Numerical Analysis of Multilevel Monte Carlo Path Simulation

Mike Giles

(joint work with Kristian Debrabant & Andreas Rößler, TU Darmstadt)

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

IMACS Seminar on Monte Carlo Methods

Brussels, Sept 7 – 11, 2009

Approach

Given a scalar SDE driven by a Brownian diffusion

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t),$$

to estimate $\mathbb{E}[P]$ where the path-dependent payoff P can be approximated by $\widehat{P}_l$ using 2^l uniform timesteps, we use

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

$\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ is estimated using N_l simulations with same $W(t)$ for both $\widehat{P}_l$ and $\widehat{P}_{l-1}$,

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right)$$

Approach

Using independent samples for each level, the variance of the combined estimator is

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

and the computational cost is proportional to $\displaystyle\sum_{l=0}^{L} N_l\, h_l^{-1}$

Hence, the variance is minimised for a fixed computational cost by choosing N_l to be proportional to $\sqrt{V_l\, h_l}$.

Approach

Since

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

can choose

- constant of proportionality for N_l so that $\mathbb{V}[\widehat{Y}] \approx \frac{1}{2}\varepsilon^2$
- finest level L so that $\left(\mathbb{E}[\widehat{P}_L - P]\right)^2 \approx \frac{1}{2}\varepsilon^2$

to get Mean Square Error approximately equal to ε^2

MLMC Theorem

Theorem: *Let* P *be a functional of the solution of a stochastic o.d.e., and* $\widehat{P}_l$ *the discrete approximation using a timestep* $h_l = 2^{-l}\, T$.

If *there exist independent estimators* $\widehat{Y}_l$ *based on* N_l *Monte Carlo samples, with computational complexity (cost)* C_l, *and positive constants* $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ *such that*

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

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

iii) $\mathbb{V}[\widehat{Y}_l] \leq c_2\, N_l^{-1} h_l^\beta$

iv) $C_l \leq c_3\, N_l\, h_l^{-1}$

MLMC Theorem

then *there exists a positive constant* c_4 *such that for any* $\varepsilon\!<\!e^{-1}$ *there are values* L *and* N_l *for which the multilevel estimator*

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

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

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > 1, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4\, \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

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)

Numerical Analysis

If P is a Lipschitz function of $S(T)$, value of underlying path simulation at a fixed time, the strong convergence property

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

implies that $\mathbb{V}[\widehat{P}_l - P] = O(h_l^{2\gamma})$ and hence

$$V_l \;\equiv\; \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l^{2\gamma}).$$

Therefore $\beta = 1$ for Euler-Maruyama discretisation,
and $\beta = 2$ for the Milstein discretisation.

However, in general, good strong convergence is neither necessary nor sufficient for good convergence for V_l.

Numerics and Analysis

option	Euler numerics	Euler analysis	Milstein numerics	Milstein analysis
Lipschitz	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
Asian	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
lookback	$O(h)$	$O(h)$	$O(h^2)$	$o(h^{2-\delta})$
barrier	$O(h^{1/2})$	$o(h^{1/2-\delta})$	$O(h^{3/2})$	$o(h^{3/2-\delta})$
digital	$O(h^{1/2})$	$O(h^{1/2}\log h)$	$O(h^{3/2})$	$o(h^{3/2-\delta})$

Table: V_l convergence observed numerically (for GBM) and proved analytically (for more general SDEs) for both the Euler and Milstein discretisations. δ can be any strictly positive constant.

Numerical Analysis

Analysis for Euler discretisations:

- lookback and barrier options: Giles, Higham & Mao (*Finance & Stochastics, 2009*)
  - lookback analysis follows from strong convergence
  - barrier analysis shows dominant contribution comes from paths which are near the barrier; uses asymptotic analysis, first proving that “extreme” paths have negligible contribution
  - similar analysis for digital options gives $O(h^{1/2-\delta})$ bound instead of $O(h^{1/2}\log h)$
- digital options: Avikainen (*Finance & Stochastics, 2009*)
  - method of analysis is quite different

Numerical Analysis

Analysis for Milstein discretisations:

- work in progress by Giles, Debrabant & Rößler
- uses boundedness of all moments to bound the contribution to V_l from "extreme" paths
(e.g. for which $\max_n |\Delta W_n| > h^{1/2-\delta}$ for some $\delta > 0$)
- uses asymptotic analysis to bound the contribution from paths which are not "extreme"

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

$$
\begin{aligned}
\widehat{S}(t) &= \widehat{S}_n + \lambda(t)(\widehat{S}_{n+1} - \widehat{S}_n) \\
&\qquad + b_n \Big(W(t) - W_n - \lambda(t)(W_{n+1} - W_n) \Big),
\end{aligned}
$$

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

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

Milstein Scheme

The Brownian interpolant is different from the standard Kloeden-Platen interpolant defined as

$$\begin{aligned}\widehat{S}_{KP}(t) \;&=\; \widehat{S}_n + a_n\,(t-t_n) + b_n\,(W(t)-W_n) \\ &\quad + \tfrac{1}{2}\,b'_n\,b_n\,\left((W(t)-W_n)^2 - (t-t_n)\right),\end{aligned}$$

for which, under the usual conditions,

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

Milstein Scheme

Theorem: Under standard conditions,

i)

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

ii)

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

iii)

$$\mathbb{E}\left[\left(\int_0^T \widehat{S}(t)-\widehat{S}_{KP}(t)\,\mathrm{d}t\right)^2\right] = O(h^3).$$

Milstein Scheme

Proof:

$$\widehat{S}(t) - \widehat{S}_{KP}(t) = \tfrac{1}{2}\, b_n' b_n\, Y(t),$$

where

$$\begin{aligned} Y(t) &= \lambda\,(W_{n+1}\!-\!W_n)^2 - (W(t)\!-\!W_n)^2 \\ &= \lambda\,(1\!-\!\lambda)\,(W_{n+1}\!-\!W_n)^2 - \left(W(t) - W_n - \lambda\,(W_{n+1}\!-\!W_n)\right)^2 \\ &\quad - 2\,\lambda\,(W_{n+1}\!-\!W_n)\,(W(t) - W_n - \lambda\,(W_{n+1}\!-\!W_n))\ . \end{aligned}$$

with $\lambda(t) = (t\!-\!t_n)/(t_{n+1}\!-\!t_n)$ as before.

Milstein Scheme

i) Assertion follows from

$$\mathbb{E}\left[\sup_{[0,T]} \left|\hat{S}(t) - \hat{S}_{KP}(t)\right|^m\right]$$

$$\leq \quad 2^{-m} \sqrt{\mathbb{E}\left[\max_n |b'_n b_n|^{2m}\right] \; \mathbb{E}\left[\sup_{[0,T]} |Y(t)|^{2m}\right]}$$

and an extreme value theory bound for $\mathbb{E}\left[\sup_{[0,T]} |Y(t)|^{2m}\right]$

ii) Assertion follows from

$$\mathbb{E}\left[\,\left|\hat{S}(t) - \hat{S}_{KP}(t)\right|^m\right] = 2^{-m}\ \mathbb{E}[\,|b'_n b_n|^m]\ \mathbb{E}[\,|Y|^m]$$

Milstein Scheme

iii) Defining $X_n := \int_{t_n}^{t_{n+1}} Y(t)\,\mathrm{d}t$ gives

$$\mathbb{E}\left[\left(\int_0^T (\hat{S}(t) - \hat{S}_{KP}(t))\,\mathrm{d}t\right)^2\right] = \tfrac{1}{4}\,\mathbb{E}\left[\left(\sum_{n=0}^{N-1} b'_n b_n X_n\right)^2\right].$$

The X_n are iid random variables, and $\mathbb{E}[b'_m b_m X_m b'_n b_n X_n] = 0$ for $n\!>\!m$, so

$$\mathbb{E}\left[\left(\int_0^T (\hat{S}(t) - \hat{S}_{KP}(t))\,\mathrm{d}t\right)^2\right] = \tfrac{1}{4}\,\mathbb{E}[X^2] \sum_{n=0}^{N-1} \mathbb{E}[(b'_n b_n)^2].$$

Result then follows from $\mathbb{E}[X^2] = O(h^4)$.

Milstein Scheme

The variance convergence for the Asian option comes directly from the last result.

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.

Extreme Paths

Lemma: If X_l is a random variable on level l, and $\mathbb{E}[\,|X_l|^m] \leq C_m$ is uniformly bounded, then, for any $\delta > 0$,

$$\mathbb{P}[\,|X_l| > h_l^{-\delta}] = o(h_l^p), \qquad \forall p > 0.$$

Proof: Markov inequality $\mathbb{P}[\,|X_l|^m \!>\! h_l^{-m\delta}] \;<\; h_l^{-m\delta}\, \mathbb{E}[\,|X_l|^m]$

Lemma: If Y_l is a random variable on level l, $\mathbb{E}[\,Y_l^2]$ is uniformly bounded, and the the indicator function $\mathbf{1}_{E_l}$ satisfies $\mathbb{E}[\mathbf{1}_{E_l}] = o(h_l^p),\ \forall p > 0$ then

$$\mathbb{E}[\,|Y_l|\,\mathbf{1}_{E_l}] = o(h_l^p), \qquad \forall p > 0.$$

Proof: Hölder inequality $\mathbb{E}[\,|Y_l|\,\mathbf{1}_{E_l}] \leq \sqrt{\mathbb{E}[\,Y_l^2]\;\mathbb{E}[\mathbf{1}_{E_l}]}$

Extreme Paths

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

$$\begin{aligned}
\max_n \left(\max(|S(nh)|,\ |\widehat{S}_n^f|,\ |\widehat{S}_n^c|) \right) &> h^{-\gamma} \\
\max_n \left(\max(|S(nh) - \widehat{S}_n^c|,\ |S(nh) - \widehat{S}_n^f|,\ |\widehat{S}_n^f - \widehat{S}_n^c|) \right) &> h^{1-\gamma} \\
\max_n |\Delta W_n| &> h^{1/2-\gamma}
\end{aligned}$$

is $o(h^p)$ for all $p > 0$.

Non-extreme paths

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

$$
\begin{aligned}
\max_n |\widehat{S}^f_n - \widehat{S}^f_{n-1}| &\leq c_1\, h^{1/2-2\gamma} \\
\max_n |b^f_n - b^f_{n-1}| &\leq c_2\, h^{1/2-2\gamma} \\
\max_n \left(|b^f_n| + |b^c_n|\right) &\leq c_3\, h^{-\gamma} \\
\max_n |b^f_n - b^c_n| &\leq c_4\, h^{1/2-2\gamma}
\end{aligned}
$$

where b^c_n is defined to equal b^c_{n-1} if n is odd.

Proof: follows from Lemmas and standard bounds.

Lookback Option

Consider a lookback option which is a Lipschitz function of the minimum and final values.

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

On the fine path, the minimum over one timestep is

$$\widehat{S}^f_{n,min} = \tfrac{1}{2}\left(\widehat{S}^f_n + \widehat{S}^f_{n+1} - \sqrt{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)^2 - 2\,(b^f_n)^2\, h_l \log U_n}\;\right)$$

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

For the coarse path, first define $\widehat{S}^c_n$ for odd n using conditional Brownian interpolation, then use the same expression for the minimum with same U_n

Lookback Option

The difference in minimum values is bounded by

$$\begin{aligned}\left|\widehat{S}^f_{min} - \widehat{S}^c_{min}\right| &\leq \max_n \left|\widehat{S}^f_{n,min} - \widehat{S}^c_{n,min}\right| \\ &\leq \max_n \left|\widehat{S}^f_n - \widehat{S}^c_n\right| + \max_n \left|\widehat{D}^f_n - \widehat{D}^c_n\right|,\end{aligned}$$

where

$$\widehat{D}^f_n = \tfrac{1}{2}\sqrt{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)^2 - 2\,(b^f_n)^2\, h_l \log U_n}$$

and $\widehat{D}^c_n$ is defined similarly with b^c_n defined to equal b^c_{n-1} when n is odd.

Lookback Option

We then get

$$
\begin{aligned}
\left| \widehat{D}^f_n - \widehat{D}^c_n \right| \\
&= \frac{\left| (\widehat{D}^f_n)^2 - (\widehat{D}^c_n)^2 \right|}{\widehat{D}^f_n + \widehat{D}^c_n} \\
&\leq \frac{\left| (\widehat{S}^f_{n+1} - \widehat{S}^f_n)^2 - (\widehat{S}^c_{n+1} - \widehat{S}^c_n)^2 \right|}{4(\widehat{D}^f_n + \widehat{D}^c_n)} + \frac{|(b^f_n)^2 - (b^c_n)^2| \; h_l \, |\log U_n|}{2(\widehat{D}^f_n + \widehat{D}^c_n)} \\
&\leq \tfrac{1}{2} \left| |\widehat{S}^f_{n+1} - \widehat{S}^f_n| - |\widehat{S}^c_{n+1} - \widehat{S}^c_n| \right| + \frac{1}{\sqrt{2}} \left| |b^f_n| - |b^c_n| \right| \sqrt{h_l \, |\log U_n|} \\
&\leq \tfrac{1}{2} \left(\left| \widehat{S}^f_{n+1} - \widehat{S}^c_{n+1} \right| + \left| \widehat{S}^f_n - \widehat{S}^c_n \right| \right) + \frac{1}{\sqrt{2}} \, |b^f_n - b^c_n| \, \sqrt{h_l \, |\log U_n|}
\end{aligned}
$$

Lookback Option

Paths are defined to be extreme if they satisfy any of the earlier conditions, or if

$$\max_n |\log U_n| > h_l^{-\gamma}.$$

$\mathbb{E}[(\widehat{P}_l - \widehat{P}_{l-1})^4]$ is bounded and therefore extreme paths have negligible contribution to $\mathbb{E}[(\widehat{P}_l - \widehat{P}_{l-1})^2]$.

For non-extreme paths, by choosing $\gamma = \min(\frac{1}{2}, \delta/5)$, can deduce from the various inequalities that $\max_n |\widehat{S}^f_n - \widehat{S}^c_n|$ and $\max_n |\widehat{S}^f_{n,min} - \widehat{S}^c_{n,min}|$ are $o(h_l^{1-\delta/2})$, so the contribution to $\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2]$ is $o(h_l^{2-\delta})$, and hence $V_l = o(h_l^{2-\delta})$.

Barrier Option

For barrier options, split paths into 3 subsets:

- extreme paths
- paths with a minimum within $O(h^{1/2-\gamma})$ of the barrier
- rest

Assuming $\inf\limits_{[0,T]} S(t)$ has bounded density (at least near the barrier) the dominant contribution comes from the second subset, for which the $O(h)$ difference between $\widehat{S}^f, \widehat{S}^c$ leads to an $O(h^{1/2})$ difference between $\widehat{P}^f, \widehat{P}^c$.

Hence, $V_l = o(h^{3/2-\delta}), \;\; \forall \delta > 0$.

Digital Option

For digital options, again split paths into 3 subsets:

- extreme paths
- paths with final $S(T)$ within $O(h^{1/2-\gamma})$ of the strike
- rest

Assuming $S(T)$ has bounded density near the strike, the dominant contribution again comes from the second subset, where the $O(h)$ difference between $\widehat{S}^f, \widehat{S}^c$ leads to an $O(h^{1/2})$ difference between $\widehat{P}^f, \widehat{P}^c$.

Hence, again, $V_l = o(h^{3/2-\delta}), \;\; \forall \delta > 0$.

Future Work

Milstein scheme for multi-dimensional SDEs generally requires Lévy areas:

$$A_{jk,n} = \int_{t_n}^{t_{n+1}} (W_j(t) - W_j(t_n))\, \mathrm{d}W_k - (W_k(t) - W_k(t_n))\, \mathrm{d}W_j.$$

- $O(h^{1/2})$ strong convergence in general if omitted
- Can still get good convergence for Lipschitz payoffs by using $W^c(t) = \frac{1}{2}(W^{f1}(t) + W^{f2}(t))$ with two fine paths created by antithetic Brownian Bridge construction
- For barrier and digital options, need to simulate Lévy areas – tradeoff between cost and accuracy, optimum may require $O(h^{3/2})$ sub-sampling of Brownian paths, giving $O(h^{3/4})$ strong convergence

Future Work

Greeks:

- the multilevel approach should work well with pathwise sensitivities for Lipschitz payoffs
- “vibrato” treatment (a hybrid combination of pathwise sensitivity and LRM) should handle digital options and second order Greeks
- (can also incorporate the adjoint approach developed with Paul Glasserman – more efficient when many Greeks are wanted for one payoff function)

Future Work

Digital options:

- current treatment uses conditional expectation one timestep before maturity, which smooths the payoff
- in multivariate cases without a known conditional expectation, can use “splitting” (with multiple independent samples for the final timestep of each path) to estimate the conditional expectation
- alternatively, the “vibrato” idea can be generalised, leading to the introduction of a Radon-Nikodym derivative due to a change of measure

Future Work

Other processes:

- finite activity jump diffusion models
  - relatively straightforward if jump rate is not path-dependent
  - trickier if jump rate is path-dependent, but can again use a Radon-Nikodym derivative to force fine and coarse paths to jump at same times
- variance gamma and other Lévy processes
  - standard variance gamma model can be simulated exactly, so multilevel only helpful for path-dependent options – easily analysed?
  - could be harder to analyse if a local volatility surface is introduced

Future Work

The multilevel approach also works well for SPDEs arising in finance

$$\mathrm{d}p = -\mu \, \frac{\partial p}{\partial x} \, \mathrm{d}t + \frac{1}{2} \, \frac{\partial^2 p}{\partial x^2} \, \mathrm{d}t + \sqrt{\rho} \, \frac{\partial p}{\partial x} \, \mathrm{d}W,$$

and oil reservoir modelling

$$\nabla \cdot \left(\kappa \, \nabla p \right) = 0,$$

where $\log \kappa$ is a stochastic field.

However, numerical analysis for these looks very challenging.

Conclusions

- have made progress in numerical analysis of multilevel Monte Carlo path simulation
- excluding the significance of “extreme” paths and using asymptotic analysis for the rest seems a flexible approach to numerical analysis

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