Multilevel Monte Carlo methods for finance

Mike Giles

Mathematical Institute, University of Oxford
Oxford-Man Institute of Quantitative Finance

HPCFinance Final Conference

March 14, 2016

Outline and objectives

- key ideas
- application to basket options
- extensions to Greeks and Lévy processes
- future application to VaR

I hope to emphasise:

- the simplicity of the idea – easy to add to existing codes
- scope for improved performance through being creative
- lots of people working on a variety of applications

Generic Problem

Suppose we have an option with payoff P on multiple underlying assets, each of which satisfies an SDE with general drift and volatility terms:

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

Will simulate these using the Milstein scheme:

$$\widehat{S}_{n+1} = \widehat{S}_n + a\,h + b\,\Delta W_n + \tfrac{1}{2}\,b\,b' \left((\Delta W_n)^2 - h \right)$$

which gives first order weak and strong convergence:

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

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

Standard MC Approach

Mean Square Error is $O(N^{-1} + h^2)$

- first term comes from variance of estimator
- second term comes from bias due to weak convergence

To make this $O(\varepsilon^2)$ requires

$$N = O(\varepsilon^{-2}), \quad h = O(\varepsilon) \quad \Longrightarrow \quad \text{cost} = O(N\,h^{-1}) = O(\varepsilon^{-3})$$

Aim is to improve this to $O(\varepsilon^{-2})$, by combining simulations with different numbers of timesteps

Control variate

Classic approach to MC variance reduction: approximate $\mathbb{E}[f]$ using

$$N^{-1} \sum_{n=1}^{N} \left\{ f(\omega^{(n)}) - \lambda \left(g(\omega^{(n)}) - \mathbb{E}[g] \right) \right\}$$

where

- control variate g has known expectation $\mathbb{E}[g]$
- g is well correlated with f, and optimal value for λ can be estimated by a few samples

For the optimal value of λ, the variance is reduced by factor $(1-\rho^2)$, where ρ is the correlation between f and g.

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[f_1]$ but it is much cheaper to simulate $f_0 \approx f_1$, then since

$$\mathbb{E}[f_1] = \mathbb{E}[f_0] + \mathbb{E}[f_1 - f_0]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} f_0^{(0,n)} \; + \; N_1^{-1} \sum_{n=1}^{N_1} \left(f_1^{(1,n)} - f_0^{(1,n)} \right)$$

Two differences from standard control variate method:

- $\mathbb{E}[f_0]$ is not known, so has to be estimated
- $\lambda = 1$

Benefit: if $f_1 - f_0$ is small, won't need many samples to accurately estimate $\mathbb{E}[f_1 - f_0]$, so cost will be reduced greatly.

Multilevel Monte Carlo

Natural generalisation: given a sequence $f_0, f_1, \ldots, f_L$

$$\mathbb{E}[f_L] = \mathbb{E}[f_0] + \sum_{\ell=1}^{L} \mathbb{E}[f_\ell - f_{\ell-1}]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} f_0^{(0,n)} \; + \; \sum_{\ell=1}^{L} \left\{ N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(f_\ell^{(\ell,n)} - f_{\ell-1}^{(\ell,n)} \right) \right\}$$

with independent estimation for each level

Multilevel Monte Carlo

If we define

- C_0, V_0 to be cost and variance of f_0
- C_ℓ, V_ℓ to be cost and variance of $f_\ell - f_{\ell-1}$

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:

- 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

Motivated by computational finance applications, in 2006 I introduced MLMC for SDEs (stochastic differential equations).

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

Level ℓ corresponds to approximation using 2^ℓ timesteps, giving approximate payoff $\widehat{P}_\ell$.

Choice of finest level L depends on weak error (bias).

Multilevel decomposition gives

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

Multilevel Path Simulation

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

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

To make RMS error less than ε

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

Multilevel Path Simulation

For the Milstein discretisation and a European option with a Lipschitz payoff function

$$\mathbb{E}[\sup_t(\widehat{S}_\ell - S)^2] = O(h_\ell^2) \quad \Longrightarrow \quad \mathbb{E}[(\widehat{P}_\ell - P)^2] = O(h_\ell^2)$$

$$\Longrightarrow \quad \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] = O(h_\ell^2)$$

and the optimal N_ℓ is asymptotically proportional to $h_\ell^{3/2}$.

To make the combined variance $O(\varepsilon^2)$ requires

$$N_\ell = O(\varepsilon^{-2} h_\ell^{3/2})$$

and hence we obtain an $O(\varepsilon^2)$ MSE for an $O(\varepsilon^{-2})$ computational cost.

MLMC Theorem

(Slight generalisation of original version)

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

MLMC Theorem allows a lot of freedom in constructing the multilevel estimator. I sometimes use different approximations on the coarse and fine levels:

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

The telescoping sum still works provided

$$\mathbb{E}\left[\widehat{P}_\ell^f\right] \ = \ \mathbb{E}\left[\widehat{P}_\ell^c\right].$$

Given this constraint, can be creative to reduce the variance

$$\mathbb{V}\left[\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c\right].$$

Basket options

Basket of 5 underlying assets, modelled by Geometric Brownian Motion

$$\mathrm{d}S_i = r\, S_i\, \mathrm{d}t + \sigma_i\, S_i\, \mathrm{d}W_i$$

with correlation between 5 driving Brownian motions

Three different payoffs on arithmetic average of assets:

- standard call:

$$P = \exp(-rT)\ \max(S(T)-K, 0)$$

- lookback:

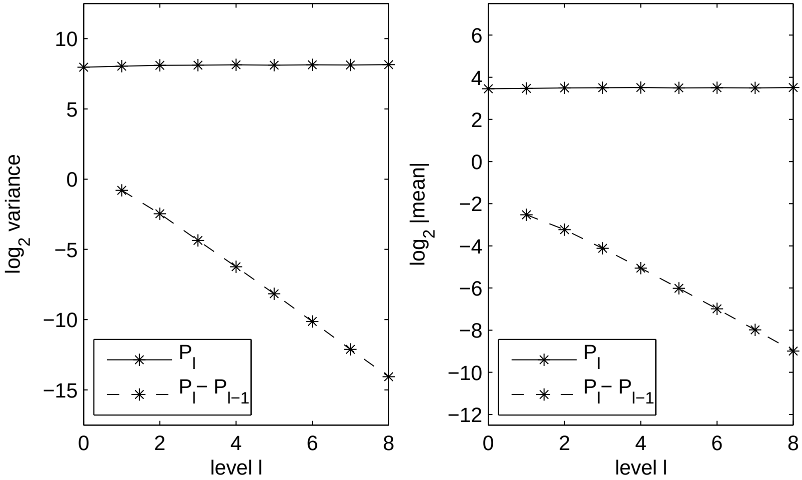
$$P = \exp(-rT)\ (S(T)-\min_t S_t)$$

- digital call:

$$P = \exp(-rT)\ \mathbf{1}_{S(T)>K}$$

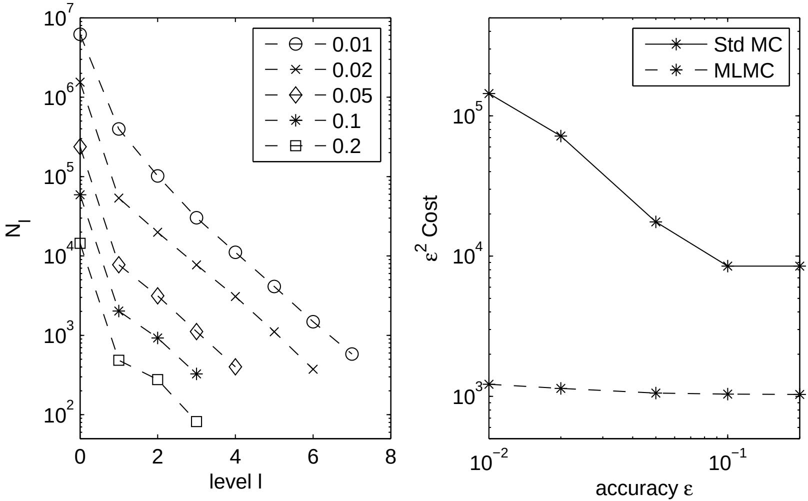
Basket options

Standard call option:



Basket options

Standard call option:



Lookback options

Payoff depends on the minimum attained by the path $S(t)$.

If the numerical approximation uses the minimum of the values at the discrete simulation times

$$\widehat{S}_{min} \equiv \min_j \widehat{S}_j$$

then we have two problems:

- $O(\sqrt{h})$ weak convergence
- $\widehat{S}_{\ell,min} - \widehat{S}_{\ell-1,min} = O(\sqrt{h_\ell})$ which leads to $V_\ell = O(h_\ell)$

Lookback options

To fix this, define a Brownian Bridge interpolation conditional on the endpoints for each timestep, with constant drift and volatility.

For the fine path, standard result for the sampling from the distribution of the minimum of a Brownian Bridge gives

$$\widehat{S}_{min} = \min_j \ \tfrac{1}{2}\left(\widehat{S}_j + \widehat{S}_{j-1} - \sqrt{(\widehat{S}_j - \widehat{S}_{j-1})^2 - 2\,h\,b_j^2 \log U_j}\ \right)$$

where the U_j are independent $U(0,1)$ random variables.

This gives $O(h)$ weak convergence, but if we do something similar for the coarse path with a different set of $U's$ the variance will still be poor.

Lookback options

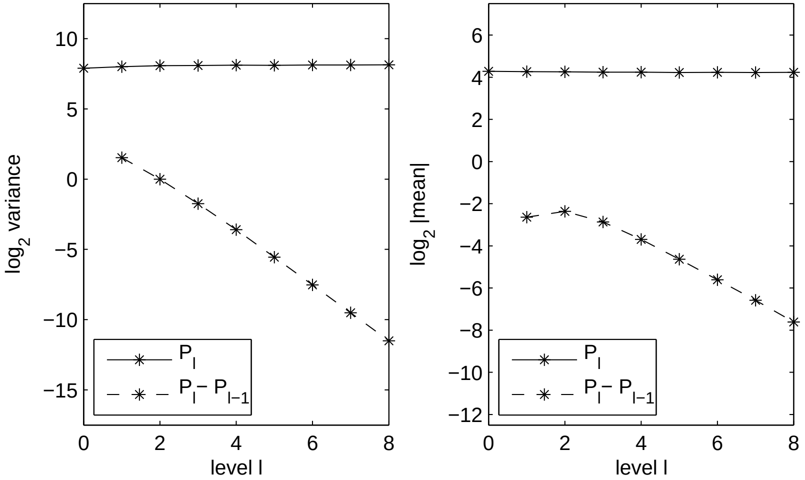
Instead, do the following:

- sample from the mid-point of the Brownian Bridge interpolation for the coarse timestep, using the Brownian path information from the fine path – this mid-point value is within $O(h_\ell)$ of the fine path simulation
- sample from the minima of each half of the coarse timestep using the same $U's$ as fine path
- take the minimum of the two minima, and then the minimum over all coarse timesteps.

This leads to an $O(h_\ell)$ difference in the computed minima for the coarse and fine paths, and is valid because the distribution for the coarse path minimum has not been altered.

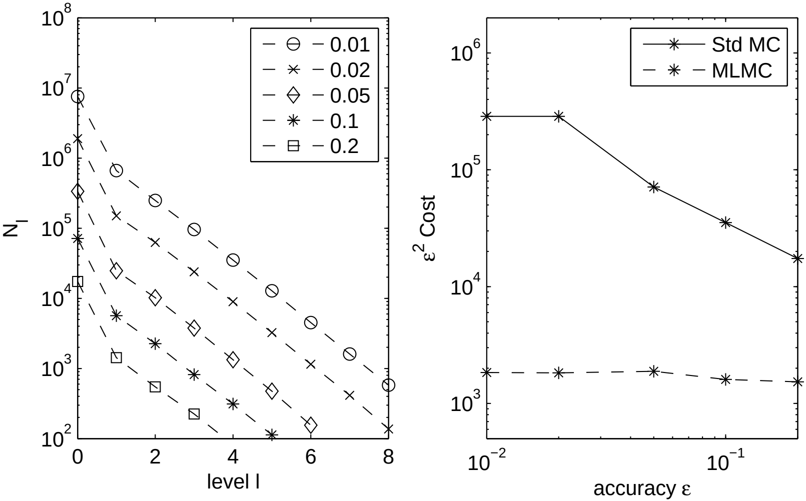
Basket options

Lookback option:



Basket options

Lookback option:



Digital options

In a digital option, the payoff is a discontinuous function of the final state.

Using the Milstein approximation, first order strong convergence means that $O(h_\ell)$ of the simulations have coarse and fine paths on opposite sides of a discontinuity.

Hence,

$$\widehat{P}_\ell - \widehat{P}_{\ell-1} = \begin{cases} O(1), & \text{with probability } O(h_\ell) \\ O(h_\ell), & \text{with probability } O(1) \end{cases}$$

so

$$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] = O(h_\ell), \quad \mathbb{E}[(\widehat{P}_\ell - \widehat{P}_{\ell-1})^2] = O(h_\ell),$$

and hence $V_\ell = O(h_\ell)$, not $O(h_\ell^2)$

Digital options

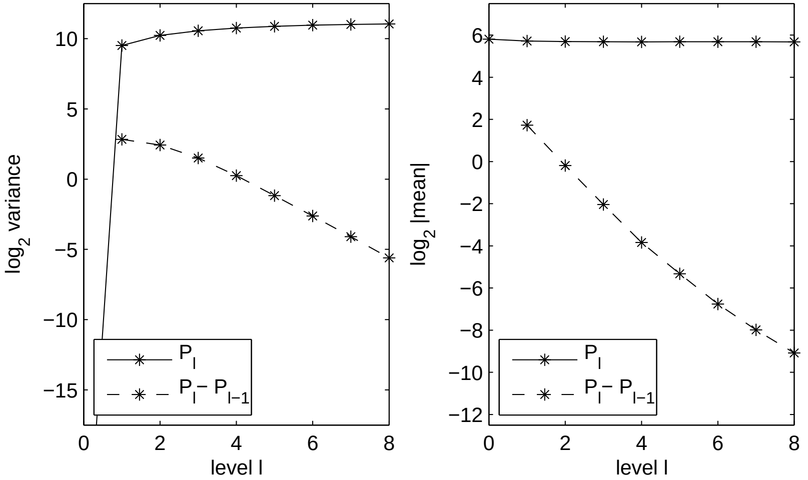
Three fixes:

- Conditional expectation: using the Euler discretisation instead of Milstein for the final timestep, conditional on all but the final Brownian increment, the final state has a Gaussian distribution, with a known analytic conditional expectation in simple cases
- Splitting: split each path simulation into M paths by trying M different values for the Brownian increment for the last fine path timestep
- Change of measure: when the expectation is not known, can use a change of measure so the coarse path takes the same final state as the fine path — difference in the "payoff" now comes from the Radon-Nikodym derivative

These all effectively smooth the payoff – end up with $V_\ell = O(h_\ell^{3/2})$.

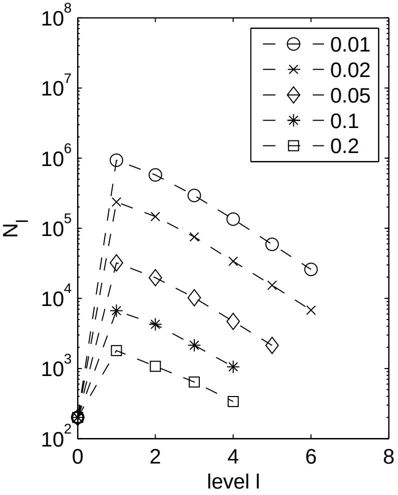
Basket options

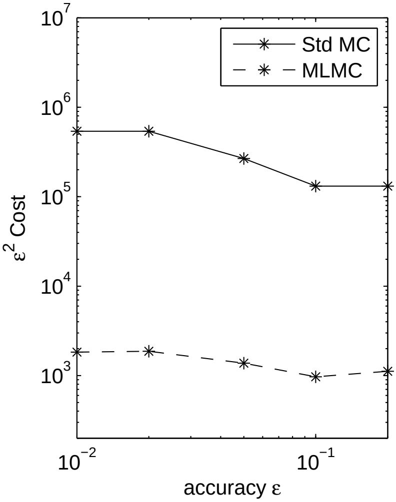
Digital call option:



Basket options

Digital call option:





Numerical 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)

Euler analysis due to G, Higham & Mao (2009) and Avikainen (2009).
Milstein analysis due to G, Debrabant & Rößler (2012).

Greeks and jump diffusion

Greeks (Burgos, 2011)

- MLMC combines well with pathwise sensitivity analysis for Greeks
- main concern is reduced regularity of "payoff"
- techniques are similar to handling digital options

Finite activity rate Merton-style jump diffusion (Xia, 2011)

- if constant rate, no problem — use jump-adapted discretisation and coarse and fine paths jump at the same time
- if path-dependent rate, then it's trickier
  - use jump-adapted discretisation plus thinning (Glasserman & Merener)
  - could lead to fine and coarse paths jumping at different times $\Longrightarrow$ poor variance
  - instead use a change of measure to force jumps to be at the same time

Lévy processes

Infinite activity rate, general Lévy processes
(Dereich 2010; Marxen 2010; Dereich & Heidenreich 2011)

- on level ℓ, simulate jumps bigger than δ_ℓ ($\delta_\ell \to 0$ as $\ell \to \infty$)
- either neglect smaller jumps or use a Gaussian approximation
- multilevel problem: discrepancy in treatment of jumps which are bigger than δ_ℓ but smaller than $\delta_{\ell-1}$

Exact simulation (Xia, 2014)

- with some popular exponential-Lévy models (variance-gamma, NIG) possible to directly simulate Lévy increments over fine timesteps
- sum them pairwise to get corresponding increments for coarse path
- very helpful for path-dependent options (Asian, lookback, barrier)

New application: VaR

Value-at-risk calculation seems a great candidate for an MLMC treatment.

VaR:

- outer simulation of multiple risk factors Z over a risk horizon $[0, H]$
- evaluation of loss in portfolio value at H compared to present time
- various measures of risk:
  - $\mathrm{VaR}_\alpha = \inf\{x : \mathbb{P}[L\!>\!x] < \alpha\}$
  - $\mathrm{CVaR}_\alpha = \alpha^{-1}\, \mathbb{E}\,[\, L\ \mathbf{1}(L\!>\!\mathrm{VaR}_\alpha)] = \mathbb{E}\,[\, L \,|\, L\!>\!\mathrm{VaR}_\alpha]$
  - other risk measures based on distribution of L

New application: VaR

The portfolio usually contains many options:

- many are simple vanilla options with values, conditional on Z, given by closed-form Black-Scholes formulas

- some are exotic options with values given by nested simulation, conditional on Z.

  i.e. for given Z need to
  - simulate multiple Brownian paths W
  - compute underlying assets S
  - average the payoff to approximate risk-neutral conditional expectation

New application: VaR

In applying MLMC ideas, there are several ways in which we can get less accurate simulations at greatly reduced cost:

- approximate option values using quadratic delta-gamma approximation
- sub-sample portfolio (i.e. pick a random sub-sample of the options in the portfolio instead of evaluating all options)
- vary number of Brownian paths used for conditional expectation
- vary number of timesteps used for SDE simulation

New application: VaR

Blatant sales pitch!

- Starting new project with Sascha Desmettre, Ralf Korn and Klaus Ritter at TU Kaiserslautern
- Very keen to engage with finance industry – looking for banks, pension/insurance companies who can help to define the challenges
- Wouldn't say no to some research funding too!

Conclusions

- multilevel idea is very simple
- challenge can be how to apply it in new situations
- discontinuous payoffs cause some difficulties, but there is a lot of experience now in coping with this
- there are also “tricks” which can be used in situations with poor strong convergence
- being used for an increasingly wide range of applications; biggest computational savings when coarsest (helpful) approximation is much cheaper than finest
- in computational finance, VaR may prove to be the application with the greatest MLMC benefits

References

Webpage for my research/papers:

`people.maths.ox.ac.uk/gilesm/mlmc.html`

Webpage for new 70-page *Acta Numerica* review and MATLAB test codes:

`people.maths.ox.ac.uk/gilesm/acta/`

– contains references to almost all MLMC research

MLMC Community

Webpage: `people.maths.ox.ac.uk/gilesm/mlmc_community.html`

Abo Academi (Avikainen) – numerical analysis
Basel (Harbrecht) – elliptic SPDEs, sparse grids
Bath (Kyprianou, Scheichl, Shardlow, Yates) – elliptic SPDEs, MCMC, Lévy-driven SDEs, stochastic chemical modelling
Chalmers (Lang) – SPDEs
Duisburg (Belomestny) – Bermudan and American options
Edinburgh (Davie, Szpruch) – SDEs, numerical analysis
EPFL (Abdulle) – stiff SDEs and SPDEs
ETH Zürich (Jenny, Jentzen, Schwab) – SPDEs, multilevel QMC
Frankfurt (Gerstner, Kloeden) – numerical analysis, fractional Brownian motion
Fraunhofer ITWM (Iliev) – SPDEs in engineering
Hong Kong (Chen) – Brownian meanders, nested simulation in finance
IIT Chicago (Hickernell) – SDEs, infinite-dimensional integration, complexity analysis
Kaiserslautern (Heinrich, Korn, Ritter) – finance, SDEs, parametric integration, complexity analysis
KAUST (Tempone, von Schwerin) – adaptive time-stepping, stochastic chemical modelling
Kiel (Gnewuch) – randomized multilevel QMC
LPMA (Frikha, Lemaire, Pagès) – numerical analysis, multilevel extrapolation, finance applications
Mannheim (Neuenkirch) – numerical analysis, fractional Brownian motion
MIT (Peraire) – uncertainty quantification, SPDEs
Munich (Hutzenthaler) – numerical analysis
Oxford (Baker, Giles, Hambly, Reisinger) – SDEs, SPDEs, numerical analysis, finance applications, stochastic chemical modelling
Passau (Müller-Gronbach) – infinite-dimensional integration, complexity analysis
Stanford (Glynn) – numerical analysis, randomized multilevel
Strathclyde (Higham, Mao) – numerical analysis, exit times, stochastic chemical modelling
Stuttgart (Barth) – SPDEs
Texas A&M (Efendiev) – SPDEs in engineering
UCLA (Caflisch) – Coulomb collisions in physics
UNSW (Dick, Kuo, Sloan) – multilevel QMC
UTS (Baldeaux) – multilevel QMC
Warwick (Stuart, Teckentrup) – MCMC for SPDEs
WIAS (Friz, Schoenmakers) – rough paths, fractional Brownian motion, Bermudan options
Wisconsin (Anderson) – numerical analysis, stochastic chemical modelling