Monte Carlo Methods for Uncertainty Quantification

Mike Giles

Mathematical Institute, University of Oxford

Contemporary Numerical Techniques

Lecture outline

Lecture 3: financial SDE applications

- financial models
- approximating SDEs
- weak and strong convergence
- mean square error decomposition
- multilevel Monte Carlo

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
- ...

SDEs in Finance

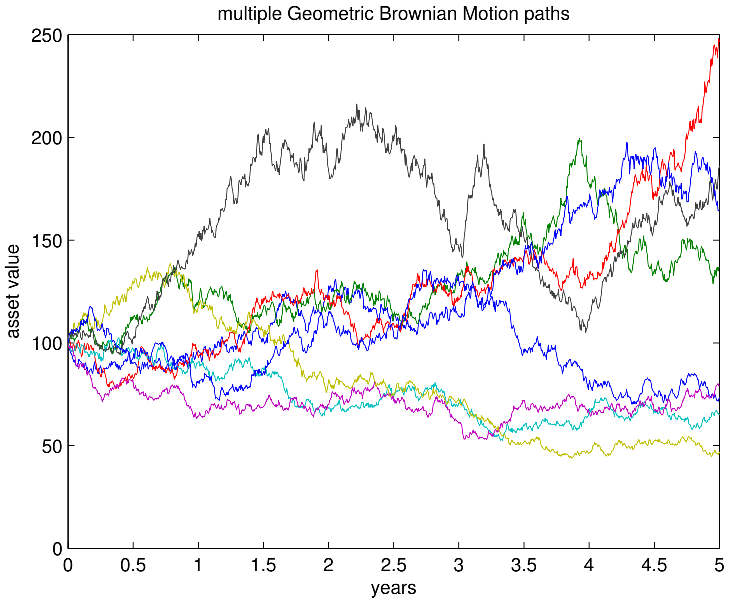
Stochastic differential equations are just ordinary differential equations plus an additional random source term.

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

The aim is **not** to predict exactly what will happen in the future, but to predict the probability of a range of possible things that **might** happen, and compute some averages, or the probability of an excessive loss.

This is really just uncertainty quantification, and they've been doing it for quite a while because they have so much uncertainty.

SDEs in Finance



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

Stochastic differential equation with general drift and volatility terms:

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

$W(t)$ is a Wiener variable with the properties that for any $q<r<s<t$, $W(t)-W(s)$ is Normally distributed with mean 0 and variance $t-s$, independent of $W(r)-W(q)$.

In many finance applications, we want to compute the expected value of an option dependent on the terminal state $P(S(T))$

Other options depend on the average, minimum or maximum over the whole time interval.

Euler discretisation

Given the generic SDE:

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

the Euler discretisation with timestep h is:

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

where ΔW_n are Normal with mean 0, variance h.

- How good is this approximation?
- How do the errors behave as $h \to 0$?

These are much harder questions when working with SDEs instead of ODEs.

Weak convergence

For most finance applications, what matters is the **weak** order of convergence, defined by the error in the expected value of the payoff.

For a European option, the weak order is m if

$$\mathbb{E}\left[f(S(T))\right] - \mathbb{E}\left[f(\widehat{S}_N)\right] = O(h^m)$$

The Euler scheme has order 1 weak convergence, so the discretisation "bias" is asymptotically proportional to h.

Strong convergence

In some Monte Carlo applications, what matters is the **strong** order of convergence, defined by the average error in approximating each individual path.

For the generic SDE, the strong order is m if

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

The Euler scheme has order 1/2 strong convergence.

The leading order errors are as likely to be positive as negative, and so cancel out – this is why the weak order is higher.

Exotic options

- Lookback option: $P = \left(S(T) - \min_{0<t<T} S(t)\right)$

  Approximation $\widehat{S}_{min} = \min_n \widehat{S}_n$ gives $O(h^{1/2})$ weak convergence

- Barrier option (down-and-out call):
  $P = \mathbf{1}(\min_{0<t<T} S(t) > B)\ \max(0, S(T)-K)$

  Approximation using $\widehat{S}_{min}$ gives $O(h^{1/2})$ weak convergence

It is possible to improve these (using something called a Brownian Bridge construction) and recover first order weak convergence.

Key point: getting high order convergence is **very** difficult.

Mean Square Error

Finally, how to decide whether it is better to increase the number of timesteps (reducing the weak error) or the number of paths (reducing the Monte Carlo sampling error)?

If the true option value is $V = \mathbb{E}[f]$

and the discrete approximation is $\widehat{V} = \mathbb{E}[\widehat{f}]$

and the Monte Carlo estimate is $\widehat{Y} = \frac{1}{N}\sum_{n=1}^{N} \widehat{f}^{(n)}$

then . . .

Mean Square Error

...the Mean Square Error is

$$\mathbb{E}\left[\left(\widehat{Y}-V\right)^2\right] = \mathbb{E}\left[\left(\widehat{Y}-\mathbb{E}[\widehat{f}] + \mathbb{E}[\widehat{f}]-\mathbb{E}[f]\right)^2\right]$$
$$= \mathbb{E}\left[(\widehat{Y}-\mathbb{E}[\widehat{f}])^2\right] + (\mathbb{E}[\widehat{f}]-\mathbb{E}[f])^2$$
$$= N^{-1}\mathbb{V}[\widehat{f}] + \left(\mathbb{E}[\widehat{f}]-\mathbb{E}[f]\right)^2$$

- first term is due to the variance of estimator
- second term is square of bias due to weak error

Hence the cost to achieve a RMS error of ε requires $N=O(\varepsilon^{-2})$, and $M=O(\varepsilon^{-1})$ timesteps (so that weak error is $O(\varepsilon)$) and hence the total cost is $O(\varepsilon^{-3})$.

Multilevel Monte Carlo

When solving finite difference equations coming from approximating PDEs, multigrid combines calculations on a nested sequence of grids to get the accuracy of the finest grid at a much lower computational cost.

Multilevel Monte Carlo uses 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.

Can also be viewed as a recursive control variate strategy.

Multilevel MC Approach

Consider multiple sets of simulations with different timesteps $h_\ell = 2^{-\ell}\, T,\ \ell = 0,1,\ldots,L$, and payoff approximation $\widehat{P}_\ell$ on level ℓ.

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

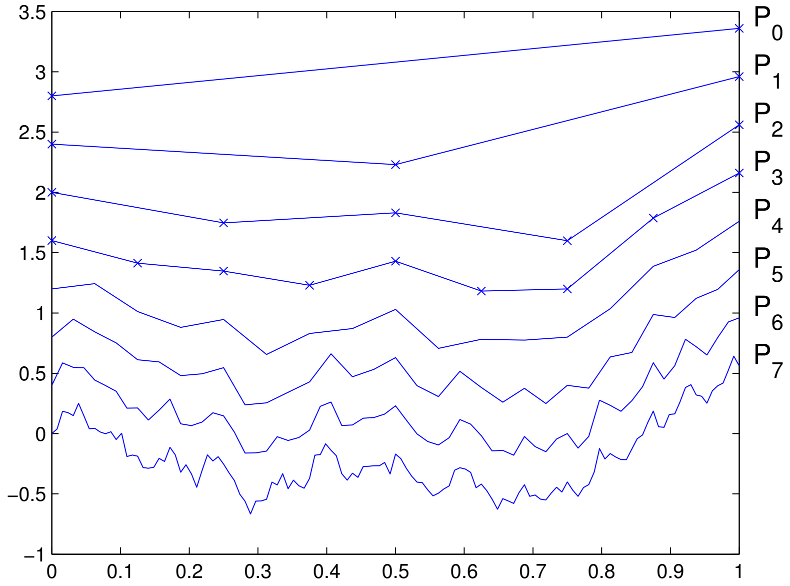
Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key point: approximate $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ using N_ℓ simulations with $\widehat{P}_\ell$ and $\widehat{P}_{\ell-1}$ obtained using same Brownian path.

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} \left(\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)}\right)$$

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

$$\mathbb{V}\left[\sum_{\ell=0}^{L} \widehat{Y}_\ell\right] = \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell, \qquad V_\ell \equiv \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}],$$

and the computational cost is proportional to $\displaystyle\sum_{\ell=0}^{L} N_\ell\, h_\ell^{-1}$.

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

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 function

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

and the optimal N_ℓ is asymptotically proportional to h_ℓ.

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

$$N_\ell = O(\varepsilon^{-2} L\, h_\ell).$$

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

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:

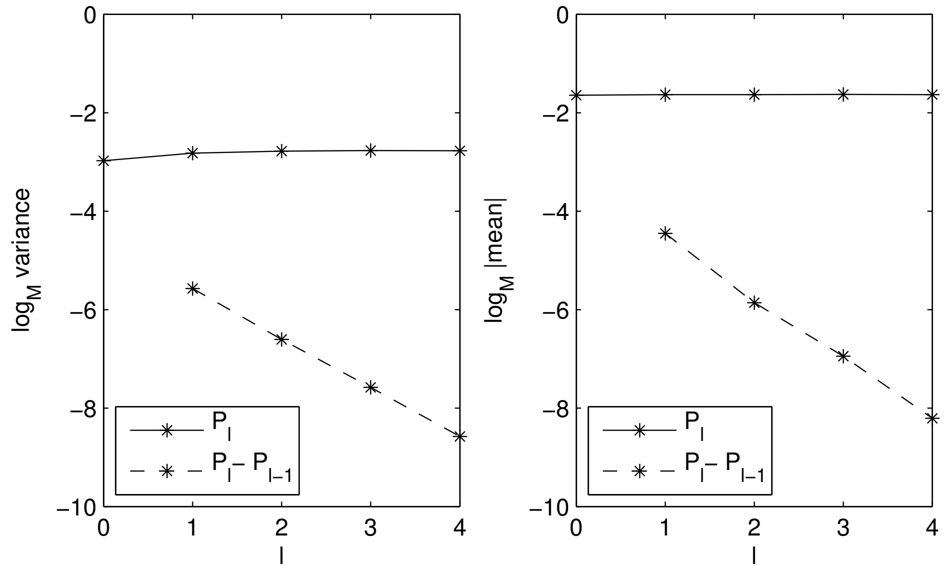
$$\begin{aligned} \mathrm{d}S &= r\, S\, \mathrm{d}t + \sqrt{V}\, S\, \mathrm{d}W_1, \qquad 0 < t < 1 \\ \mathrm{d}V &= \lambda\, (\sigma^2 - V)\, \mathrm{d}t + \xi\, \sqrt{V}\ \mathrm{d}W_2, \end{aligned}$$

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

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

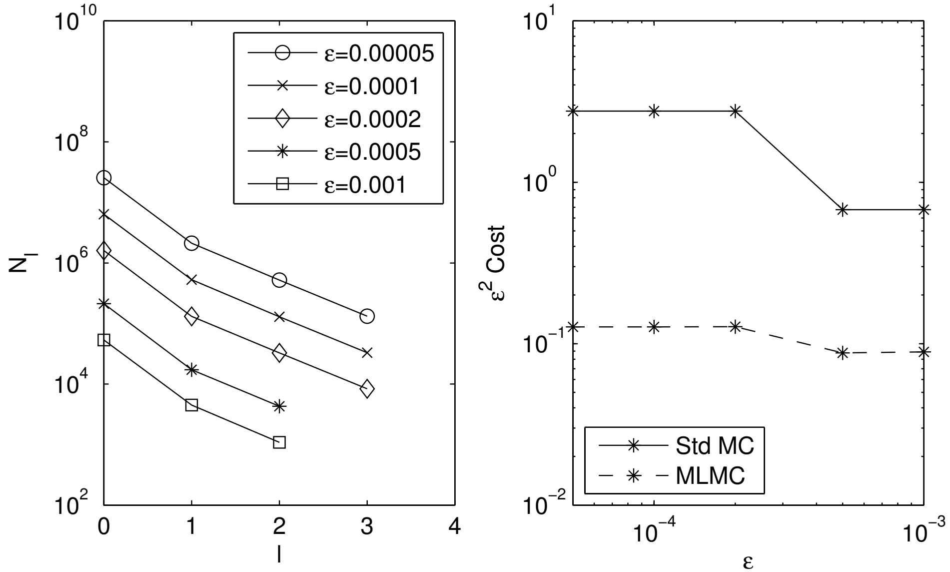
Results

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



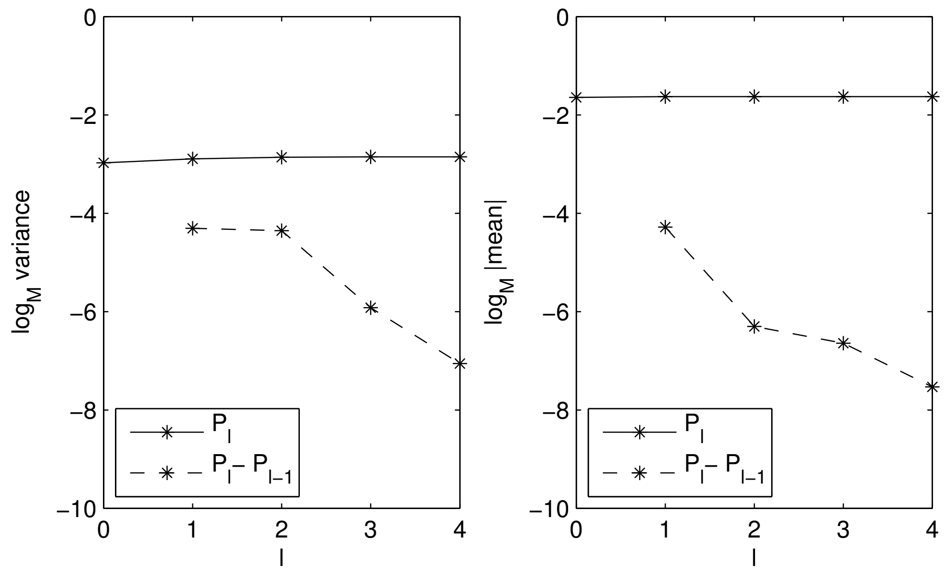
Results

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



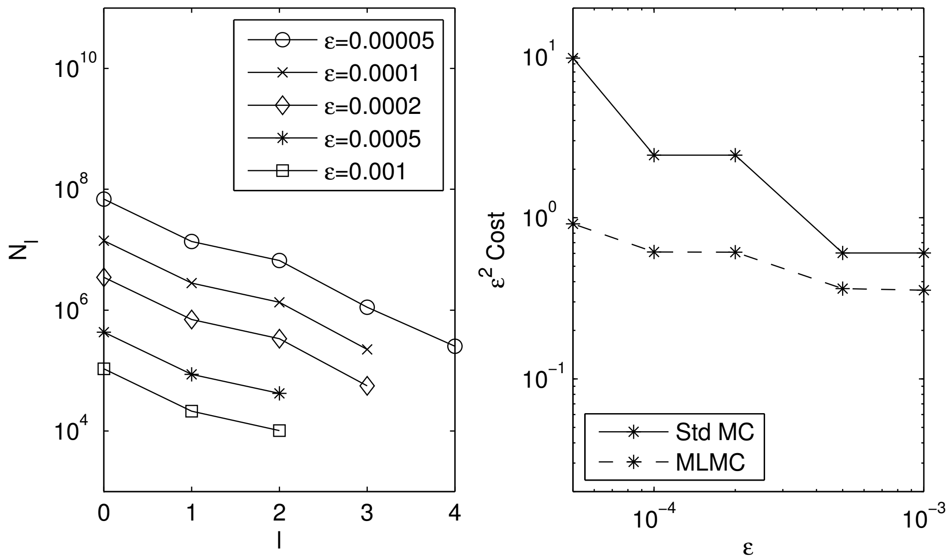
Results

Heston model: European call



Results

Heston model: European call



References

M.B. Giles, "Multi-level Monte Carlo path simulation", Operations Research, 56(3):607-617, 2008.

M.B. Giles. "Improved multilevel Monte Carlo convergence using the Milstein scheme", pages 343-358 in Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, 2008.

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

people.maths.ox.ac.uk/gilesm/mlmc_community.html