Multilevel Monte Carlo Simulation

Mike Giles
mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute
Oxford-Man Institute of Quantitative Finance

Multilevel MC Approach

Suppose we want to estimate $\mathbb{E}[P]$ where $P(\omega)$ can be simulated numerically with different levels of accuracy, and corresponding costs, giving $\hat{P}_l$, $l = 0, 1, \ldots, L$.

$$\mathbb{E}[\hat{P}_L] = \mathbb{E}[\hat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\hat{P}_l - \hat{P}_{l-1}]$$

Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key idea: approximate $\mathbb{E}[\hat{P}_l - \hat{P}_{l-1}]$ using $N_l$ simulations with $\hat{P}_l$ and $\hat{P}_{l-1}$ obtained using same underlying sample $\omega$.

$$\hat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left( \hat{P}^{(i)}_l - \hat{P}^{(i)}_{l-1} \right)$$
Using independent samples for each level, the variance of the combined estimator is

\[
\mathbb{V} \left[ \sum_{l=0}^{L} \hat{Y}_l \right] = \sum_{l=0}^{L} N_l^{-1} V_l, \quad V_l \equiv \begin{cases} 
\mathbb{V}[\hat{P}_l - \hat{P}_{l-1}], & l > 0 \\
\mathbb{V}[\hat{P}_0], & l = 0 
\end{cases}
\]

and the computational cost is \( \sum_{l=0}^{L} N_l C_l \), where \( C_l \) is the cost of a single sample.

Hence, the variance is minimised for a fixed computational cost by choosing \( N_l \) to be proportional to \( \sqrt{V_l/C_l} \).
Multilevel MC Approach

Since

\[
\mathbb{E} \left[ (\hat{Y} - \mathbb{E}[P])^2 \right] = \mathbb{V}[\hat{Y}] + \left( \mathbb{E}[\hat{P}_L] - \mathbb{E}[P] \right)^2
\]

can choose

- constant of proportionality for \( N_l \) so that \( \mathbb{V}[\hat{Y}] \approx \frac{1}{2} \varepsilon^2 \)

- finest level \( L \) so that \( \left( \mathbb{E}[\hat{P}_L - P] \right)^2 \approx \frac{1}{2} \varepsilon^2 \)

to get Mean Square Error equal to \( \varepsilon^2 \)
Previous work

- First paper (Operations Research, 2006 – 2008) applied idea to SDE path simulation using Euler-Maruyama discretisation

- Second paper (MCQMC 2006 – 2007) used Milstein discretisation for scalar SDEs – improved strong convergence gives improved multilevel variance convergence

- Multilevel method is a generalisation of two-level control variate method of Kebaier (2005), and similar to ideas of Speight (2009)

- Also related to multilevel parametric integration by Heinrich (2001)
Multilevel Theorem

**Theorem:** Given multilevel estimators $\hat{Y}_l$ based on $N_l$ samples, each with cost $C_l$, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ with $\alpha \geq \frac{1}{2} \gamma$, such that

i) $\left| \mathbb{E}[\hat{P}_l - P] \right| \leq c_1 2^{-\alpha l}$

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

iii) $\mathbb{V}[\hat{Y}_l] \leq c_2 N_l^{-1} 2^{-\beta l}$

iv) $C_l \leq c_3 2^{\gamma l}$
then there is constant $c_4$ such that for any $\varepsilon < e^{-1}$ there are values $L$ and $N_l$ for which the multilevel estimator

$$\hat{Y} = \sum_{l=0}^{L} \hat{Y}_l,$$

with Mean Square Error $MSE \equiv \mathbb{E} \left[ (\hat{Y} - \mathbb{E}[P])^2 \right] < \varepsilon^2$

with a computational cost $C$ with bound

$$C \leq \begin{cases} 
    c_4 \varepsilon^{-2}, & \beta > \gamma, \\
    c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\
    c_4 \varepsilon^{-2} - (\gamma - \beta) / \alpha, & 0 < \beta < \gamma.
\end{cases}$$
In multilevel path simulations for scalar SDEs such as
\[ dS = a(S, t) \, dt + b(S, t) \, dW, \quad 0 \leq t \leq T, \]
each level typically uses twice as many timesteps as the previous, so \( \gamma = 1. \)

Question then is: what is \( \beta? \)

\[ V_l \propto 2^{-\beta l} \propto h_l^\beta \]

where \( h_l \) is timestep on level \( l. \)
For applications in which $P$ is a Lipschitz function of $S(T)$, value of underlying path simulation at a fixed time, strong convergence property

$$\left(\mathbb{E} \left[ (\hat{S}_N - S(T))^2 \right] \right)^{1/2} = O(h^\omega)$$

implies that

$$\nabla[\hat{P}_l - P] = O(h_2^{2\omega})$$

and hence

$$\nabla[\hat{P}_l - \hat{P}_{l-1}] = O(h_2^{2\omega})$$

and therefore $\beta = 2\omega$. 
## Multilevel path simulation

<table>
<thead>
<tr>
<th>option</th>
<th>Euler numerics</th>
<th>Euler analysis</th>
<th>Milstein numerics</th>
<th>Milstein analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz</td>
<td>$O(h)$</td>
<td>$O(h)$</td>
<td>$O(h^2)$</td>
<td>$O(h^2)$</td>
</tr>
<tr>
<td>Asian</td>
<td>$O(h)$</td>
<td>$O(h)$</td>
<td>$O(h^2)$</td>
<td>$O(h^2)$</td>
</tr>
<tr>
<td>lookback</td>
<td>$O(h)$</td>
<td>$O(h)$</td>
<td>$O(h^2)$</td>
<td>$o(h^{3/2})$</td>
</tr>
<tr>
<td>digital</td>
<td>$O(h^{1/2})$</td>
<td>$O(h^{1/2} \log h)$</td>
<td>$O(h^3/2)$</td>
<td>$o(h^{3/2})$</td>
</tr>
</tbody>
</table>

Table: convergence for $V_I$ as observed numerically and proved analytically for both the Euler and Milstein discretisations. $\delta$ can be any strictly positive constant.
Multilevel path simulation

Analysis for Euler discretisations:
- lookback and barrier: Giles, Higham & Mao (*Finance & Stochastics*, 2009)

Analysis for Milstein discretisations:
- Giles, Debrabant & Rößler (TU Darmstadt)
- multilevel estimator for path-dependent options based on conditional Brownian interpolation within timesteps (or extrapolation in final timestep)
Milstein Scheme

Brownian interpolation: within each timestep, model the behaviour as simple Brownian motion (i.e. constant drift and volatility) conditional on the two end-points

\[
\hat{S}(t) = \hat{S}_n + \lambda(t)(\hat{S}_{n+1} - \hat{S}_n) + b_n \left( W(t) - W_n - \lambda(t)(W_{n+1} - W_n) \right),
\]

where \( \lambda(t) = \frac{t - t_n}{t_{n+1} - t_n} \).

There then exist analytic results for the distribution of the min/max/average over each timestep, and probability of crossing a barrier.
Milstein Scheme

**Theorem:** Under standard conditions,

\[
\mathbb{E} \left[ \sup_{[0,T]} \left| \hat{S}(t) - S(t) \right|^m \right] = O((h \log h)^m),
\]

\[
\sup_{[0,T]} \mathbb{E} \left[ \left| \hat{S}(t) - S(t) \right|^m \right] = O(h^m),
\]

\[
\mathbb{E} \left[ \left( \int_0^T \hat{S}(t) - S(t) \, dt \right)^2 \right] = O(h^3).
\]
Milstein Scheme

The variance convergence for the Asian option comes directly from this.

Will now outline the analysis for the lookback option – the barrier is similar but more complicated.

The digital option is based on a Brownian extrapolation from one timestep before the end – the analysis is similar.

The analysis for the lookback, barrier and digital options uses the idea of “extreme” paths which are highly improbable – the variance comes mainly from non-extreme paths for which one can use asymptotic analysis.
Milstein Scheme

Computing $\hat{P}_l - \hat{P}_{l-1}$ requires a fine and coarse path simulation for the same underlying Brownian motion.

On the fine path, the minimum over one timestep is

$$\tilde{S}_{n,min}^f = \frac{1}{2} \left( \hat{S}_n^f + \hat{S}_{n+1}^f - \sqrt{\left( \hat{S}_{n+1}^f - \hat{S}_n^f \right)^2 - 2 \left( b_n^f \right)^2 h_l \log U_n} \right)$$

where $U_m$ is a $(0,1]$ uniform random variable.

For the coarse path, first define $\hat{S}_n^c$ for odd $n$ using conditional Brownian interpolation, then use the same expression for the minimum with same $U_n$. 
Milstein Scheme

**Theorem:** For any $\gamma > 0$, the probability that $W(t)$, its increments $\Delta W_n$ and the corresponding SDE solution $S(t)$ and approximations $\hat{S}_n^f$ and $\hat{S}_n^c$ satisfy any of the following “extreme” conditions

$$\max_n \left( \max(|S(nh)|, |\hat{S}_n^f|, |\hat{S}_n^c|) \right) > h^{-\gamma}$$

$$\max_n \left( \max(|S(nh) - \hat{S}_n^c|, |S(nh) - \hat{S}_n^f|, |\hat{S}_n^f - \hat{S}_n^c|) \right) > h^{1-\gamma}$$

$$\max_n |\Delta W_n| > h^{1/2 - \gamma}$$

is $o(h^p)$ for all $p > 0$. 
Furthermore, there exist constants $c_1, c_2, c_3, c_4$ such that if none of these conditions is satisfied, and $\gamma < \frac{1}{2}$, then

\[
\max_n |\hat{S}_n^f - \hat{S}_{n-1}^f| \leq c_1 h^{1/2 - 2\gamma}
\]

\[
\max_n |b_n^f - b_{n-1}^f| \leq c_2 h^{1/2 - 2\gamma}
\]

\[
\max_n \left( |b_n^f| + |b_n^c| \right) \leq c_3 h^{-\gamma}
\]

\[
\max_n |b_n^f - b_n^c| \leq c_4 h^{1/2 - 2\gamma}
\]

where $b_n^c$ is defined to equal $b_{n-1}^c$ if $n$ is odd.
Milstein Scheme

The lookback analysis splits paths into:

- **“extreme”** paths, which have such low probability that their contribution to the variance is negligible \((o(h^p) \text{ for any } p > 0)\)

- non-extreme paths for which it can be proved that

\[
\left| \hat{S}_{f,\min} - \hat{S}^c_{\min} \right| \leq \max_n \left| \hat{S}^f_{n,\min} - \hat{S}^c_{n,\min} \right|
\]

\[
= o(h^{1-\delta/2})
\]

for any \(\delta > 0\), and hence \(V_l = o(h_l^{2-\delta})\).
Currently working with Christoph Reisinger on an SPDE application which arises in CDO modelling (Bush, Hambly, Haworth & Reisinger)

\[
dp = -\mu \frac{\partial p}{\partial x} \, dt + \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \, dt + \sqrt{\rho} \frac{\partial p}{\partial x} \, dW
\]

with absorbing boundary \( p(0, t) = 0 \)

- derived in limit as number of firms \( \rightarrow \infty \)
- \( x \) is distance to default
- \( p(x, t) \) is probability density function
- \( dW \) term corresponds to systemic risk
- \( \partial^2 p/\partial x^2 \) comes from idiosyncratic risk
SPDE application

- numerical discretisation combines Milstein time-marching with central difference approximations
- coarsest level of approximation uses 1 timestep per quarter, and 10 spatial points
- each finer level uses four times as many timesteps, and twice as many spatial points – ratio is due to numerical stability constraints
- mean-square stability theory, with and without absorbing boundary
- computational cost $C_l \propto 8^l$
- numerical results suggest variance $V_l \propto 8^{-l}$
- can prove $V_l \propto 16^{-l}$ when no absorbing boundary
Fractional loss on equity tranche of a 5-year CDO:

\[
\log_2 \text{variance} \quad \log_2 |\text{mean}| \\
\begin{array}{c}
\cdot \quad P_l \\
- \cdot \quad P_l - P_{l-1}
\end{array}
\]
SPDE application

Fractional loss on equity tranche of a 5-year CDO:

![Graph showing the relationship between level l and accuracy ε, with cost ε^2 plotted against accuracy ε.](Multilevel Monte Carlo – p. 22/25)
Future work

- “vibrato” technique for digital options:
  - current treatment uses conditional expectation one timestep before maturity, which smooths the payoff
  - the “vibrato” idea generalises this to cases without a known conditional expectation

- Greeks:
  - the multilevel approach should work well, combining pathwise sensitivities with “vibrato” treatment to cope with lack of smoothness
  - can also incorporate the adjoint approach developed with Paul Glasserman – more efficient when many Greeks are wanted for one payoff function
Future work

- variance-gamma, CGMY and other Lévy processes:
  - given exact simulation techniques, multilevel benefit is in treating path-dependent options
  - could also handle addition of a local vol surface

- American options – the next big challenge:
  - instead of Longstaff-Schwartz approach, view it as an exercise boundary optimisation problem, and use multilevel optimisation?
Conclusions

Multilevel Monte Carlo method has already achieved
- improved order of complexity
- significant benefits for model problems

but there is still a lot more research to be done, both theoretical and applied.


Papers are available from:
www.maths.ox.ac.uk/~gilesm/finance.html