larger N and smaller weak error $\mathbb{E}[\widehat{P}] - \mathbb{E}[P]$.

SDE Path Simulation

My interest was in SDEs (stochastic differential equations) for finance, which in a simple one-dimensional case has the form

$$\mathrm{d}S_t = a(S_t, t)\, \mathrm{d}t + b(S_t, t)\, \mathrm{d}W_t$$

Here $\mathrm{d}W_t$ is the increment of a Brownian motion – Normally distributed with variance $\mathrm{d}t$.

This is usually approximated by the simple Euler-Maruyama method

$$\widehat{S}_{t_{n+1}} = \widehat{S}_{t_n} + a(\widehat{S}_{t_n}, t_n)\, h + b(\widehat{S}_{t_n}, t_n)\, \Delta W_n$$

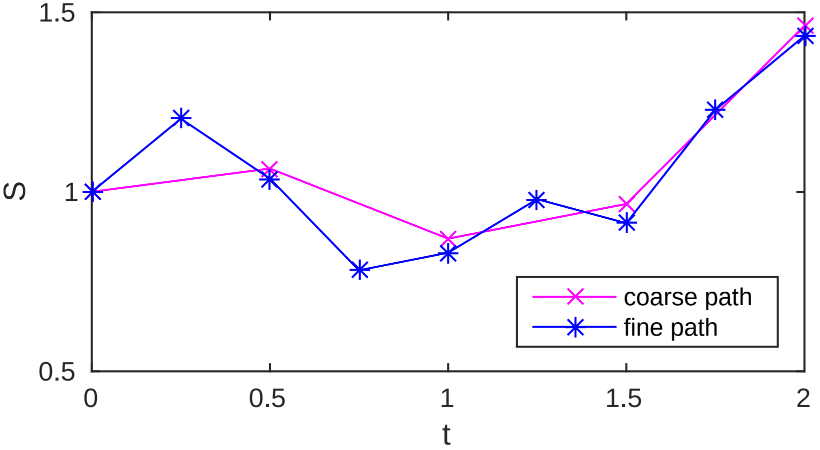
with uniform timestep h, and increments ΔW_n with variance h.

In simple applications, the output of interest is a function of the final value:

$$\widehat{P} \equiv f(\widehat{S}_T)$$

SDE Path Simulation

Geometric Brownian Motion: $\mathrm{d}S_t = r\, S_t\, \mathrm{d}t + \sigma\, S_t\, \mathrm{d}W_t$



SDE Path Simulation

Two kinds of discretisation error:

Weak error:

$$\mathbb{E}[\widehat{P}] - \mathbb{E}[P] = O(h)$$

Strong error:

$$\left(\mathbb{E}\left[\sup_{[0,T]} \left(\widehat{S}_t - S_t \right)^2 \right] \right)^{1/2} = O(h^{1/2})$$

For reasons which will become clear, I prefer to use the Milstein discretisation for which the weak and strong errors are both $O(h)$.

SDE Path Simulation

The Mean Square Error is

$$N^{-1}\,\mathbb{V}[\widehat{P}] + \left(\mathbb{E}[\widehat{P}] - \mathbb{E}[P]\right)^2 \;\approx\; a\,N^{-1} + b\,h^2$$

If we want this to be ε^2, then we need

$$N = O(\varepsilon^{-2}), \qquad h = O(\varepsilon)$$

so the total computational cost is $O(\varepsilon^{-3})$.

To improve this cost we need to

- reduce N – variance reduction or Quasi-Monte Carlo methods
- reduce the cost of each path (on average) – MLMC

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[\widehat{P}_1]$ but it is much cheaper to simulate $\widehat{P}_0 \approx \widehat{P}_1$, then since

$$\mathbb{E}[\widehat{P}_1] = \mathbb{E}[\widehat{P}_0] + \mathbb{E}[\widehat{P}_1 - \widehat{P}_0]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} \widehat{P}_0^{(0,n)} \; + \; N_1^{-1} \sum_{n=1}^{N_1} \left(\widehat{P}_1^{(1,n)} - \widehat{P}_0^{(1,n)} \right)$$

Benefit: if $\widehat{P}_1 - \widehat{P}_0$ is small, its variance will be small, so won't need many samples to accurately estimate $\mathbb{E}[\widehat{P}_1 - \widehat{P}_0]$, so cost will be reduced greatly.

Multilevel Monte Carlo

Natural generalisation: given a sequence $\widehat{P}_0, \widehat{P}_1, \ldots, \widehat{P}_L$

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

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} \widehat{P}_0^{(0,n)} \; + \; \sum_{\ell=1}^{L} \left\{ N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}_\ell^{(\ell,n)} - \widehat{P}_{\ell-1}^{(\ell,n)} \right) \right\}$$

with independent estimation for each level of correction

Multilevel Monte Carlo

If we define

- C_0, V_0 to be cost and variance of $\widehat{P}_0$
- C_ℓ, V_ℓ to be cost and variance of $\widehat{P}_\ell - \widehat{P}_{\ell-1}$

then the total cost is $\displaystyle\sum_{\ell=0}^{L} N_\ell\, C_\ell$ and the variance is $\displaystyle\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$.

Using a Lagrange multiplier μ^2 to minimise the cost for a fixed variance

$$\frac{\partial}{\partial N_\ell} \sum_{k=0}^{L} \left(N_k\, C_k + \mu^2 N_k^{-1} V_k \right) = 0$$

gives

$$N_\ell = \mu \sqrt{V_\ell / C_\ell} \quad \Longrightarrow \quad N_\ell\, C_\ell = \mu \sqrt{V_\ell\, C_\ell}$$

Multilevel Monte Carlo

Setting the total variance equal to ε^2 gives

$$\mu = \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell\, C_\ell} \right)$$

and hence, the total cost is

$$\sum_{\ell=0}^{L} N_\ell \; C_\ell \;=\; \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell C_\ell} \right)^2$$

in contrast to the standard cost which is approximately $\varepsilon^{-2}\, V_0\, C_L$.

The MLMC cost savings are therefore approximately:

- V_L/V_0, if $\sqrt{V_\ell C_\ell}$ increases with level
- C_0/C_L, if $\sqrt{V_\ell C_\ell}$ decreases with level

Multilevel Path Simulation

With SDEs, level ℓ corresponds to approximation using M^ℓ timesteps, giving approximate payoff $\widehat{P}_\ell$ at cost $C_\ell = O(h_\ell^{-1})$.

Simplest estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ for $\ell > 0$ is

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

using same driving Brownian path for both levels.

Analysis gives $\text{MSE} = \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell + \left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P] \right)^2$

To make RMS error less than ε

- choose $N_\ell \propto \sqrt{V_\ell / C_\ell}$ so total variance is less than $\frac{1}{2}\,\varepsilon^2$
- choose L so that $\left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P] \right)^2 < \frac{1}{2}\,\varepsilon^2$

Multilevel Path Simulation

For Lipschitz payoff functions $P \equiv f(S_T)$, we have

$$
\begin{aligned}
V_\ell \;\equiv\; \mathbb{V}\left[\widehat{P}_\ell - \widehat{P}_{\ell-1}\right] \;&\leq\; \mathbb{E}\left[(\widehat{P}_\ell - \widehat{P}_{\ell-1})^2\right] \\
&\leq\; K^2\, \mathbb{E}\left[(\widehat{S}_{T,\ell} - \widehat{S}_{T,\ell-1})^2\right] \\
&=\; \begin{cases} O(h_\ell), & \text{Euler-Maruyama} \\ O(h_\ell^2), & \text{Milstein} \end{cases}
\end{aligned}
$$

and hence

$$
V_\ell\, C_\ell = \begin{cases} O(1), & \text{Euler-Maruyama} \\ O(h_\ell), & \text{Milstein} \end{cases}
$$

MLMC Theorem

(Slight generalisation of version in 2008 *Operations Research* paper)

If there exist independent estimators $\widehat{Y}_\ell$ based on N_ℓ Monte Carlo samples, each costing C_ℓ, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$ and

i) $\left| \mathbb{E}[\widehat{P}_\ell - P] \right| \leq c_1 \, 2^{-\alpha \, \ell}$

ii) $\mathbb{E}[\widehat{Y}_\ell] = \begin{cases} \mathbb{E}[\widehat{P}_0], & \ell = 0 \\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_\ell] \leq c_2 \, N_\ell^{-1} 2^{-\beta \, \ell}$

iv) $\mathbb{E}[C_\ell] \leq c_3 \, 2^{\gamma \, \ell}$

MLMC Theorem

then there exists a positive constant c_4 such that for any $\varepsilon < 1$ there exist L and N_ℓ for which the multilevel estimator

$$\widehat{Y} = \sum_{\ell=0}^{L} \widehat{Y}_\ell,$$

has a mean-square-error with bound $\mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$

with an expected computational cost C with bound

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > \gamma, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4\, \varepsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

MLMC Theorem

Two observations of optimality:

- MC simulation needs $O(\varepsilon^{-2})$ samples to achieve RMS accuracy ε. When $\beta > \gamma$, the cost is optimal — $O(1)$ cost per sample on average.

  (Would need multilevel QMC to further reduce costs)

- When $\beta < \gamma$, another interesting case is when $\beta = 2\alpha$, which corresponds to $\mathbb{E}[\widehat{Y}_\ell]$ and $\sqrt{\mathbb{E}[\widehat{Y}_\ell^2]}$ being of the same order as $\ell \to \infty$. In this case, the total cost is $O(\varepsilon^{-\gamma/\alpha})$, which is the cost of a single sample on the finest level — again optimal.

MLMC work on SDEs

- Milstein discretisation for path-dependent options – G (2008)
- numerical analysis – G, Higham, Mao (2009), Avikainen (2009), G, Debrabant, Rößler (2012)
- financial sensitivities ("Greeks") – Burgos (2011)
- jump-diffusion models – Xia (2011)
- Lévy processes – Dereich (2010), Marxen (2010), Dereich & Heidenreich (2011), Xia (2013), Kyprianou (2014)
- American options – Belomestny & Schoenmakers (2011)
- Milstein in higher dimensions without Lévy areas – G, Szpruch (2014)
- adaptive timesteps – Hoel, von Schwerin, Szepessy, Tempone (2012), G, Lester, Whittle (2014), Fang, G (2016)

MLMC for SPDEs

- quite natural application, with better cost savings than SDEs due to higher dimensionality
- range of applications
  - Graubner & Ritter (Darmstadt) – parabolic
  - G, Reisinger (Oxford) – parabolic
  - Cliffe, G, Scheichl, Teckentrup (Bath/Nottingham) – elliptic
  - Barth, Jenny, Lang, Meyer, Mishra, Müller, Schwab, Sukys, Zollinger (ETH Zürich) – elliptic, parabolic, hyperbolic
  - Harbrecht, Peters (Basel) – elliptic
  - Efendiev (Texas A&M) – numerical homogenization
  - Vidal-Codina, G, Peraire (MIT) – reduced basis approximation

Other MLMC applications

- parametric integration, integral equations (Heinrich)
- multilevel QMC (Dick, G, Kuo, Scheichl, Schwab, Sloan)
- stochastic chemical reactions (Anderson & Higham, Tempone)
- mixed precision computation on FPGAs (Korn, Ritter, Wehn)
- MLMC for MCMC (Scheichl, Schwab, Stuart, Teckentrup)
- Coulomb collisions in plasma (Caflisch)
- invariant distribution of contractive Markov process (Glynn & Rhee)
- invariant distribution of contractive SDEs (G, Lester & Whittle)

These are all discussed in my 70-page Acta Numerica review article (2015).

Nested simulation

Nested simulation is interested in the estimation of

$$\mathbb{E}\left[g\left(\mathbb{E}[f(X,Y)\,|\,X]\right)\right]$$

for independent random variables X, Y.

If each individual $f(X,Y)$ can be sampled at unit cost then an MLMC treatment can use 2^ℓ samples on level ℓ.

For given sample X, a good "antithetic" estimator is

$$Z_\ell \;=\; g(\overline{f}) \;-\; \tfrac{1}{2}\left(g(\overline{f}^{(a)}) + g(\overline{f}^{(b)})\right)$$

where

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

Nested simulation

Note that

$$\overline{f} = \tfrac{1}{2}\left(\overline{f}^{(a)} + \overline{f}^{(b)}\right),$$

so that

$$\begin{aligned}\overline{f}^{(a)} &= \overline{f} + \tfrac{1}{2}\left(\overline{f}^{(a)} - \overline{f}^{(b)}\right),\\ \overline{f}^{(b)} &= \overline{f} - \tfrac{1}{2}\left(\overline{f}^{(a)} - \overline{f}^{(b)}\right).\end{aligned}$$

Doing a Taylor series expansion about $\overline{f}$ then gives

$$Z_\ell \approx \tfrac{1}{2}\, g''(\overline{f})\ \left(\overline{f}^{(a)} - \overline{f}^{(b)}\right)^2 = O(2^{-\ell})$$

which gives $\alpha = 1, \beta = 2, \gamma = 1$, and hence an $O(\varepsilon^{-2})$ complexity.

This has been used for pedestrian "flow" by Haji-Ali (2012) and credit modelling by Bujok, Hambly & Reisinger (2015).

EVPPI

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

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

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

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

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

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

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

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

MLMC treatment

Takashi Goda (University of Tokyo, 2016) proposed an MLMC estimator using 2^ℓ samples on level ℓ for conditional expectation.

For given sample X, define

$$Z_\ell \;=\; \tfrac{1}{2}\left(\max_d \overline{f_d}^{(a)} + \max_d \overline{f_d}^{(b)} \right) \;-\; \max_d \overline{f_d}$$

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

$$\mathbb{E}[Z_\ell] = \mathbb{E}[\max_d \overline{f_d}_{,2^{\ell-1}}] - \mathbb{E}[\max_d \overline{f_d}_{,2^\ell}]$$

where $\overline{f_d}_{,2^\ell}$ is an average of 2^ℓ inner samples, and hence

$$\begin{aligned}\sum_{\ell=1}^{L} \mathbb{E}[Z_\ell] &= \mathbb{E}[\max_d f] - \mathbb{E}[\max_d \overline{f_d}_{,2^L}] \\ &\to \mathbb{E}[\max_d f] - \mathbb{E}\left[\max_d \mathbb{E}[f(X,Y)\,|\,X]\right]\end{aligned}$$

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

MLMC treatment

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

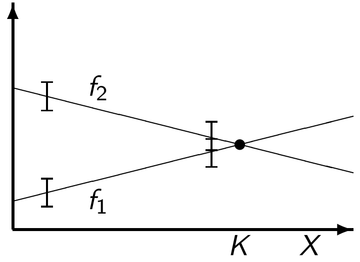
Define

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

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

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

Numerical analysis



Heuristic analysis:

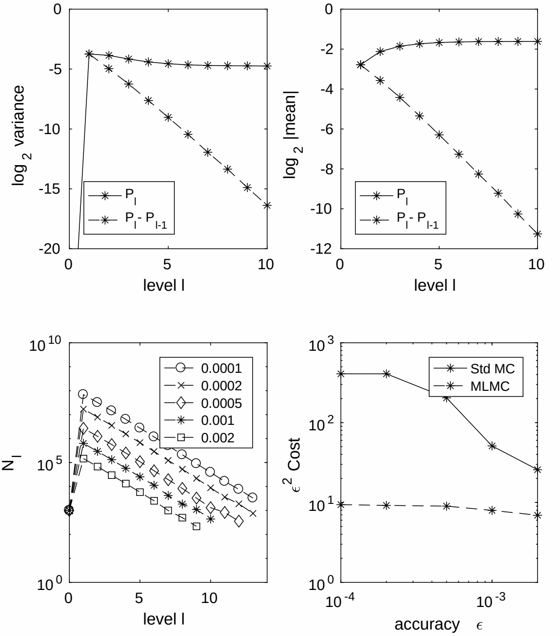
- $\overline{f_d}^{(a)} - \overline{f_d}^{(b)} = O(2^{-\ell/2})$, due to CLT
- $O(2^{-\ell/2})$ probability of being within $O(2^{-\ell/2})$ of K
- under this condition, $Z_\ell = O(2^{-\ell/2})$; otherwise, $Z_\ell = 0$
- hence $\mathbb{E}[Z_\ell] = O(2^{-\ell})$ and $\mathbb{E}[Z_\ell^2] = O(2^{-3\ell/2})$, so $\alpha\!=\!1, \beta\!=\!3/2$.

It is possible to make this rigorous given some assumptions.

Numerical 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.



MCMC extension

With some models it is not possible to generate iid samples of $f(X,Y)$, but instead can generate a MCMC sequence of samples which is asymptotically from the correct distribution.

In this case, for given sample X, define

$$Z_\ell \;=\; \max_d \overline{f_d}^{(c)} \;-\; \max_d \overline{f_d}$$

where

- $\overline{f_d}^{(c)}$ is an average of first $2^{\ell-1}$ MCMC samples $f_d(X,Y)$;
- $\overline{f_d}$ is an average over the first 2^ℓ MCMC samples.

MCMC extension

Under certain conditions, MCMC analysis gives

$$\overline{f_d}^{(c)} - \overline{f_d} = O(2^{-\ell/2})$$

similar to what one would get from iid samples.

This leads to $Z_\ell = O(2^{-\ell/2})$ for all samples.

Hence we have $\alpha=1, \beta=1, \gamma=1$ giving $O(\varepsilon^{-2}|\log\varepsilon|^2)$ complexity.

There's a slight possibility of improved complexity if the MCMC chains are contracting – need to investigate this.

Conclusions

- MLMC is a major direction in Monte Carlo research, with lots of applications
- in Oxford we have developed supporting software in MATLAB, C/C++, R and Python
- MLMC works well for nested simulation, and gives optimal $O(\varepsilon^{-2})$ complexity for EVPPI estimation
- supporting numerical analysis has already been developed
- can use MCMC sampling for inner estimate (new)
- Dr. Takashi Goda from Univ. of Tokyo is collaborating in the project

Webpages:

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

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