Multilevel Monte Carlo for Discontinuous Payoffs

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

University of Warwick, Nov 6th, 2009

SDEs in Finance

In computational finance, stochastic differential equations are used to model the behaviour of

- stocks
- interest rates
- exchange rates
- weather
- electricity/gas demand
- crude oil prices
- . . .

The stochastic term accounts for the uncertainty of unpredictable day-to-day events.

SDEs in Finance

Examples:

- Geometric Brownian motion (Black-Scholes model for stock prices)

$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W$$

- Cox-Ingersoll-Ross model (interest rates)

$$\mathrm{d}r = \alpha(b-r)\,\mathrm{d}t + \sigma\,\sqrt{r}\,\mathrm{d}W$$

- Heston stochastic volatility model (stock prices)

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

with correlation ρ between $\mathrm{d}W_1$ and $\mathrm{d}W_2$

Generic Problem

SDE with general drift and volatility terms:

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

In many finance applications, we want to compute the expected value of an option dependent on the terminal state

$$P \equiv f(S(T))$$

Initially, will assume the “payoff” function $f(U)$ has a uniform Lipschitz bound,

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

Standard MC Approach

Euler discretisation with timestep h:

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

and Milstein discretisation for a scalar SDE:

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

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

$$\widehat{Y} = N^{-1} \sum_{i=1}^{N} \widehat{P}^{(i)}$$

Standard MC Approach

The mean square error is defined as

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] &= \mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[\widehat{P}] \; + \; \mathbb{E}[\widehat{P}] - \mathbb{E}[P]\right)^2\right] \\ &= \mathbb{E}\left[(\widehat{Y} - \mathbb{E}[\widehat{P}])^2\right] + (\mathbb{E}[\widehat{P}] - \mathbb{E}[P])^2 \\ &= N^{-1}\mathbb{V}[\widehat{P}] + \left(\mathbb{E}[\widehat{P}] - \mathbb{E}[P]\right)^2\end{aligned}$$

- first term is due to variance of estimator
- second term is due to bias due to finite timestep
  – weak convergence

Standard MC Approach

Weak convergence:

- error in the expected value, $\mathbb{E}[\widehat{P}] - \mathbb{E}[P]$
- most important error in most applications
- $O(h)$ for both the Euler and Milstein discretisations

Strong convergence:

- error in path approximation

$$\sqrt{\mathbb{E}\left[\left\|\widehat{S}_{T/h} - S(T)\right\|^2\right]} \quad \text{or} \quad \sqrt{\mathbb{E}\left[\max_{0<t<T}\left\|\widehat{S}(t) - S(t)\right\|^2\right]}$$

- usually not relevant, but important for multilevel method
- $O(h^{1/2})$ for the Euler discretisation
  $O(h)$ for the Milstein discretisation

Standard MC Approach

Combined mean-square-error is $O(N^{-1} + h^2)$.

To make this equal to ε^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 $O(\varepsilon^{-2})$, by combining simulations with different numbers of timesteps – same accuracy as finest calculations, but at a much lower computational cost.

Note: objective is equivalent to using $O(1)$ timesteps per path, on average.

Multilevel MC Approach

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

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

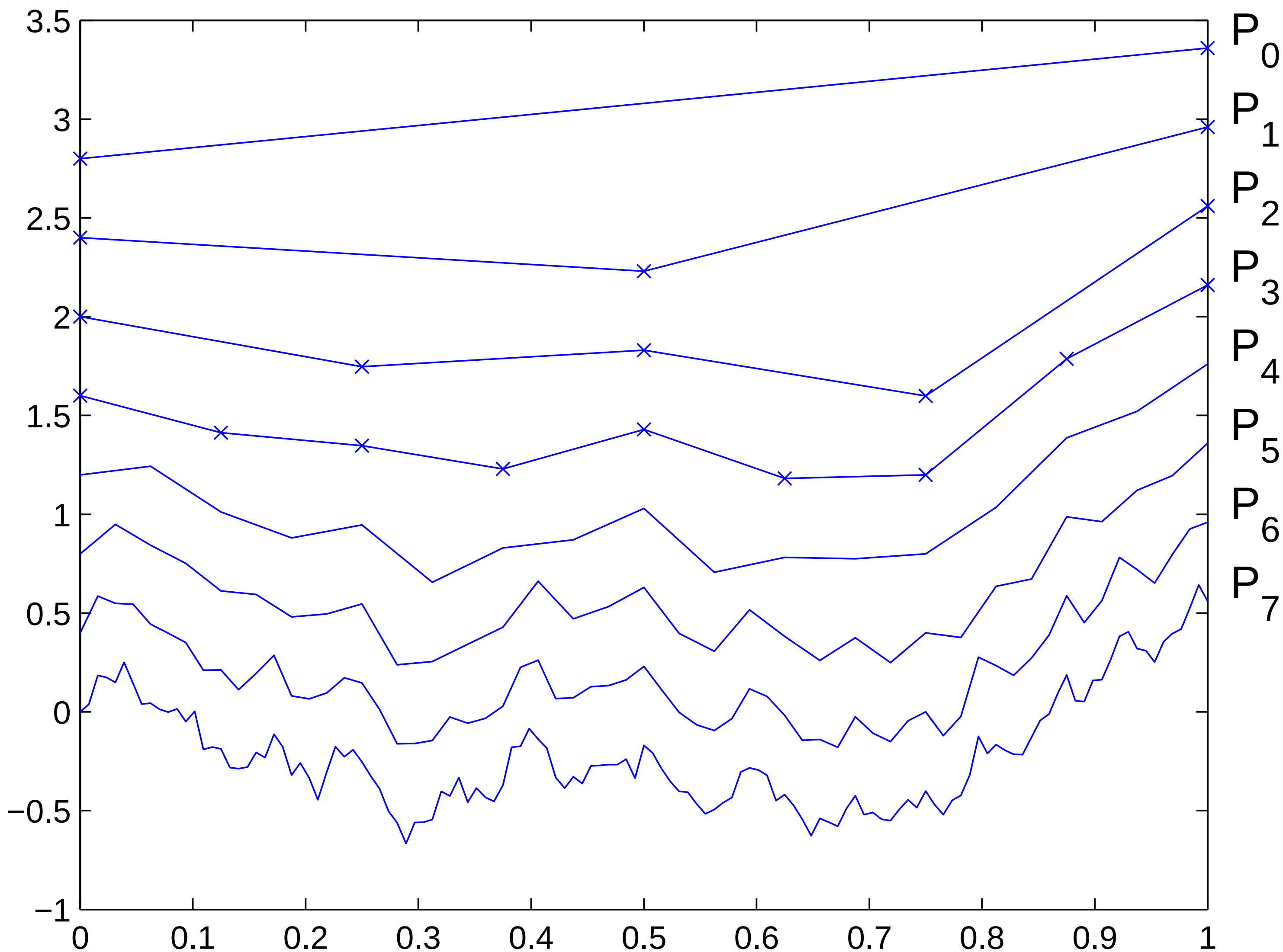
Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key point: approximate $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ 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

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

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 $\sqrt{V_l\, h_l}$.

The constant of proportionality can be chosen so that the combined variance is $O(\varepsilon^2)$.

Multilevel MC Approach

For the Euler discretisation and the Lipschitz payoff

$$\mathbb{V}[\widehat{P} - P] \;\leq\; \mathbb{E}[(\widehat{P} - P)^2] \;\leq\; c^2\, \mathbb{E}\left[\left|\widehat{S}_N - S(T)\right|^2\right] \;=\; O(h)$$

so $\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l)$ and the optimal N_l is $O(h_l)$.

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

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

To make the bias $O(\varepsilon)$ needs $h_L = O(\varepsilon) \implies L = O(\log_2 \varepsilon^{-1})$

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

For the Milstein discretisation

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

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

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

$$N_l = O(\varepsilon^{-2}\, h_l^{3/2})$$

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

Results

Geometric Brownian motion:

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

$T=1,\quad S(0)=100,\quad r=0.05,\quad \sigma=0.2$

European call option with discounted payoff

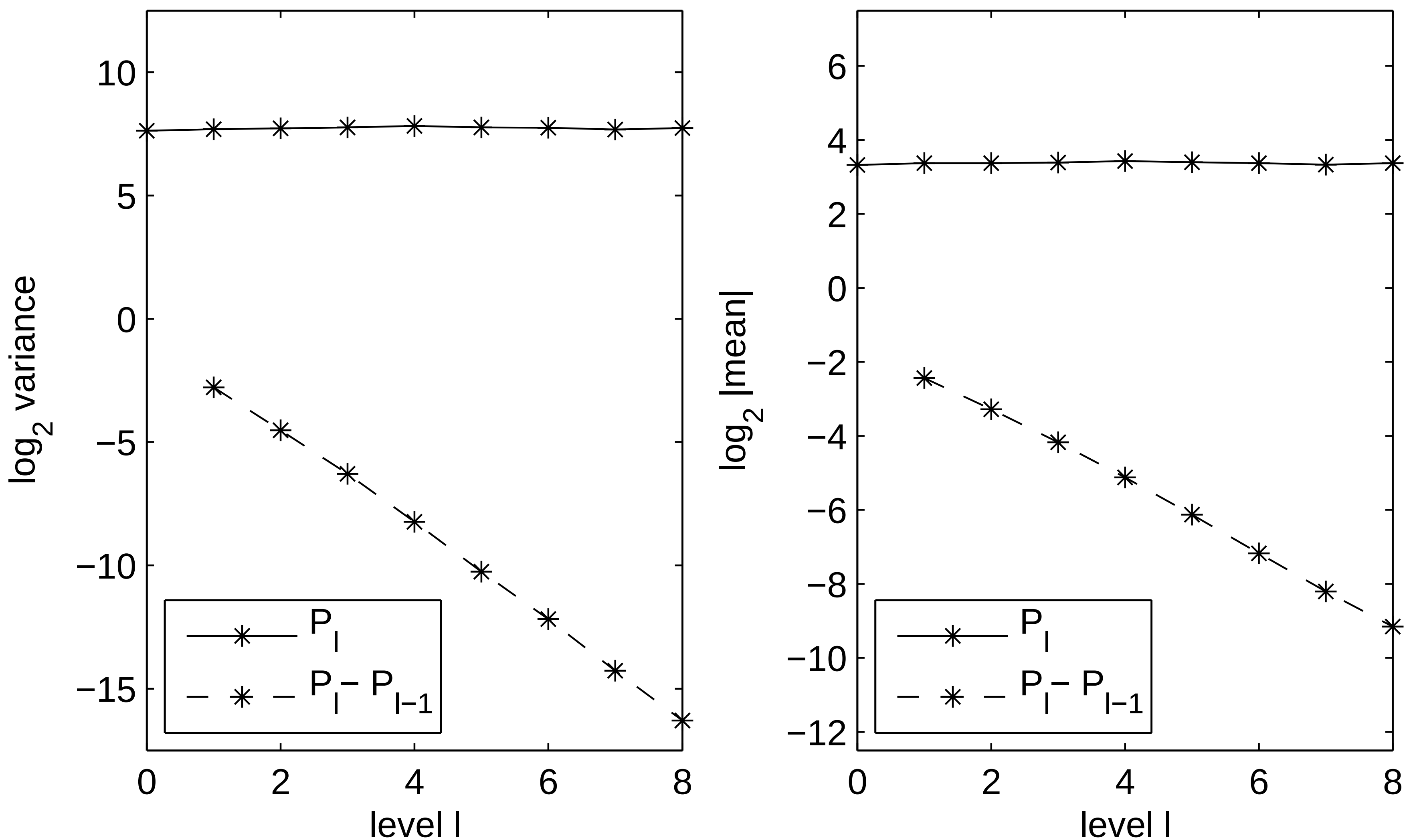
$$\exp(-rT)\ \max(S(T)-K,0)$$

with $K=100$.

Numerical results use the Milstein discretisation

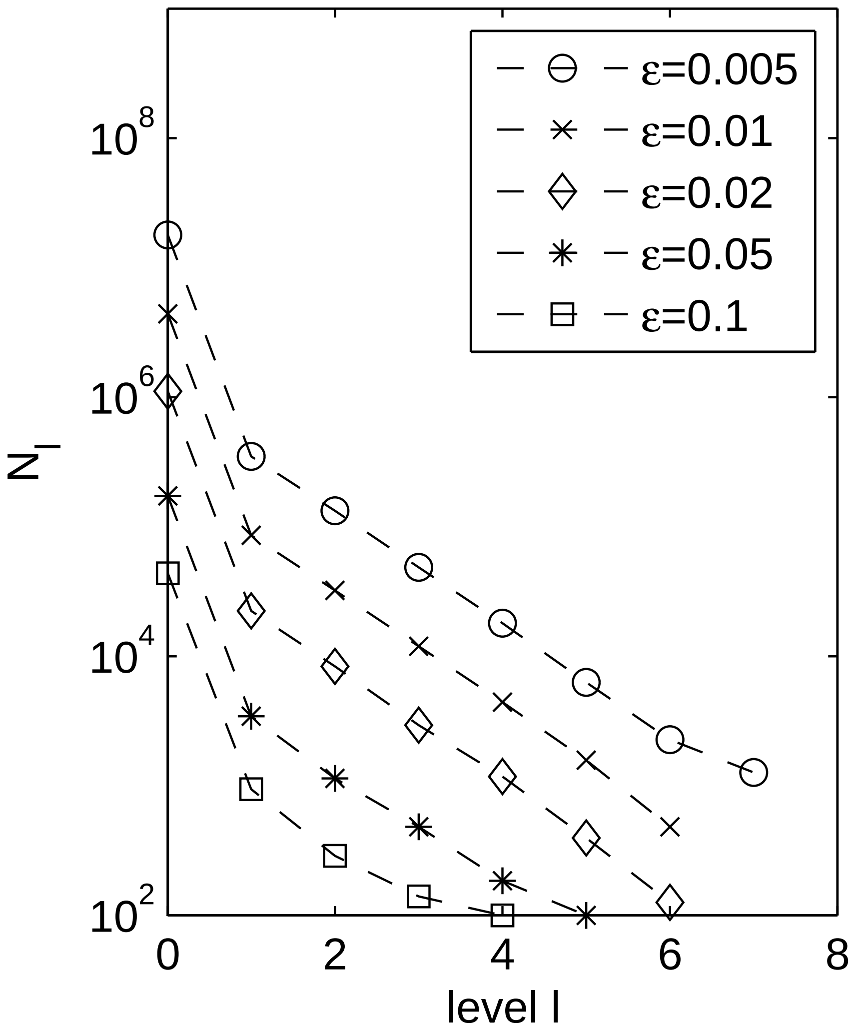
MLMC Results

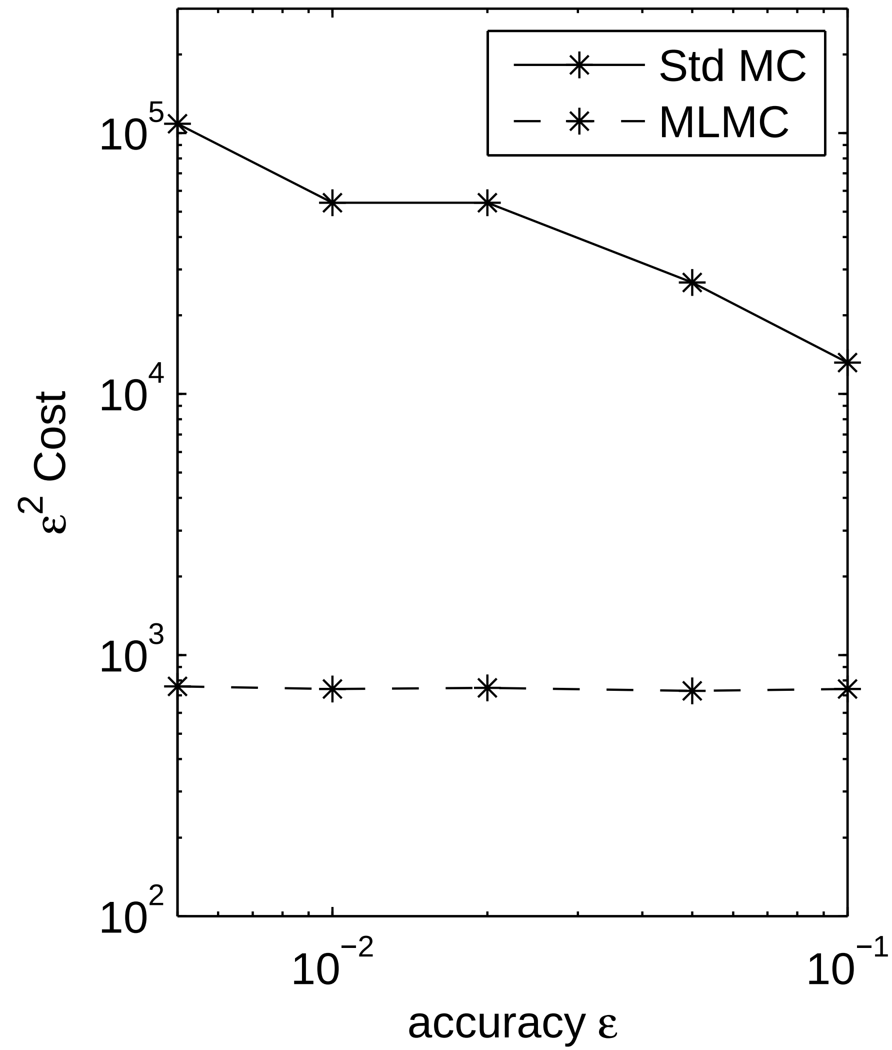
GBM: European call, $\exp(-rT)\,\max(S(T)-K,0)$



MLMC Results

GBM: European call, $\exp(-rT)\ \max(S(T)-K, 0)$





Multilevel MC Approach

Theorem: *Let P be a functional of the solution of an s.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) $\left| \mathbb{E}[\widehat{P}_l - P] \right| \leq c_1\, h_l^\alpha$

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*, 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*

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

Digital Options

What if we don't have the Lipschitz property?

A digital call payoff has the form

$$f(S(T)) = \begin{cases} 1, & S(T) > K \\ 0, & S(T) \leq K \end{cases}$$

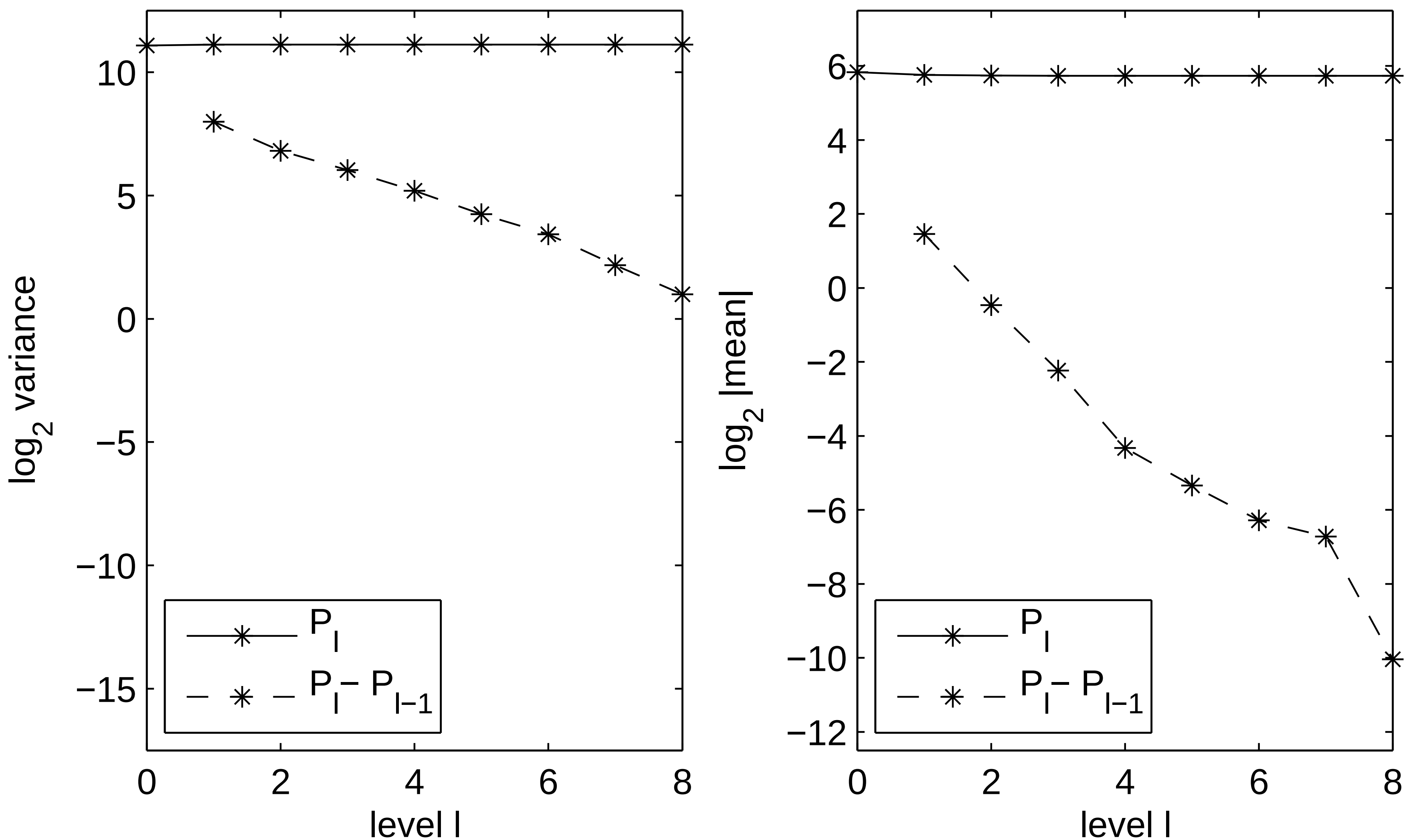
When using the Milstein discretisation

- in most cases, fine and coarse paths are on same side of K, so $\widehat{P}_l - \widehat{P}_{l-1} = 0$
- for $O(h_l)$ of the paths, fine and coarse paths end up on different sides of K so $\widehat{P}_l - \widehat{P}_{l-1} = \pm 1$

Hence $\mathbb{E}[(\widehat{P}_l - \widehat{P}_{l-1})^2]$ and $\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}]$ are both $O(h_l)$.

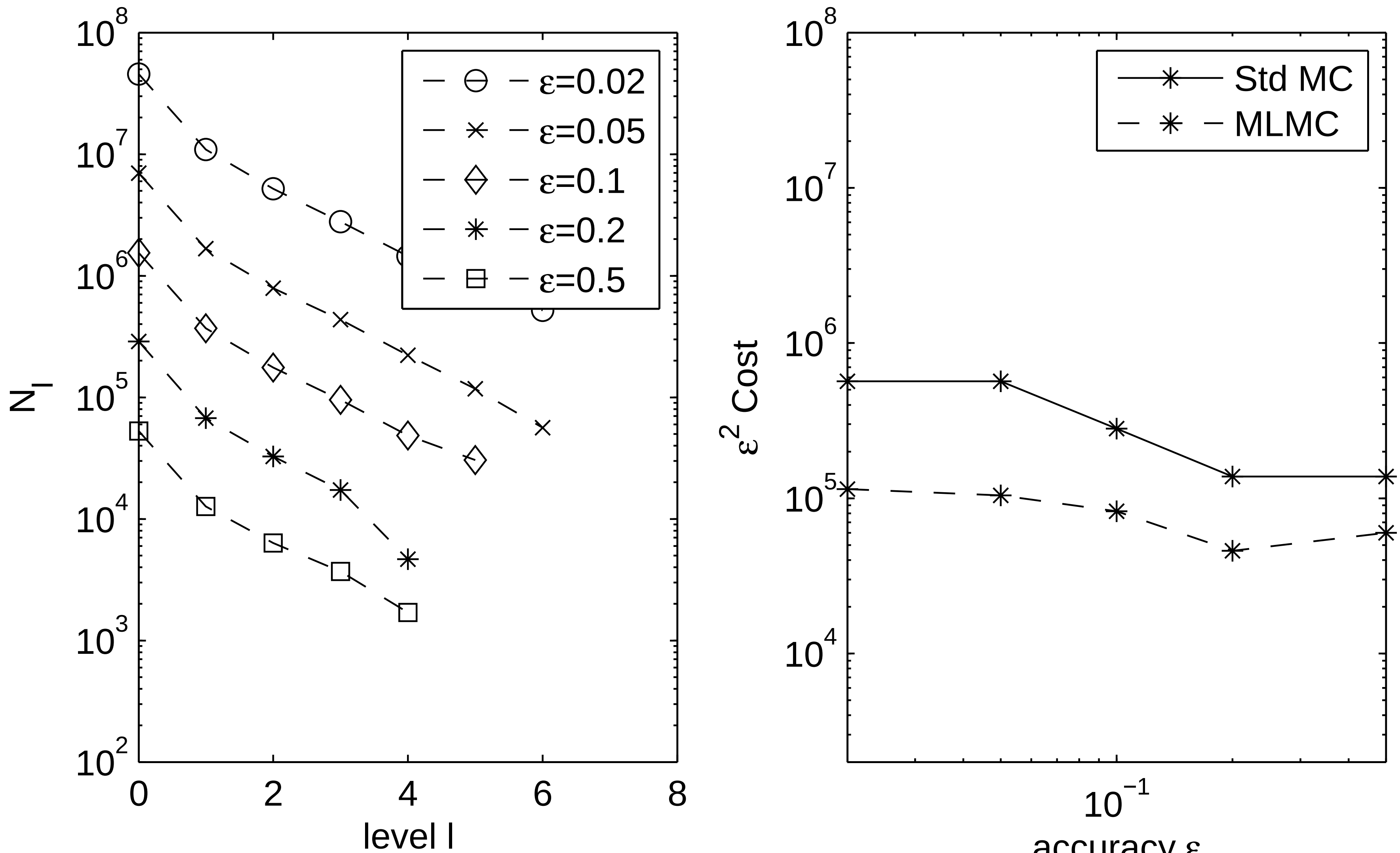
MLMC Results

GBM: digital call $K \exp(-rT)\, \mathbf{1}\{S(T)>K\}$



MLMC Results

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T) > K\}$



Digital Options

What is needed is to smooth the payoff.

On the fine path simulation, can stop one timestep before the end and use a conditional expectation for the final value.

$$\widehat{P}_l = \mathbb{E}_Z[f(\widehat{S}_N)\,|\,\widehat{S}_{N-1}]$$

where (for a scalar SDE)

$$\widehat{S}_N = \widehat{S}_{N-1} + a_{N-1}\,h + b_{N-1}\sqrt{h}\;Z$$

The key is that we know that

$$\mathbb{E}_Z[f(\widehat{S}_N)|\widehat{S}_{N-1}] = \Phi\left(\frac{\widehat{S}_{N-1} + a_{N-1}h - K}{b_{N-1}\sqrt{h}}\right)$$

where $\Phi()$ is the cumulative Normal distribution function.

Digital Options

What about the coarse path?

Could use

$$\widehat{P}_{l-1} = \mathbb{E}[f(\widehat{S}^c_N) \,|\, \widehat{S}^c_{N-2}] = \Phi\left(\frac{\widehat{S}^c_{N-2} + 2\,a^c_{N-2}h - K}{b^c_{N-1}\sqrt{2h}}\right)$$

but this gives $\widehat{P}_l - \widehat{P}_{l-1} = O(1)$ for paths near K, so no benefit

Instead, we want to define $\widehat{P}_{l-1}$ so that

- $\widehat{P}_l - \widehat{P}_{l-1}$ is small
- $\mathbb{E}[\widehat{P}_{l-1}] = \mathbb{E}[f(\widehat{S}^c_N) \,|\, \widehat{S}^c_{N-2}]$

Digital Options

Starting from

$$\widehat{S}^c_N = \widehat{S}^c_{N-2} + 2\, a^c_{N-2}\, h + b^c_{N-2} \left(\Delta W + \sqrt{h}\ Z \right)$$

where ΔW for first half timestep is same as for fine path, set

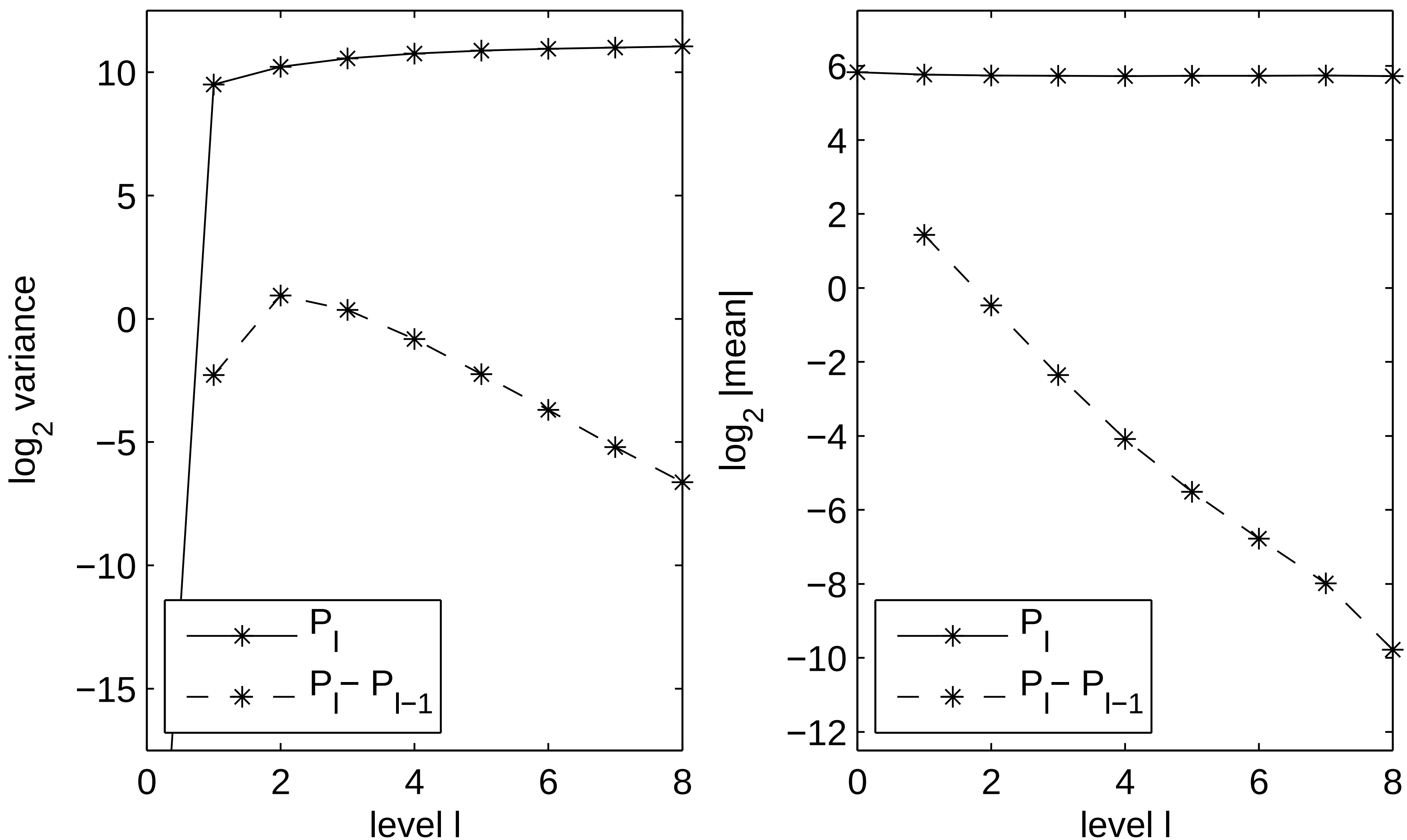
$$\begin{aligned} \widehat{P}_{l-1} &= \mathbb{E}_Z[f(\widehat{S}^c_N) \,|\, \widehat{S}^c_{N-2}, \Delta W] \\ &= \Phi\left(\frac{\widehat{S}^c_{N-2} + 2\, a^c_{N-2} h + b^c_{N-2} \Delta W - K}{b^c_{N-2} \sqrt{h}} \right) \end{aligned}$$

for which

- $\widehat{P}_l - \widehat{P}_{l-1} = O(h^{1/2})$ for paths near K
- $\mathbb{E}_{\Delta W}\left\{ \mathbb{E}_Z[f(\widehat{S}^c_N) \,|\, \widehat{S}^c_{N-2}, \Delta W] \right\} = \mathbb{E}[f(\widehat{S}^c_N) \,|\, \widehat{S}^c_{N-2}]$

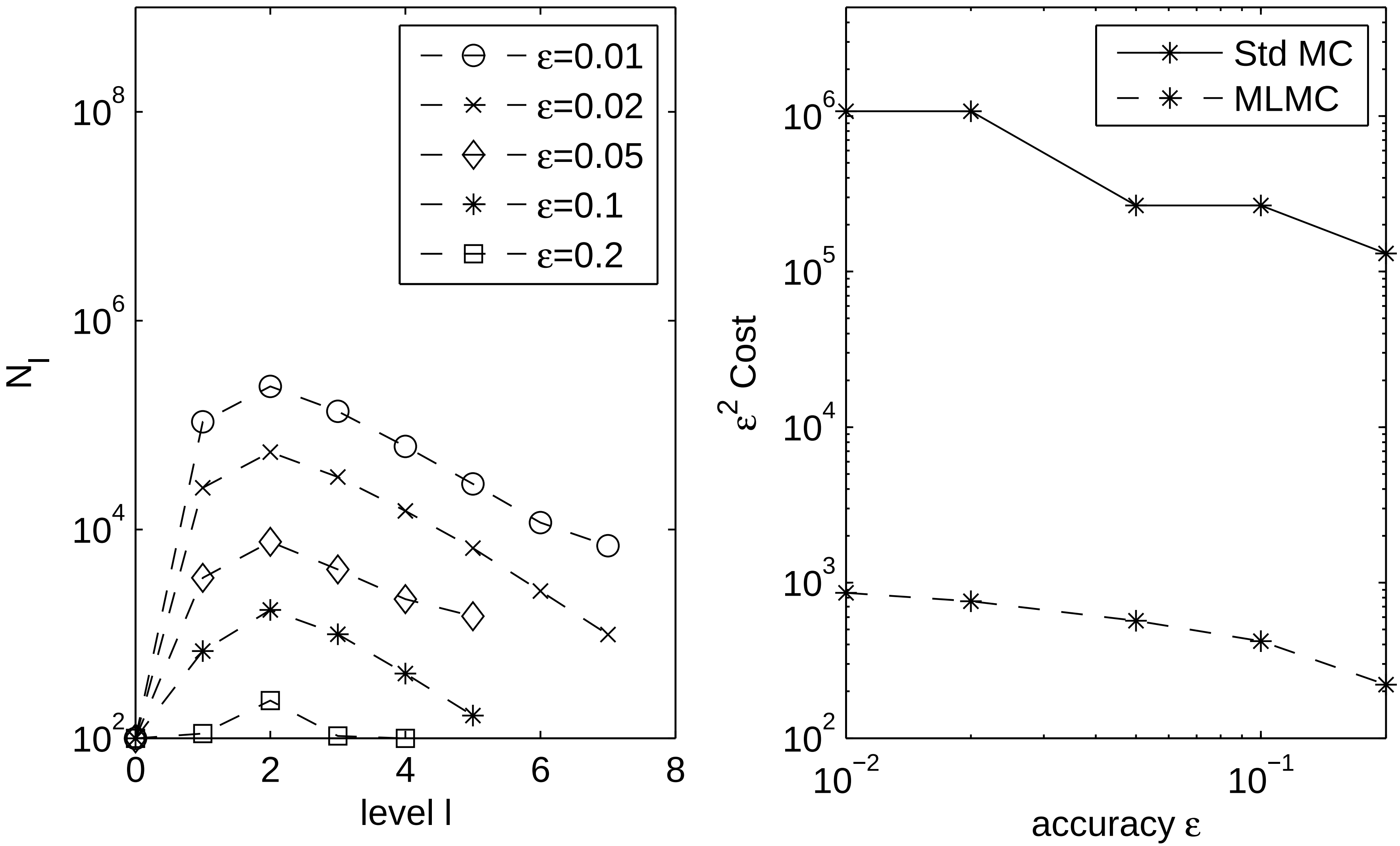
MLMC Results

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T)>K\}$



MLMC Results

GBM: digital call $K\,\exp(-rT)\,\mathbf{1}\{S(T)>K\}$



Digital Options

What if we don’t have an analytic expression for the conditional expectation?

Or if the payoff function is provided as a “black-box”?

Two solutions:

- use a change of measure
- use “splitting”

Change of Measure

If we have two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ with probability density functions $p_{\mathbb{P}}(x)$ and $p_{\mathbb{Q}}(x)$ then

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}[f(x)] &= \int p_{\mathbb{P}}(x)\ f(x)\,\mathrm{d}x \\
&= \int p_{\mathbb{Q}}(x)\ \frac{p_{\mathbb{P}}(x)}{p_{\mathbb{Q}}(x)}\ f(x)\,\mathrm{d}x \\
&= \mathbb{E}_{\mathbb{Q}}[r(x)\,f(x)]
\end{aligned}
$$

where $r(x) = \dfrac{p_{\mathbb{P}}(x)}{p_{\mathbb{Q}}(x)}$ is the Radon-Nikodym derivative.

(This is used in importance sampling to reduce the variance when the payoff is rarely non-zero)

Change of Measure

In our case, $\mathbb{P}_c$ and $\mathbb{P}_f$ corresponds to the conditional terminal distributions for the coarse and fine paths, and $\mathbb{Q}$ is a similar Gaussian distribution

We then get

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

Also, if $f \equiv 1$ we get $\mathbb{E}_{\mathbb{Q}}[r_f - r_c] = 0$, and hence when $f \neq 1$

$$\widehat{P}_l - \widehat{P}_{l-1} = \mathbb{E}_{\mathbb{Q}}[(r_f - r_c)(f - f_0)]$$

where f_0 is any fixed value (e.g. at peak of $\mathbb{Q}$)

Change of Measure

$\mathbb{Q}$ is taken to have a mean which is the average of the means for the coarse and fine paths, and a variance which is equal to the sum of their variances (not the average)

This makes $r_f - r_c$ small in the tails where $f - f_0$ is largest.

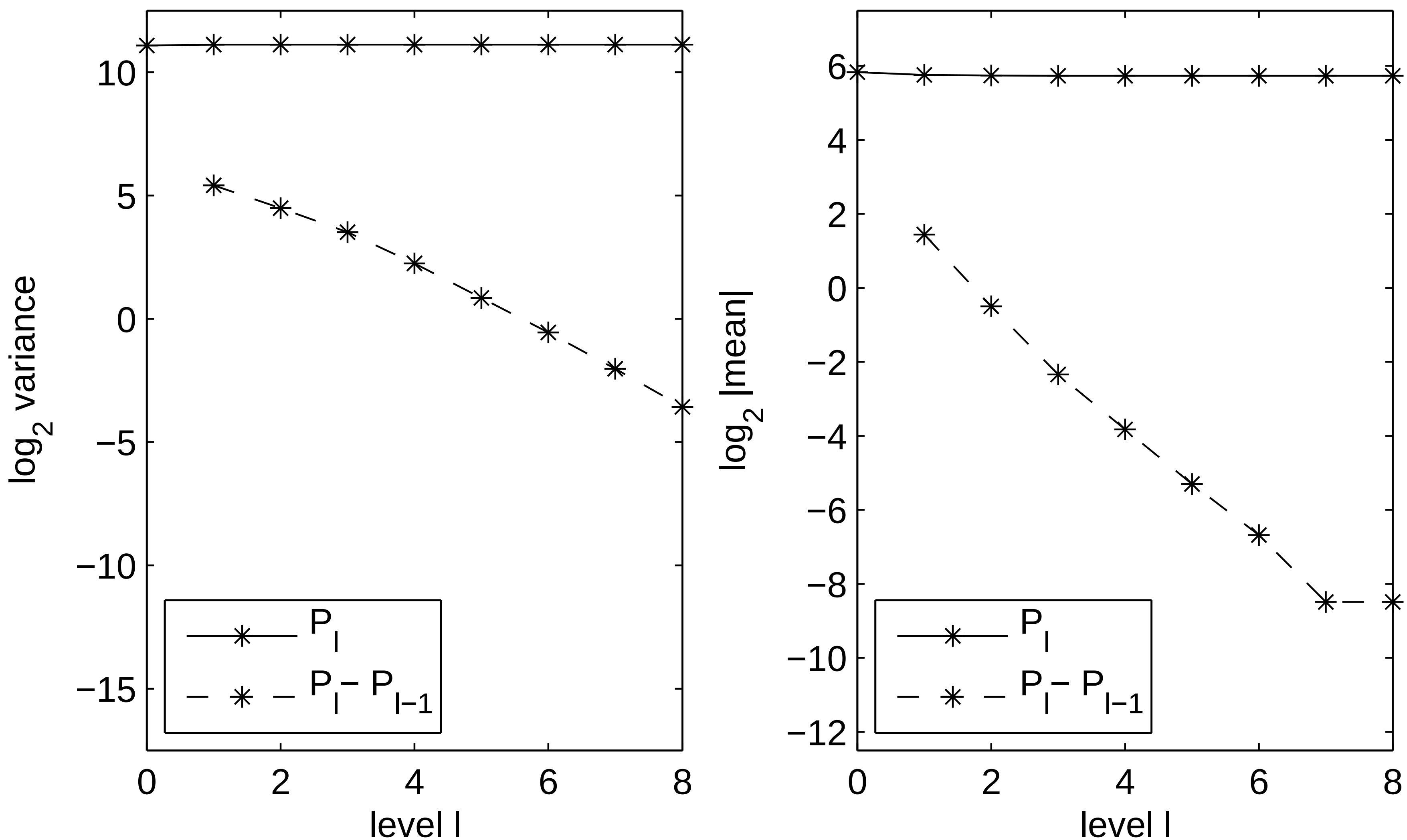
The expectation can be estimated using a simgle sample

For $O(h^{1/2})$ paths near the strike, $r_f - r_c = O(h^{1/2})$ and $f - f_0 = O(1)$, while for remainder $f - f_0 = 0$.

Hence, $\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l^{3/2})$

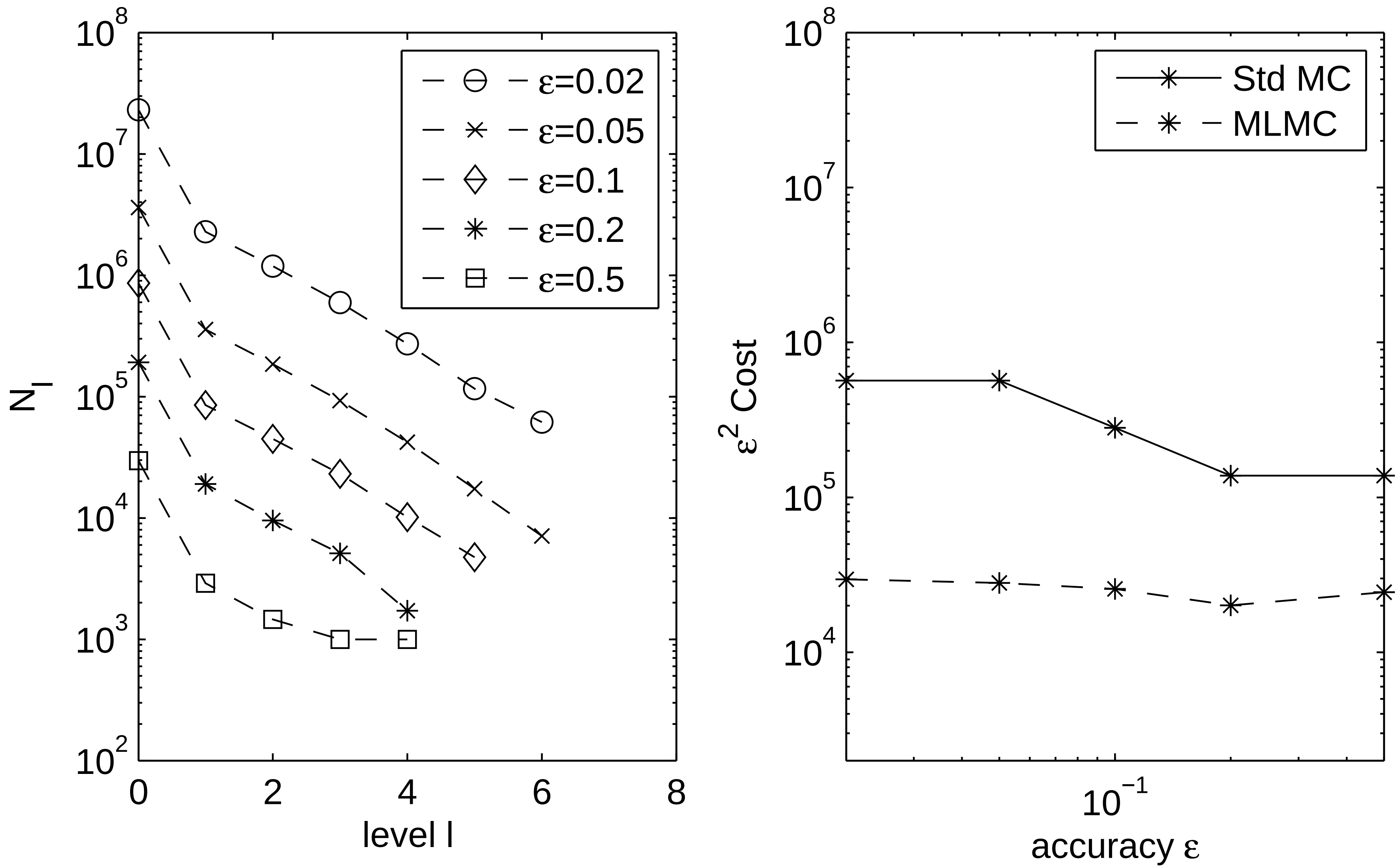
Change of Measure

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T)>K\}$



Change of Measure

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T) > K\}$



Splitting

Splitting uses multiple samples to estimate the value of the final conditional expectation.

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

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

Splitting

Going back to the original multilevel estimator (no conditional expectation) can argue that

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

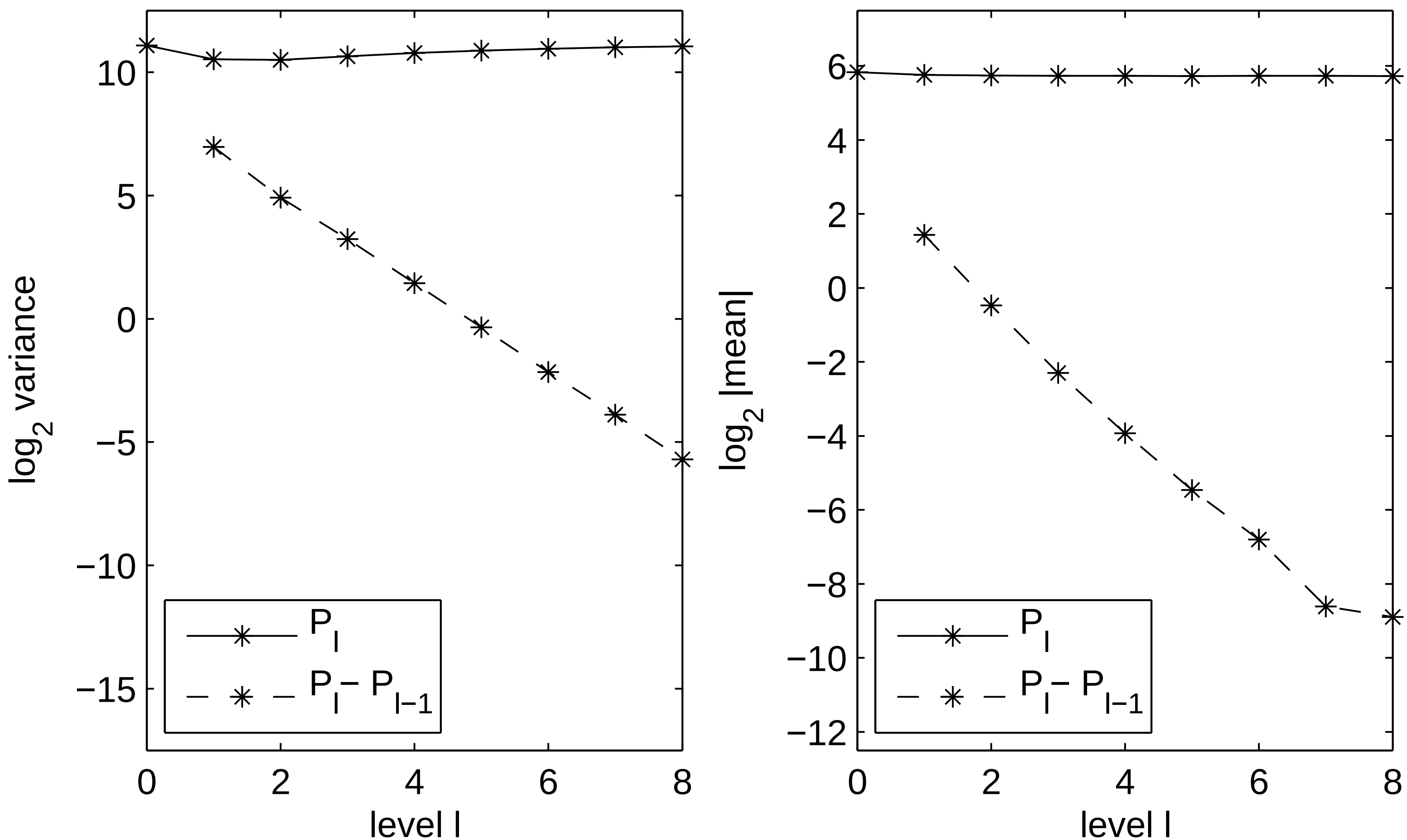
where $g(W,Z) \equiv \widehat{P}_l - \widehat{P}_{l-1}$. Hence, provided

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

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

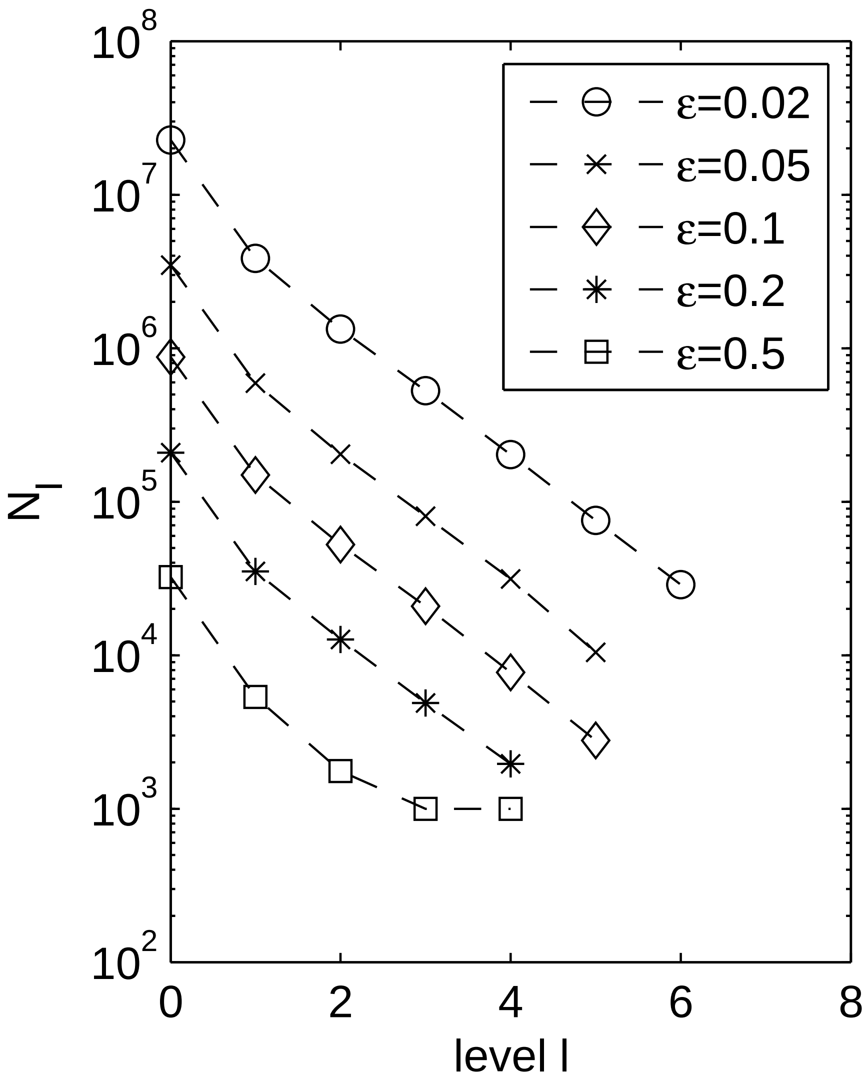
Splitting

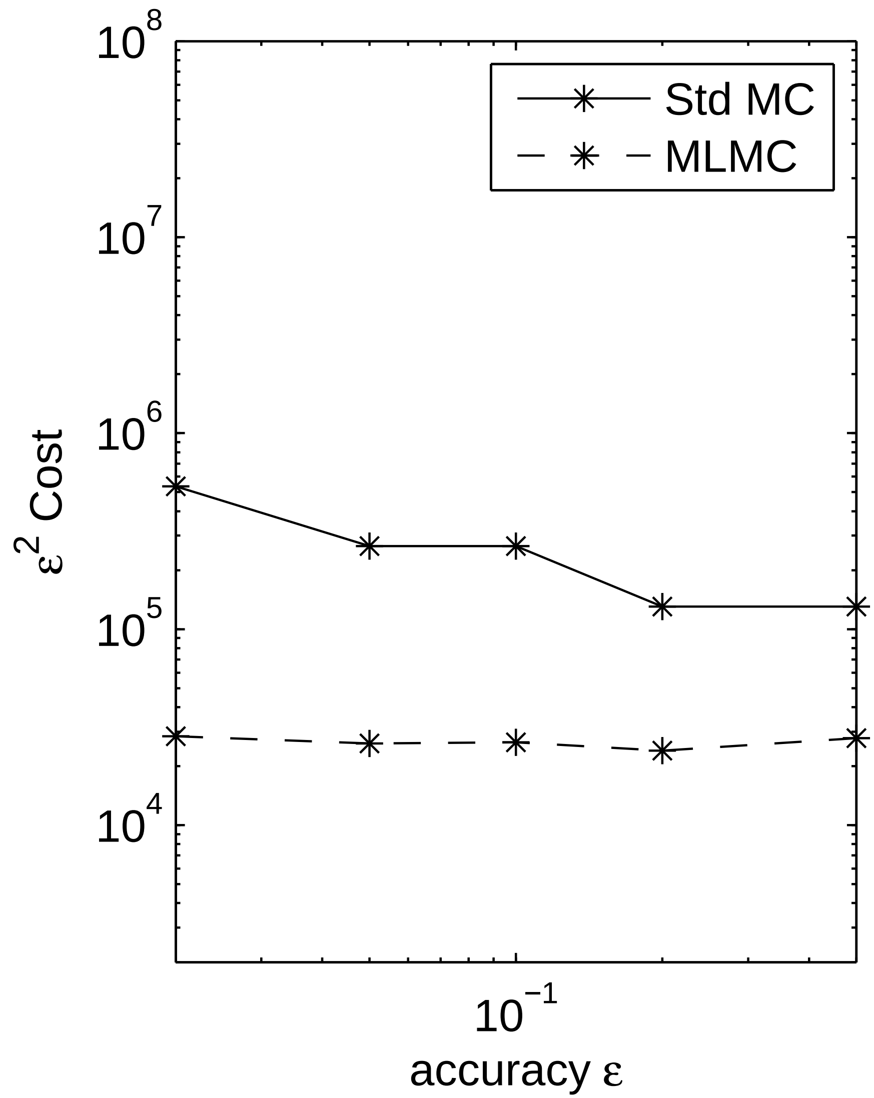
GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T)>K\}$



Splitting

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T)>K\}$





Basket Option

The technqiues extend naturally to multivariate cases.

For example, the analytic conditional expectation can be used for a basket option in which the payoff is based on a weighted average of several stocks.

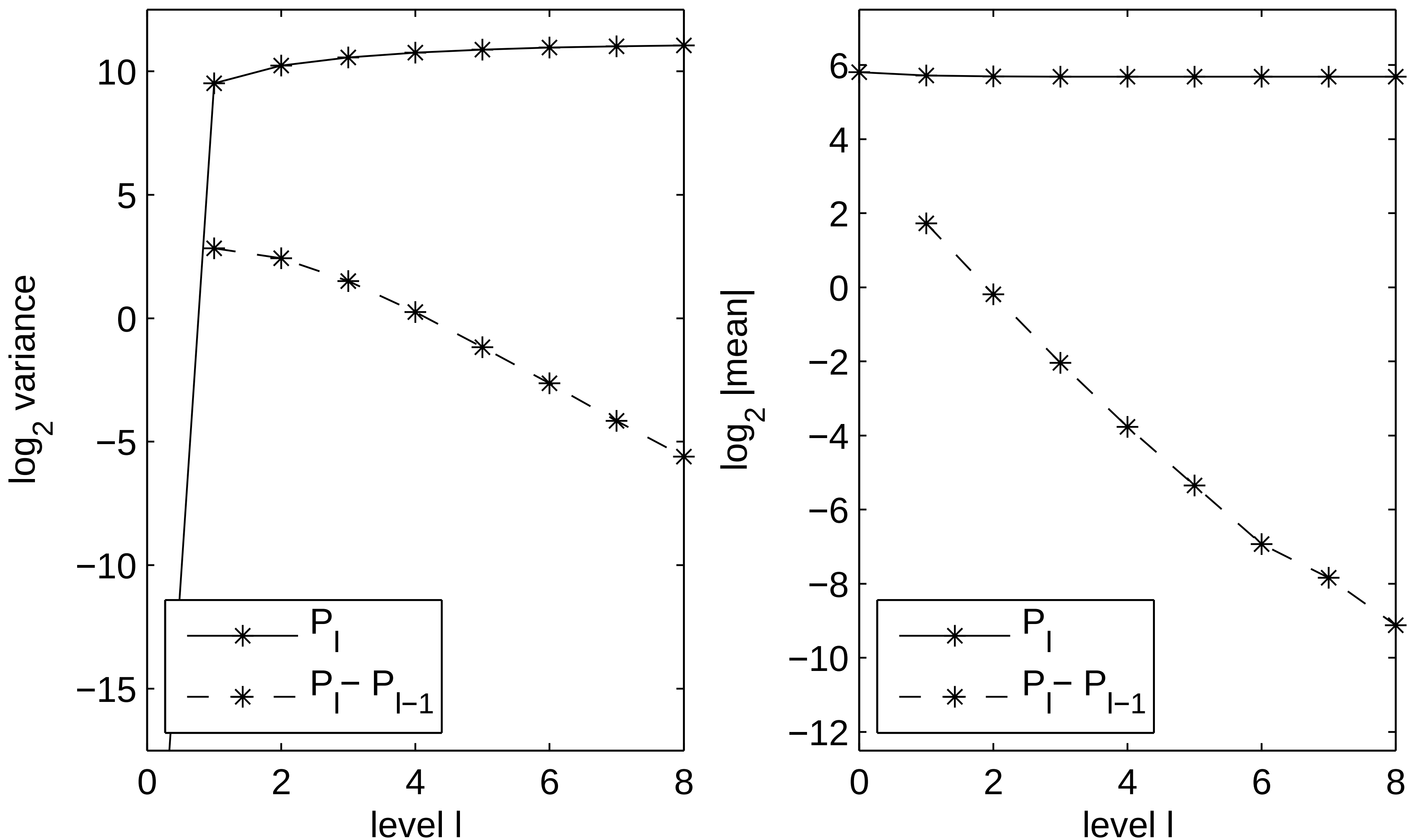
Basket of 5 underlying assets, each GBM with

$$r=0.05,\;\; T=1,\;\; S_i(0)=100,\;\; \sigma=(0.2,\,0.25,\,0.3,\,0.35,\,0.4),$$

and correlation $\rho=0.25$ between each of the driving Brownian motions.

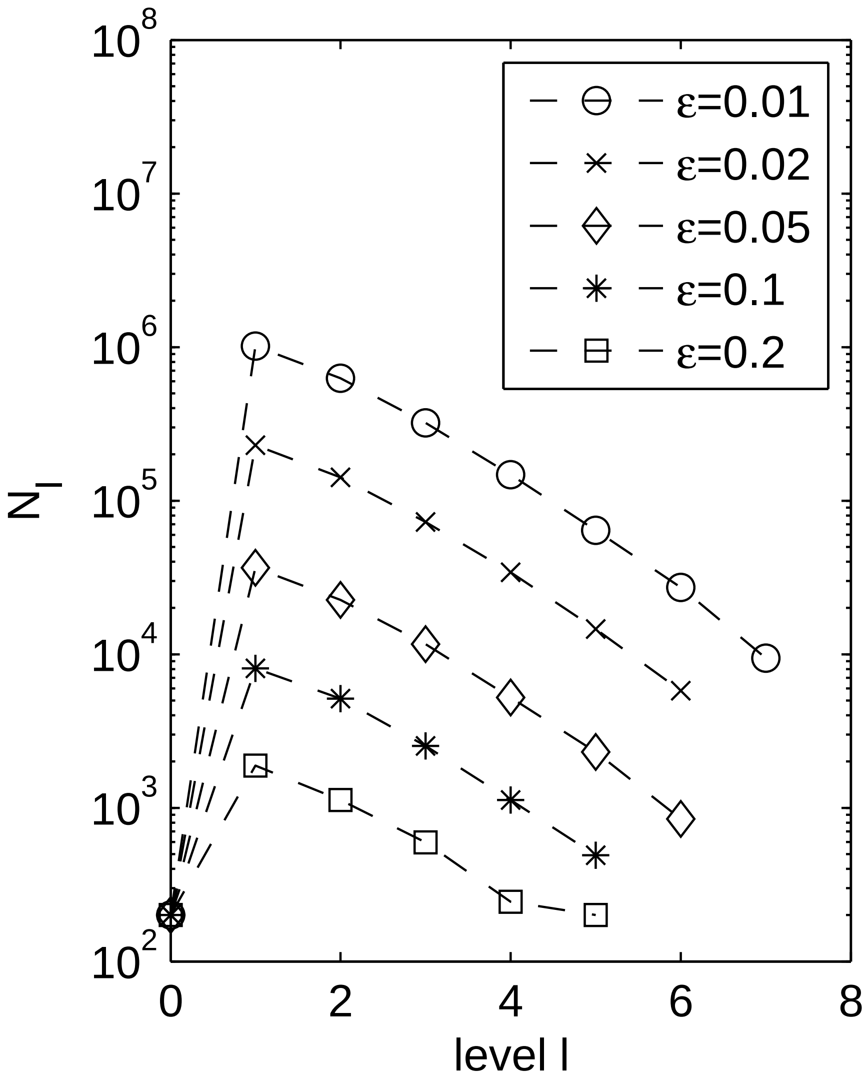
Basket Option

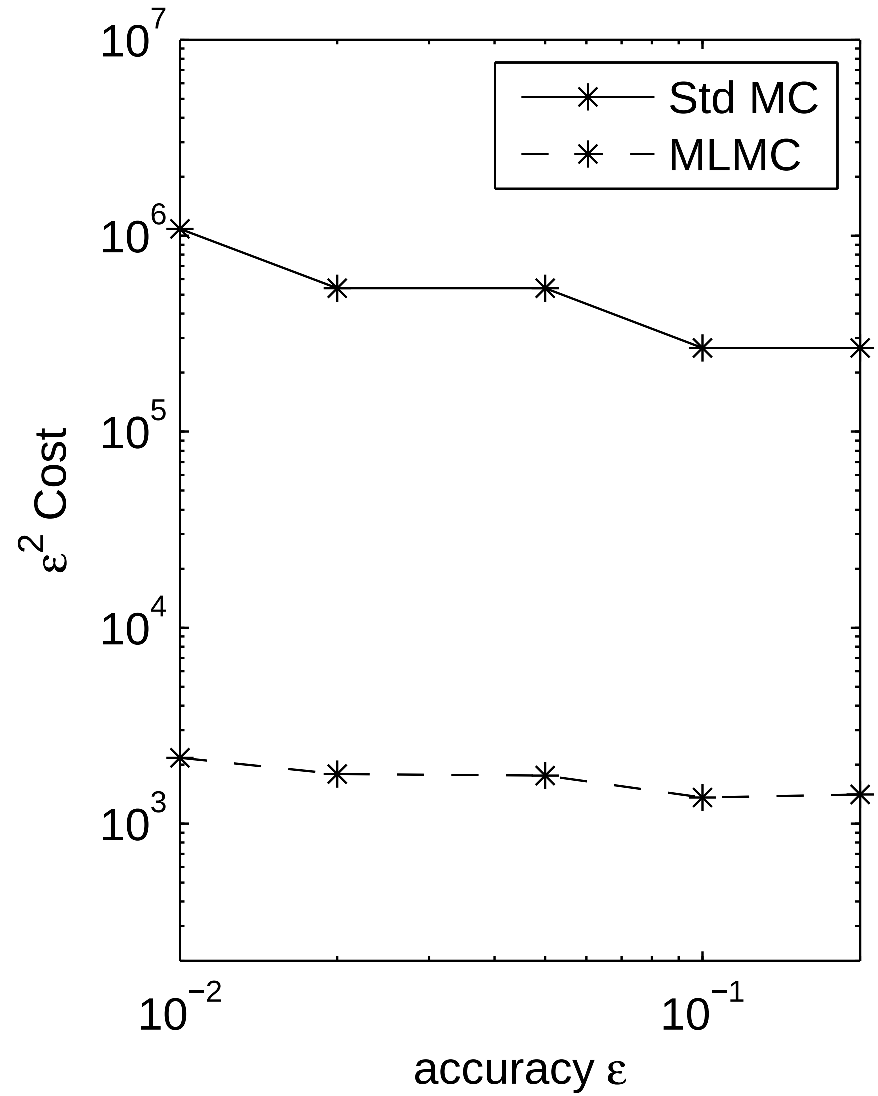
GBM: digital call on basket of 5 assets



Basket Option

GBM: digital call on basket of 5 assets





Conclusions

- Multilevel Monte Carlo method delivers an improved order of complexity for many applications
- Discontinuous payoffs pose an interesting challenge, but can be treated using conditional expectation to smooth the payoff
- For cases without analytic values, can use either "splitting" or a change of measure
- Conditional expectation and "splitting" can be analysed rigorously; the change of measure is tougher
- Can also handle cases in which the payoff depends on values at intermediate times

Papers

"Multilevel Monte Carlo path simulation",
Operations Research, 56(3):607-617, 2008.

"Improved multilevel Monte Carlo convergence using the Milstein scheme", pp. 343-358 in *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer, 2007.

"Multilevel Monte Carlo for basket options",
Winter Simulation Conference 2009

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