Variance Reduction Through Multilevel Monte Carlo Path Calculations

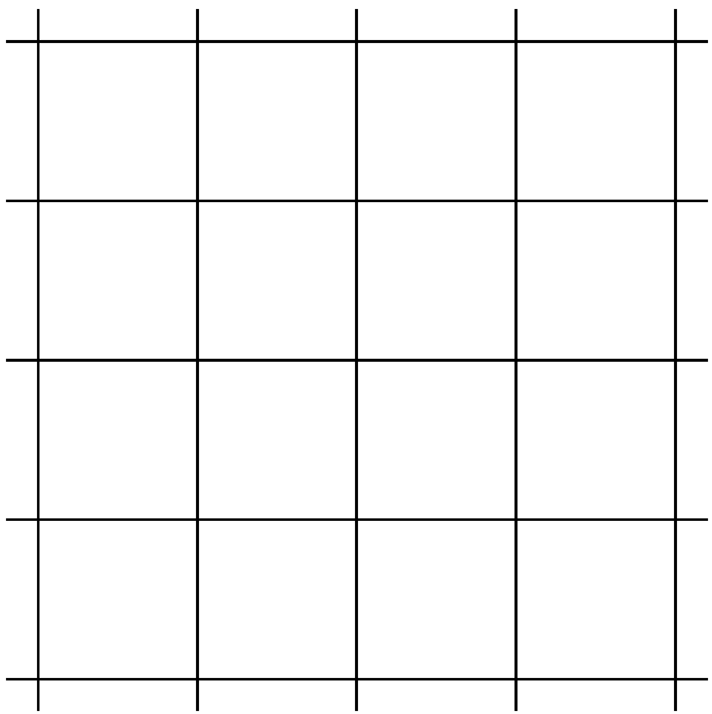
Mike Giles

giles@comlab.ox.ac.uk

Oxford University Computing Laboratory

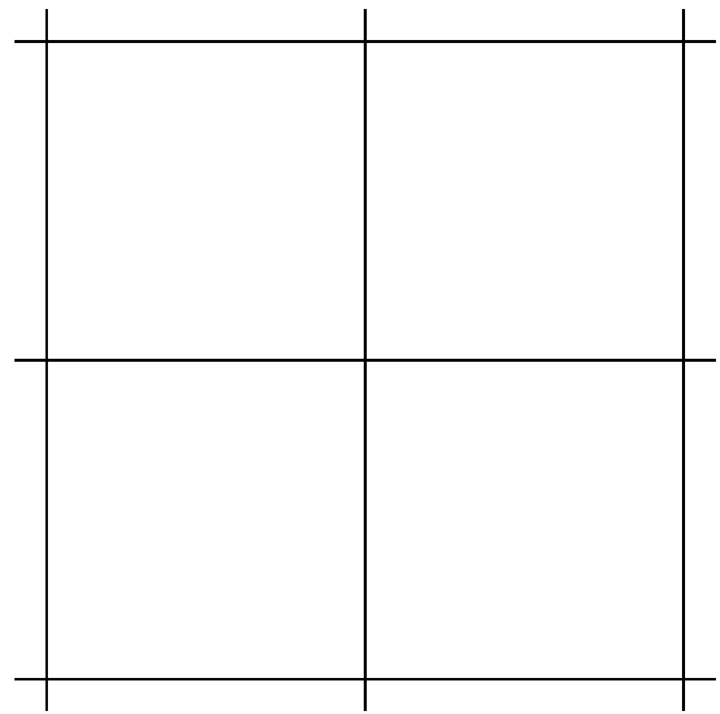
Multigrid

A powerful technique for solving PDE discretisations:



Fine grid

more accurate
more expensive



Coarse grid

less accurate
less expensive

Multigrid

Multigrid combines calculations on a nested sequence of grids to get the accuracy of the finest grid at a much lower computational cost.

We will use a similar idea to achieve variance reduction in Monte Carlo path calculations, combining simulations with different numbers of timesteps – same accuracy as finest calculations, but at a much lower computational cost.

Generic Problem

SDE with general drift and volatility terms:

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

Suppose we want to compute the expected value of an option dependent on the terminal state

$$P = f(S(T))$$

with a uniform Lipschitz bound,

$$|f(U) - f(V)| \ \leq c\, \|U - V\|\,, \quad \forall\, U, V.$$

Standard MC Approach

Euler discretisation with timestep h:

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

Simplest estimator for expected payoff is an average of N independent path simulations:

$$\widehat{Y} = N^{-1} \sum_{i=1}^{N} f(\widehat{S}^{(i)}_{T/h}).$$

- weak convergence – $O(h)$ error in expected payoff
- strong convergence – $O(h^{1/2})$ error in individual path

Standard MC Approach

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

- 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 cost to $O\left(\varepsilon^{-2}(\log \varepsilon)^2\right)$

(In 2005, Ahmed Kebaier published a two-level method which reduces the cost to $O\left(\varepsilon^{-2.5}\right)$, equivalent to a single application of Richardson extrapolation.)

Multilevel MC Approach

Consider multiple sets of simulations with different timesteps $h_l = 2^{-l}\,T,\ l = 0, 1, \ldots, L$, and payoff $\widehat{P}_l$

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

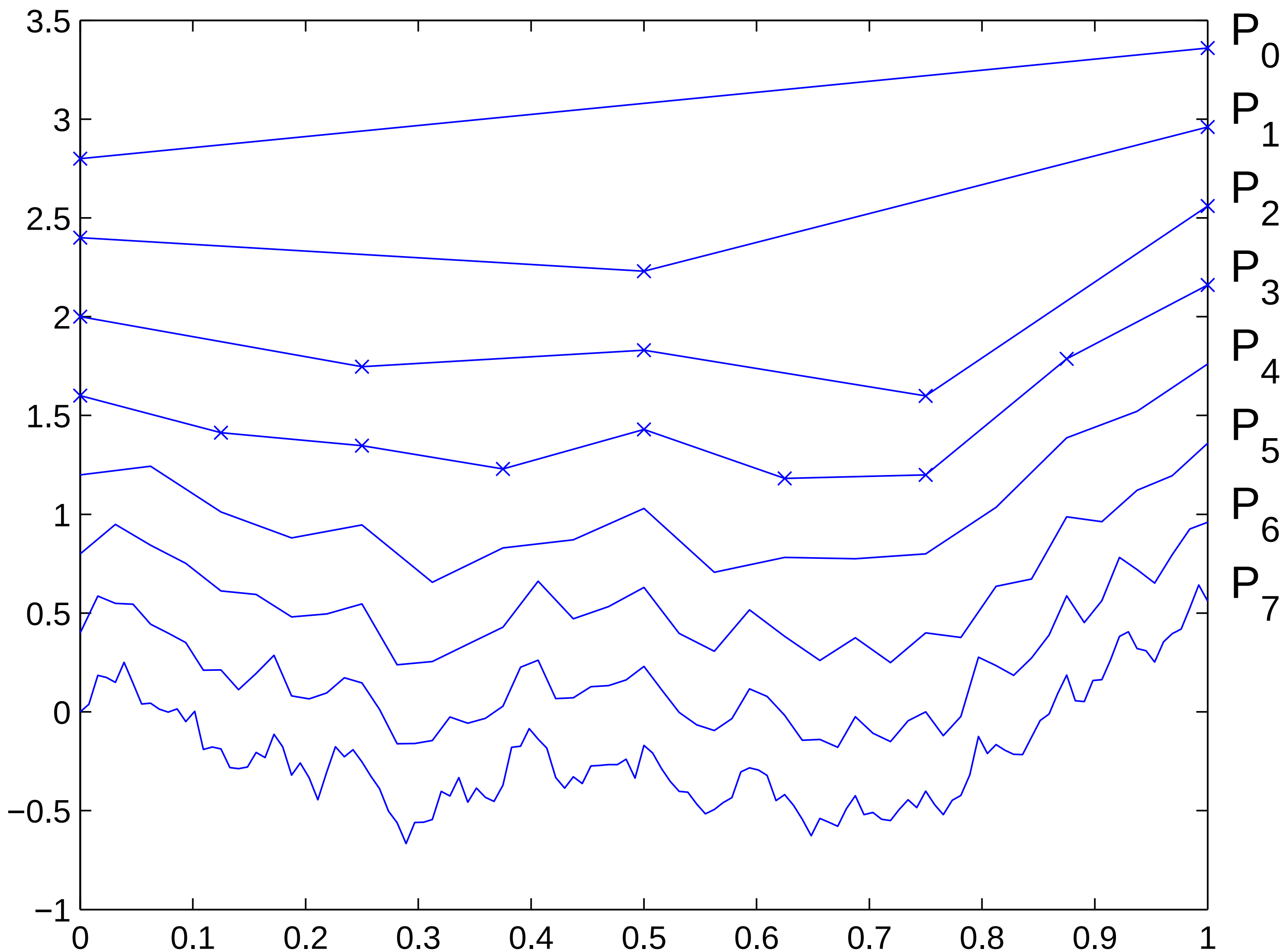
Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key point: approximate $E[\widehat{P}_l - \widehat{P}_{l-1}]$ using N_l simulations with $\widehat{P}_l$ and $\widehat{P}_{l-1}$ obtained using <u>same</u> Brownian path.

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

Multilevel MC Approach

Discrete Brownian path at different levels



Multilevel MC Approach

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

$$V\left[\sum_{l=0}^{L}\widehat{Y}_l\right] = \sum_{l=0}^{L} N_l^{-1} V_l, \qquad V_l \equiv V[\widehat{P}_l - \widehat{P}_{l-1}],$$

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 $\sqrt{V_l\, h_l}$.

Multilevel MC Approach

For the Euler discretisation and the Lipschitz payoff function

$$V[\widehat{P}_l - P] = O(h_l) \quad \Longrightarrow \quad V[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l)$$

and the optimal N_l is asymptotically proportional to h_l.

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

$$N_l = O(\varepsilon^{-2} L h_l).$$

To make the bias $O(\varepsilon)$ requires

$$L = \log_2 \varepsilon^{-1} + O(1) \quad \Longrightarrow \quad h_L = O(\varepsilon).$$

Hence, we obtain an $O(\varepsilon^2)$ MSE for a computational cost which is $O(\varepsilon^{-2} L^2) = O(\varepsilon^{-2} (\log \varepsilon)^2)$.

Multilevel MC Approach

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* $h_l = M^{-l}\,T$.

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

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

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

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

iv) C_l*, the computational complexity of* $\widehat{Y}_l$*, is bounded by*

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

Multilevel MC Approach

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 multi-level estimator*

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

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

with a computational complexity C *with bound*

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

Results

Geometric Brownian motion:

$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W, \qquad 0 < t < 1,$$

$S(0)\!=\!1$, $r\!=\!0.05$, $\sigma\!=\!0.2$

Heston model:

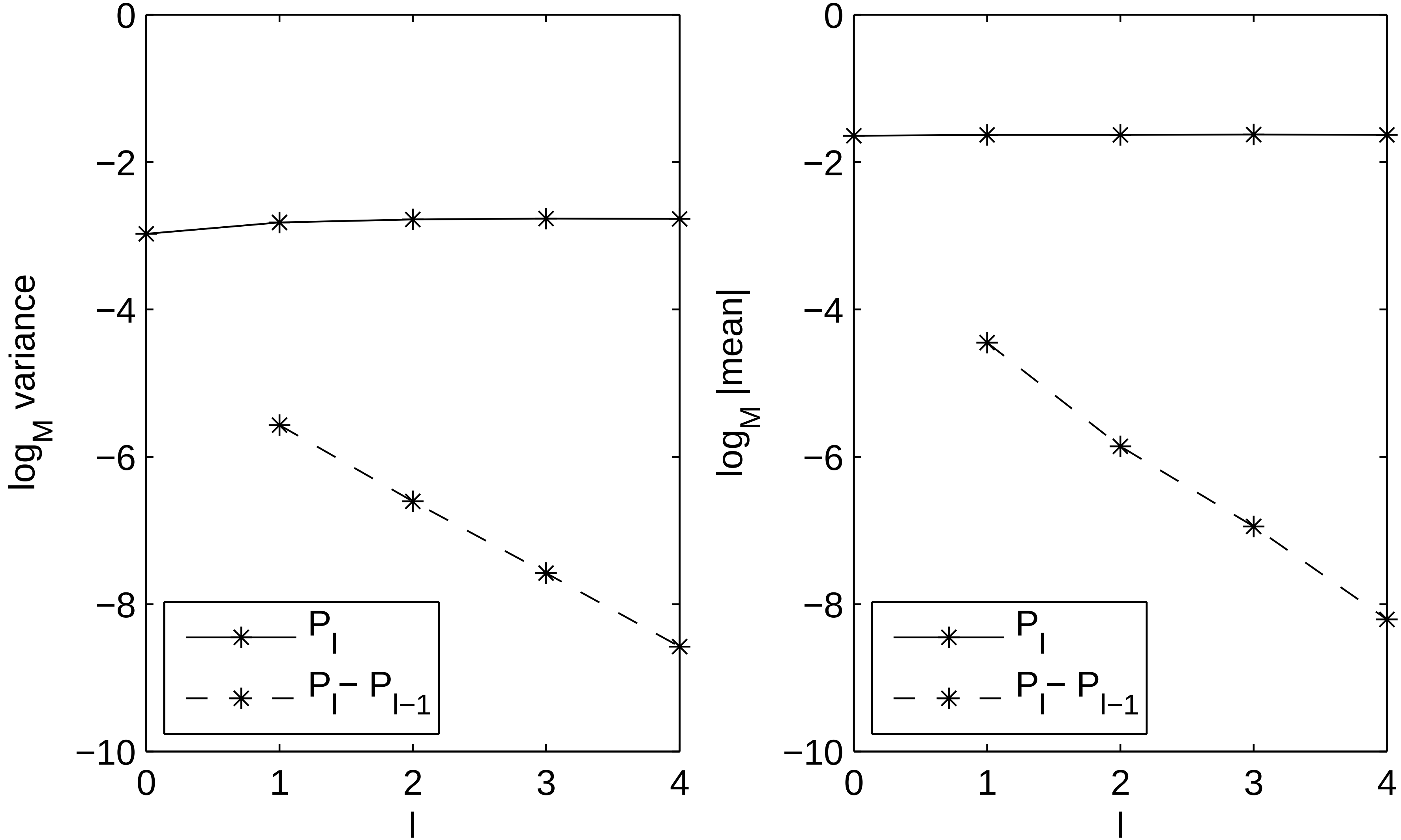
$$\begin{aligned} \mathrm{d}S &= r\,S\,\mathrm{d}t + \sqrt{V}\,S\,\mathrm{d}W_1, \qquad 0 < t < 1 \\ \mathrm{d}V &= \lambda\,(\sigma^2 - V)\,\mathrm{d}t + \xi\,\sqrt{V}\,\mathrm{d}W_2, \end{aligned}$$

$S(0)\!=\!1$, $V(0)\!=\!0.04$, $r\!=\!0.05$, $\sigma\!=\!0.2$, $\lambda\!=\!5$, $\xi\!=\!0.25$, $\rho\!=\!-0.5$

All calculations use $M\!=\!4$, more efficient than $M\!=\!2$.

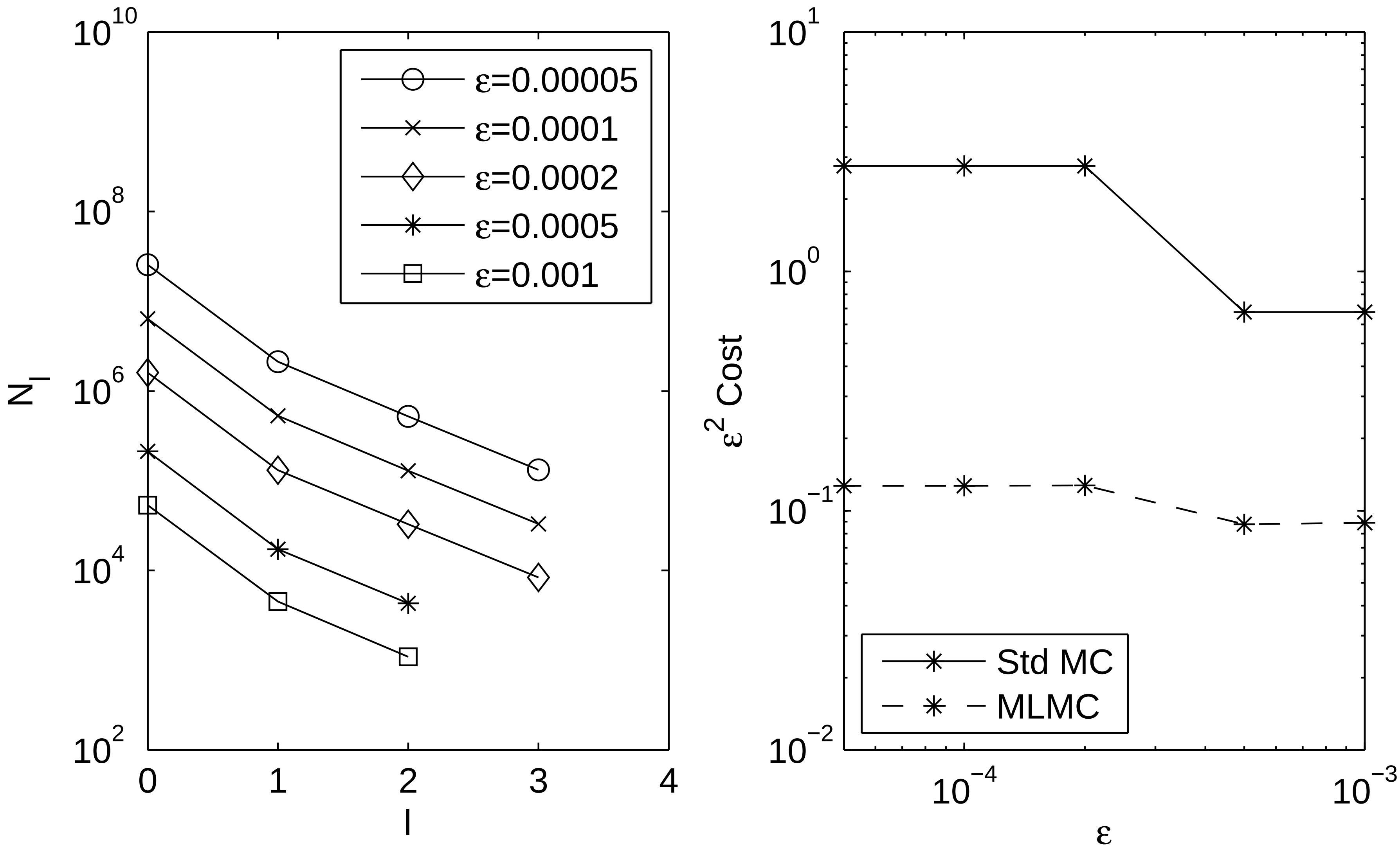
Results

GBM: European call, $\max(S(1)-1,0)$



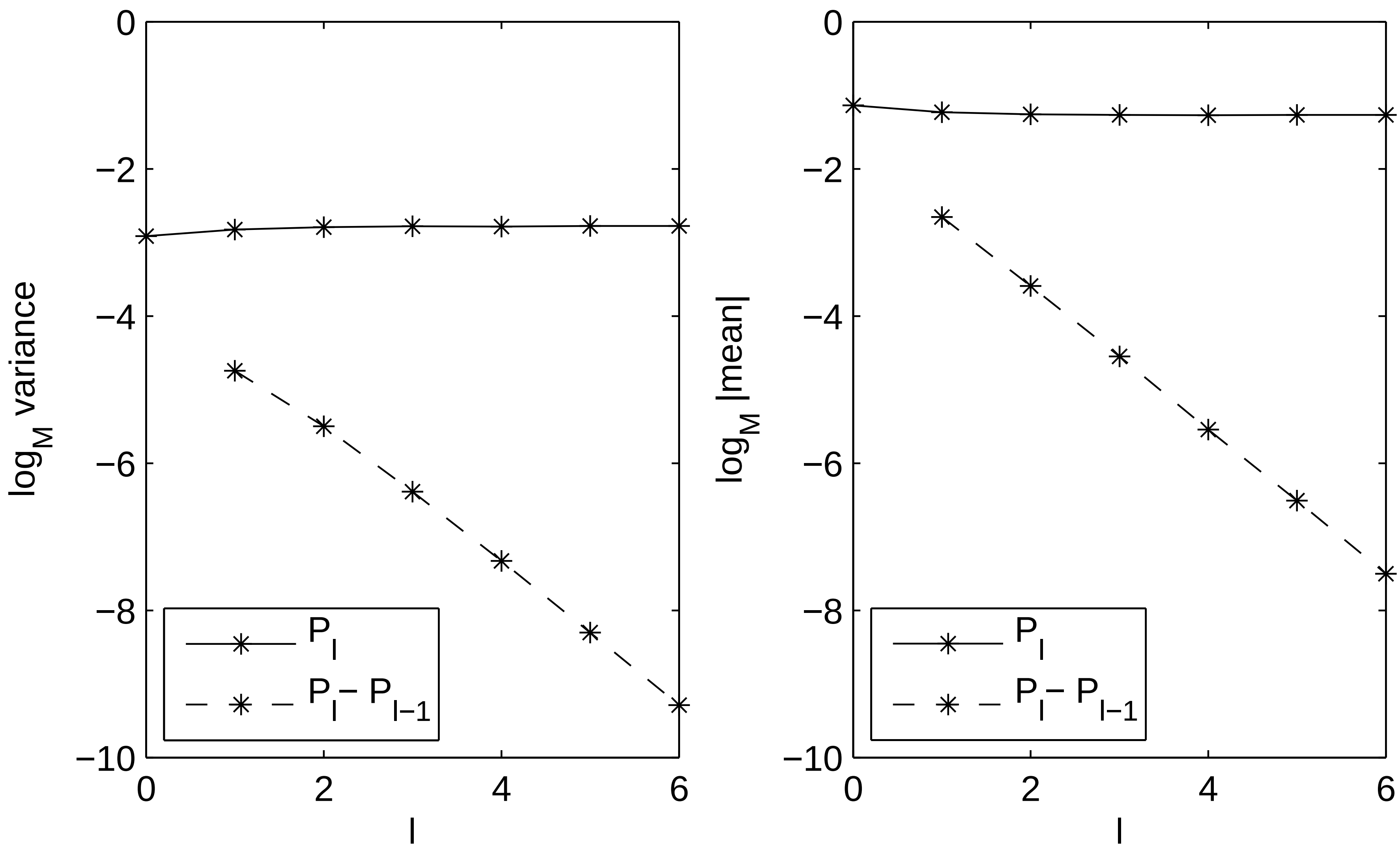
Results

GBM: European call, $\max(S(1)-1,0)$



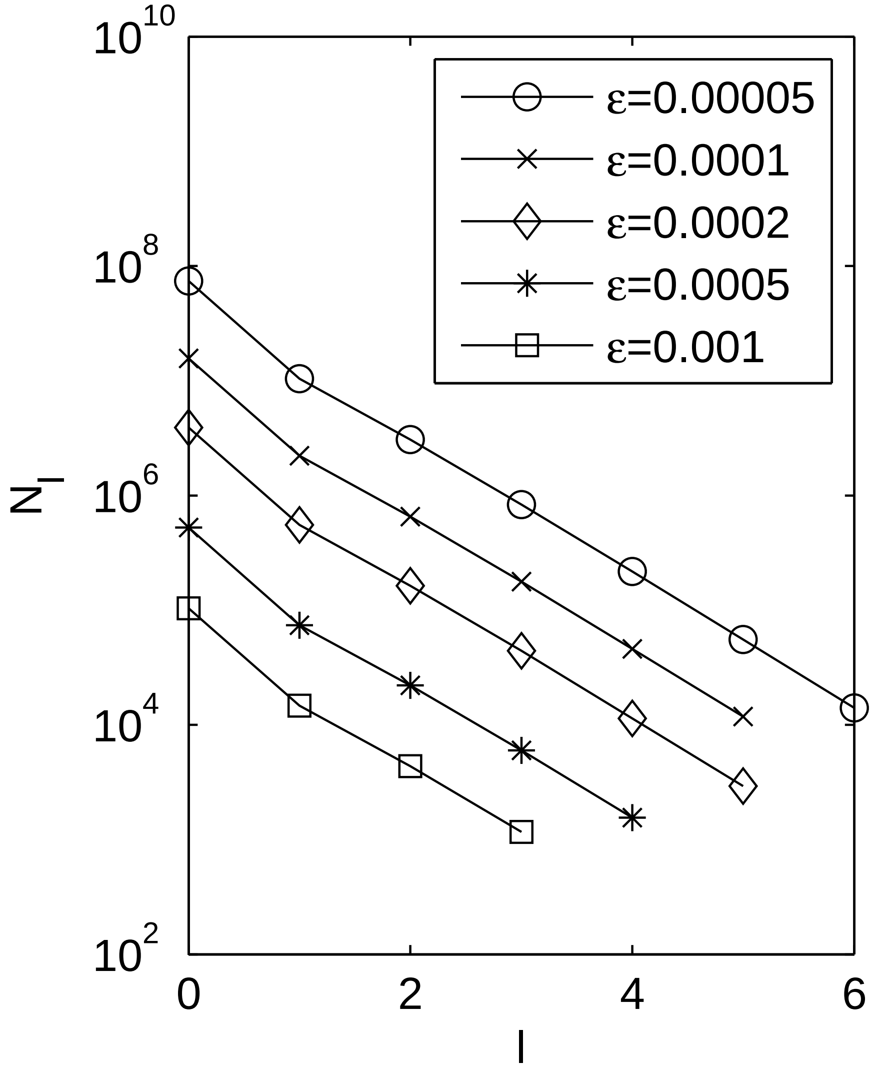
Results

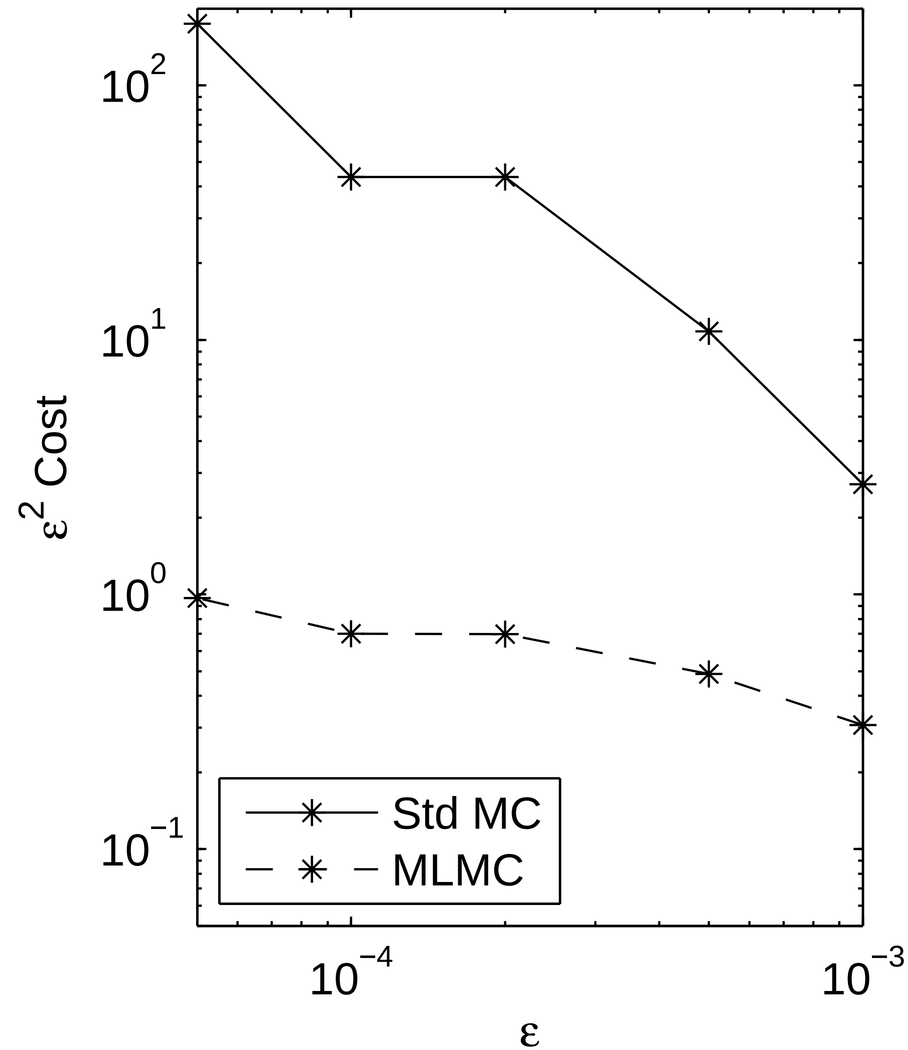
GBM: lookback option, $S(1) - \min_{0<t<1} S(t)$



Results

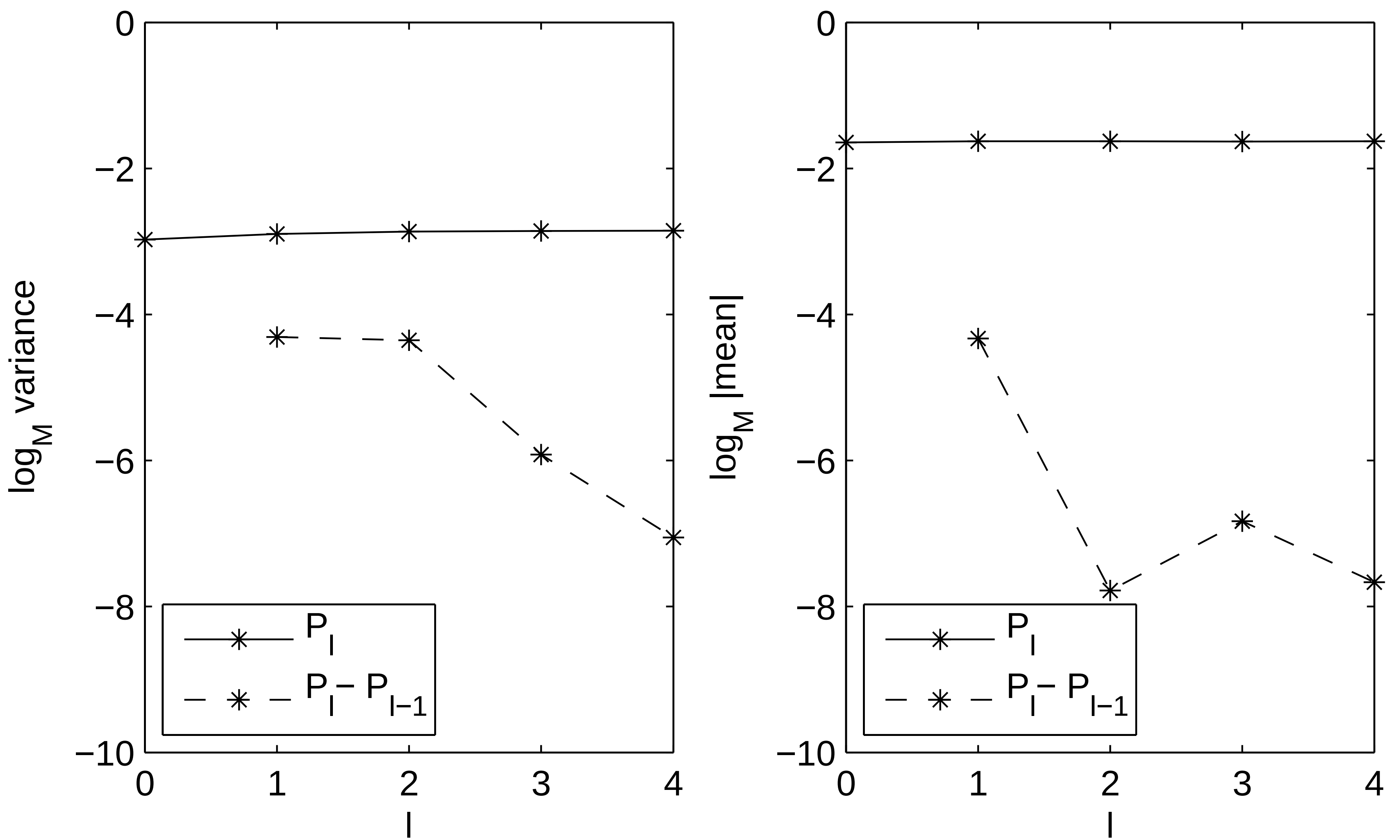
GBM: lookback option, $S(1) - \min_{0<t<1} S(t)$





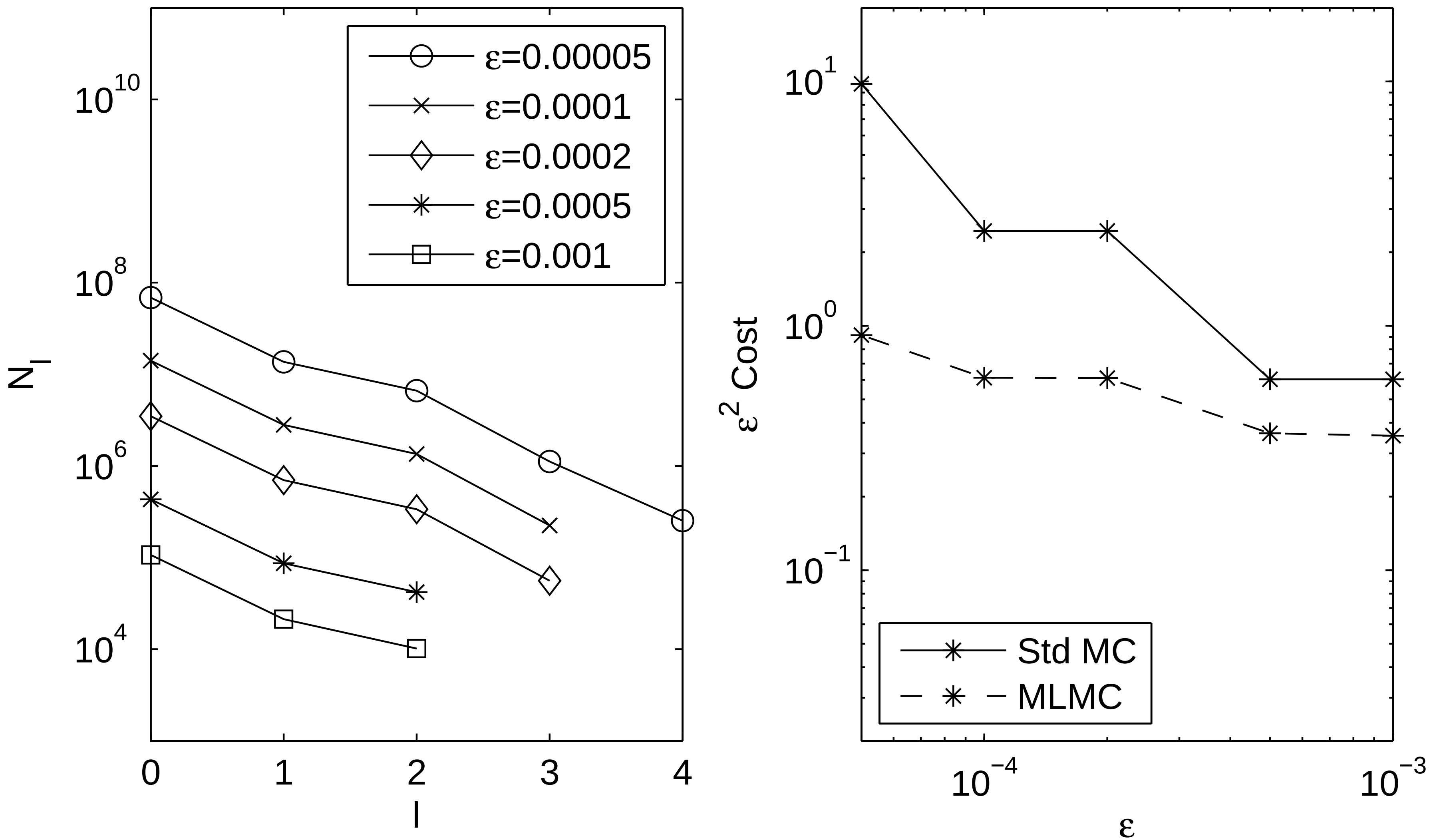
Results

Heston model: European call



Results

Heston model: European call



Comments

Results so far:

- improved order of complexity
- easy to implement
- significant benefits for model problems

Future work:

- use of Milstein method and a control variate or antithetic variables to reduce complexity to $O(\varepsilon^{-2})$
- adaptive sampling to treat discontinuous payoffs and pathwise derivatives for Greeks
- use of quasi-Monte Carlo methods
- additional variance reduction techniques

Milstein Scheme

Generic SDE:

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t), \quad 0<t<T,$$

with correlation matrix $\Omega(S,t)$ between elements of $DW(t)$.

Simplest Milstein scheme sets Lévy areas to zero to give

$$\widehat{S}_{i,n+1} = \widehat{S}_{i,n} + a_i\,h + b_{ij}\,\Delta W_{j,n} + \tfrac{1}{2}\,\frac{\partial b_{ij}}{\partial S_l}\,b_{lk}\left(\Delta W_{j,n}\,\Delta W_{k,n} - h\,\Omega_{jk}\right)$$

using implied summation convention.

Milstein Scheme

In scalar case:

- $O(h)$ strong convergence
- $O(\varepsilon^{-2})$ complexity for Lipschitz payoffs
- $O(\varepsilon^{-2}(\log \varepsilon)^2)$ complexity for digitals and Greeks

In vector case:

- still only $O(h^{1/2})$ strong convergence
- but $\widehat{S}_n - E[S \,|\, W_n] = O(h)$

Milstein Scheme

If a coarse path with timestep $2h$ is constructed using

$$\Delta W_n^c = \sqrt{2h}\, Y_n$$

where the Y_n are $N(0,1)$ random variables, and the fine path uses a Brownian Bridge construction with

$$\Delta W_n^f = \tfrac{1}{2}\sqrt{2h}\,(Y_n\!+\!Z_n), \quad \Delta W_{n+\frac{1}{2}}^f = \tfrac{1}{2}\sqrt{2h}\,(Y_n\!-\!Z_n).$$

where the Z_n are also $N(0,1)$ random variables, then perturbation analysis shows that the $O(h^{1/2})$ difference between the two paths comes from a sum of terms proportional to

$$Y_{j,n} Z_{k,n} - Y_{k,n} Z_{j,n}.$$

Milstein Scheme

Using the idea of antithetic variables, we use the estimator

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\tfrac{1}{2} \left(\widehat{P}_l^{(i)} + \widehat{P}_l^{(i)*} \right) - \widehat{P}_{l-1}^{(i)} \right),$$

where $\widehat{P}_l^{(i)*}$ is based on the same coarse path Y_n, but with Z_n replaced by $-Z_n$, which leads to the cancellation of the leading order error proportional to Z_n.

- $V[\widehat{Y}_l] = O(h^2)$ for smooth payoffs, $O(h^{3/2})$ for Lipschitz
- in both cases, gives $O(\varepsilon^{-2})$ complexity for $O(\varepsilon)$ accuracy

Adaptive sampling

With digital options, the problem is that small path changes lead to an $O(1)$ change in the payoff

For the Euler discretisation, $O(h^{1/2})$ strong convergence $\Longrightarrow O(h^{1/2})$ paths have an $O(1)$ value for $\widehat{Y}_l$

Hence,

$$V[\widehat{Y}_l] = O(h^{1/2}).$$

For improved results, need more samples of paths near payoff discontinuities.

Adaptive sampling

Two ideas for adaptive sampling are both based on Brownian Bridge constructions, using coarse timestep realisations to decide which paths are “interesting” (i.e. likely to produce a large variance)

- idea 1: start with lots of paths, and prune those which are not interesting
- idea 2: start with relatively few paths, and sub-divide those which look interesting
- in each case, need to use path weights to ensure estimator remains unbiased
- no results yet, but I think this will make digital and barrier options as efficient as Lipschitz payoffs

Quasi-Monte Carlo

Quasi-Monte Carlo methods can offer greatly improved convergence with respect to the number of samples N:

- in the best case, $O(N^{-1+\delta})$ error for arbitrary $\delta > 0$, instead of $O(N^{-1/2})$
- depends on knowledge/identification of “important dimensions” in an application
  - Brownian Bridge
  - Principal Component Analysis
- most theory doesn’t apply to financial applications because of lack of payoff smoothness
- confidence intervals can be obtained by using randomized QMC
- my plans are to start by using Sobol sequences

Other Variance Reduction

- stratified sampling – probably not, because QMC has already done a good job of leading dimensions
- control variate – probably not (except perhaps for geometric Asian) multilevel approach can be viewed as using the coarse path value as a control variate
- importance sampling – might be useful for over-sampling the tails of the Normal distributions

Final words

Results so far:

- improved order of complexity
- easy to implement
- significant benefits for model problems

Future work:

- use of Milstein method and a control variate or antithetic variables to reduce complexity to $O(\varepsilon^{-2})$
- adaptive sampling to treat discontinuous payoffs and pathwise derivatives for Greeks
- use of quasi-Monte Carlo methods
- additional variance reduction techniques

Working Paper

M.B. Giles, “Multi-level Monte Carlo path simulation”
Oxford University Computing Laboratory
Numerical Analysis Report NA-06/03

`www.comlab.ox.ac.uk/mike.giles/finance.html`

Email: `giles@comlab.ox.ac.uk`