Progress with multilevel Monte Carlo methods

Mike Giles

Des Higham & Xuerong Mao (Strathclyde)

Kristian Debrabant & Andreas Rößler (Darmstadt)

Sylvestre Burgos, Yuan Xia & Christoph Reisinger (Oxford)

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

Fields Institute Workshop on Computational Methods in Finance

Toronto, March 22–24, 2010

Research overview

Long-term objective is faster Monte Carlo simulation of path dependent options to estimate prices and Greeks

Several ingredients, not yet all combined:

- multilevel method
- quasi-Monte Carlo
- adjoint pathwise Greeks
- parallel computing on NVIDIA GPUs

Emphasis in this presentation is on multilevel method

Outline

- multilevel approach
- numerical analysis
- Greeks
- jump diffusion models
- an SPDE application
- research by others
- future plans

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 $S(T)$, the value of the underlying at maturity, 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

	Euler		Milstein	
option	numerics	analysis	numerics	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 discretisation:

- 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 discretisation for scalar SDEs:

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

Analytic results for the distribution of the min/max/average over each timestep are used to construct multilevel estimator for Asian, lookback and barrier options

Digital options use Brownian extrapolation from $\widehat{S}_{N-1}$ and take conditional expectation to effectively smooth the payoff

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,

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

but can prove that

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

Barrier Options

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 Options

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

Basket Options

The Euler discretisation, multilevel implementation and numerical analysis all extend naturally to multi-dimensional SDEs, but variance convergence is poor for exotic options

In some cases (e.g. multiple assets with uncoupled scalar SDEs) can still use the Milstein discretisation.

The multilevel construction and numerical analysis extend too for basket option based on an average of the underlying assets

Key point: weighted average of Brownian interpolations is another Brownian interpolation, so can use the same multilevel construction as before

Multivariate digitals

What if the payoff is more complicated (not based on a simple average) but depends only on the values at a discrete set of times?

For a Lipschitz payoff there's no problem

If the payoff is discontinuous, may not have an analytic value for the conditional expectation based on Brownian extrapolation (and interpolation for earlier times)

Multivariate digitals

First solution: use a change of measure (as in importance sampling)

If $\mathbb{P}_c$ and $\mathbb{P}_f$ correspond to the conditional terminal distributions for the coarse and fine paths, and $\mathbb{Q}$ is an equivalent Gaussian distribution (with a larger variance),

$$\begin{aligned}\widehat{P}_l - \widehat{P}_{l-1} &= \mathbb{E}_{\mathbb{P}_f}[f] - \mathbb{E}_{\mathbb{P}_c}[f] \\ &= \mathbb{E}_{\mathbb{Q}}[(r_f - r_c)\, f] \\ &= \mathbb{E}_{\mathbb{Q}}[(r_f - r_c)(f - f_0)]\end{aligned}$$

where r_f, r_c are the Radon-Nikodym derivatives, and f_0 is any fixed constant (e.g. at peak of $\mathbb{Q}$).

Good asymptotic behaviour, but not wonderful in practice

Multivariate digitals

Second solution: use splitting

If W and Z are independent random variables, then for any function $g(W,Z)$ the estimator

$$\widehat{Y}_{M,N} = N^{-1} \sum_{n=1}^{N} \left(M^{-1} \sum_{m=1}^{M} g(W^{(n)}, Z^{(m,n)}) \right)$$

with independent samples $W^{(n)}$ and $Z^{(m,n)}$ is an unbiased estimator for $\mathbb{E}_{W,Z}\left[g(W,Z)\right] \equiv \mathbb{E}_W\left[\mathbb{E}_Z[g(W,Z)\,|\,W]\right]$, and its variance is

$$N^{-1}\, \mathbb{V}_W\left[\mathbb{E}_Z[g(W,Z)\,|\,W]\right] + (MN)^{-1}\, \mathbb{E}_W\left[\mathbb{V}_Z[g(W,Z)\,|\,W]\right].$$

Multivariate digitals

Here W is the driving Brownian path up to $T-h$ and Z is the increment for the final timestep.
Can argue that

$$
\begin{aligned}
\mathbb{V}_W\left[\mathbb{E}_Z[g(W,Z)\,|\,W]\right] &= O(h^{3/2}) \\
\mathbb{E}_W\left[\mathbb{V}_Z[g(W,Z)\,|\,W]\right] &= O(h)
\end{aligned}
$$

where $g(W,Z) \equiv \widehat{P}_l - \widehat{P}_{l-1}$. Hence, provided

$$h^{-1/2} \ll M \ll h^{-1}$$

get same asymptotic variance as analytic expectation, and at same asymptotic cost.

In limited testing, works better than change of measure

Greeks

My preference is to compute Greeks using IPA / pathwise sensitivities (L'Ecuyer, Broadie & Glasserman) because of efficient adjoint implementation (Giles & Glasserman)

What's the problem with multilevel implementation?

- for Lipschitz payoffs, lose one order of smoothness, so for first order Greeks the payoff sensitivity looks like a digital option so use same tricks
  - compute sensitivity of conditional expectation one timestep before maturity
  - use a change of measure
  - use splitting
- for digital payoffs, can't use pathwise sensitivity analysis

Greeks

For both Lipschitz and digital options can use “vibrato” sensitivity analysis:

- can be viewed as a hybrid combination of pathwise sensitivity analysis up to $T-h$, and then Likelihood Ratio Method for the final timestep
- can also be viewed as applying the change of measure idea to a perturbed path, in the limit of infinitesimal perturbation
- either way it effectively smooths the payoff, but the variance still suffers

Greeks

Asymptotic variance for the multilevel correction using the Milstein discretisation and vibrato sensitivity analysis

	first order Greeks	second order Greeks
Lipschitz	$O(h^{3/2})$	$O(h^{1/2})$
digital	$O(h^{1/2})$	$O(h^{-1/2})$

Asymptotic computational cost to achieve $O(\varepsilon)$ RMS accuracy, assuming first order weak convergence

	first order Greeks	second order Greeks
Lipschitz	$O(\varepsilon^{-2})$	$O(\varepsilon^{-5/2})$
digital	$O(\varepsilon^{-5/2})$	$O(\varepsilon^{-7/2})$

Jump Diffusion

For finite activity jump diffusion models like Merton's, the multilevel treatment is relatively straightforward if the Poisson jump rate is constant

- use jump-adapted discretisation, adding jump times to standard uniform timestep discretisation times
- Milstein approximation of pure diffusion model between jumps, with Brownian interpolation within each timestep
- jump intervals are exponential random variables; the same values are used for coarse and fine paths
- again get $\|\widehat{S}^f - \widehat{S}_c\| \approx O(h)$

Jump Diffusion

Trickier when the jump rate is path-dependent

- can lead to coarse and fine paths jumping at slightly different times, producing big differences in payoff if one jumps before maturity and the other after
- similar to problems with digital option
- also causes difficulty for computing pathwise sensitivities

This last point holds generally – discontinuous behaviour is bad for both pathwise sensitivities (invalid) and the multilevel method (poor convergence) and the same “fix” usually works for both

Jump Diffusion

One approach to path-dependent jump rates uses a constant rate to generate candidate jump times, then uses "thinning" to select a subset (Glasserman & Merener)

- if the same uniform random variables are used for thinning coarse and fine paths then in most cases the paths will use the same jump times, but in a few they won't – same problem
- can use a change of measure to map to a thinning process with 50/50 acceptance/rejection – then using the same r.v. means the coarse and fine paths always jump at the same time
- initial numerical results look good

Jump Diffusion

Another approach defines the jump interval by a discrete approximation to

$$\int_0^\tau \lambda(S,t)\,\mathrm{d}t = -\log U$$

where U is a unit interval uniform r.v.

The coarse path can be required to jump at the same time as the fine path through a change of measure with Radon-Nikodym derivative

$$\exp\left(\int_0^T (\lambda^f - \lambda^c)\ \mathrm{d}t \right)\ \prod_{n=1}^{N_T} \frac{\lambda_n^c}{\lambda_n^f}$$

Not tested yet – currently being implemented

SPDE Application

Currently working with Christoph Reisinger on an SPDE application which arises in CDO modelling (Bush, Hambly, Haworth & Reisinger)

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

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

- derived in limit as number of firms $\longrightarrow \infty$
- x is distance to default
- $p(x,t)$ is probability density function
- $\mathrm{d}W$ 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

Other Research

- Rainer Avikainen – numerical analysis of multilevel method with Euler discretisation
- Raul Tempone, Anders Szepessy – multilevel method combined with adaptive time-stepping
  - adaptive time-stepping can be very effective in some circumstances
  - might be great for difficulties with Heston model
- Steffen Dereich, F. Heidenreich – multilevel method for Lévy processes
  - large jumps simulated individually
  - small jumps approximate by Brownian diffusion
  - small/large distinction dependent on timestep h, so changes between coarse and fine paths

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 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

American options – the next big challenge

- instead of Longstaff-Schwartz approach, view it as a global exercise boundary optimisation problem?
- parametric representation of exercise boundary
- use multilevel method (combined with adjoints?) to compute parametric sensitivities
- feed into classical gradient based optimisation
- start by optimising coarse approximations to give starting point for finer approximations (FMG in multigrid literature, multilayer optimisation – Sachs & Käbe)

Conclusions

- multilevel method being adapted to increasingly more challenging applications
- numerical analysis now supports some of the experimental findings
- many of the challenges are closely related to those faced when computing pathwise sensitivities
- haven't discussed use of QMC, but this helps greatly as most computational effort is expended on coarse levels

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