Stochastic Simulation: Lecture 7

Prof. Mike Giles

Oxford University Mathematical Institute

SDE Path Simulation

In these 2 lectures we are interested in SDEs of the form

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

in which the multi-dimensional Brownian motion W_t has covariance Σ.

The standard assumptions on the drift and diffusion functions are:

- Lipschitz in space:

$$\|a(x,t) - a(y,t)\| \leq L_a \|x - y\|, \quad \|b(x,t) - b(y,t)\| \leq L_b \|x - y\|,$$

- Hölder in time:

$$\|a(x,s) - a(x,t)\| \leq L_a(1 + \|x\|)\ |s-t|^{1/2},$$

$$\|b(x,s) - b(x,t)\| \leq L_b(1 + \|x\|)\ |s-t|^{1/2}$$

SDE Path Simulation

The simplest approximation is the forward Euler scheme, which is known as the Euler-Maruyama approximation when applied to SDEs:

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

Here h is the timestep, $\widehat{S}_n$ is the approximation to S_{nh} and the ΔW_n are i.i.d. $N(0, h\,\Sigma)$ Brownian increments.

For ODEs, the forward Euler method has $O(h)$ accuracy, and other more accurate methods would usually be preferred.

However, SDEs are very much harder to approximate so the Euler-Maruyama method is used widely in practice.

Numerical analysis is also very difficult and even the definition of "accuracy" is tricky.

Weak convergence

In many applications, mostly concerned with **weak** errors, the error in the expected value of an output quantity of interest (QoI).

If the QoI is $f(S_T)$ this is

$$\mathbb{E}[f(S_T)] - \mathbb{E}[f(\widehat{S}_{T/h})]$$

and it is of order α if

$$\mathbb{E}[f(S_T)] - \mathbb{E}[f(\widehat{S}_{T/h})] = O(h^\alpha)$$

For a path-dependent QoI, the weak error is

$$\mathbb{E}[f(S)] - \mathbb{E}[\widehat{f}(\widehat{S})]$$

where $f(S)$ is a function of the entire path S_t, and $\widehat{f}(\widehat{S})$ is a corresponding approximation.

Weak convergence

Key theoretical result (Bally and Talay, 1995):

If $p(S)$ is the p.d.f. for S_T and $\widehat{p}(S)$ is the p.d.f. for $\widehat{S}_{T/h}$ computed using the Euler-Maruyama approximation, then if $a(S,t)$ and $b(S,t)$ are Lipschitz w.r.t. S,t

$$\|p(S) - \widehat{p}(S)\|_1 = O(h)$$

and hence for bounded function f

$$\mathbb{E}[f(S_T)] - \mathbb{E}[f(\widehat{S}_{T/h})] = O(h)$$

This holds even if f is discontinuous; earlier theory only covered smooth f.

Weak convergence

Numerical demonstration: Geometric Brownian Motion

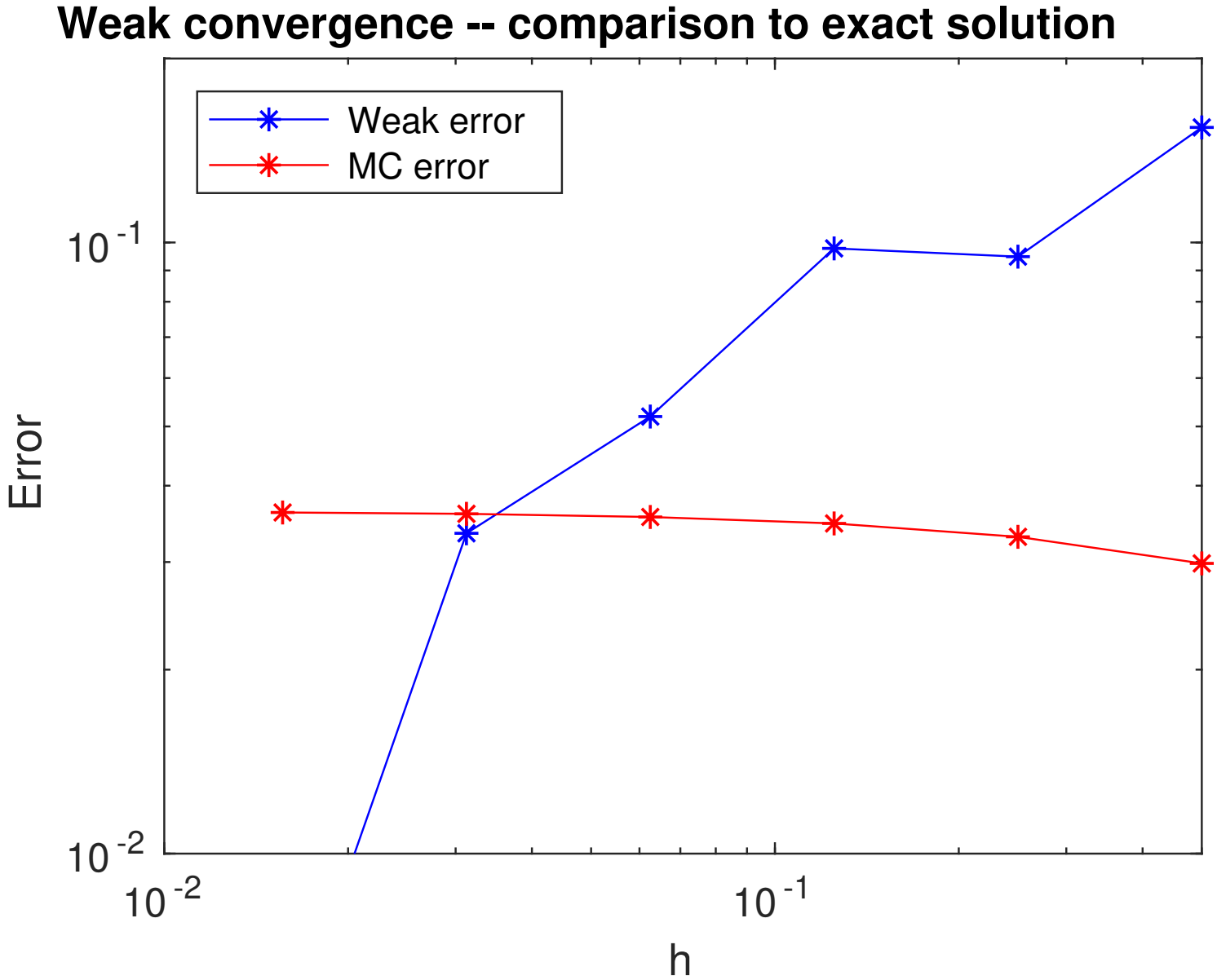
$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W$$

$r = 0.05,\ \sigma = 0.5,\ T = 1$

European call: $S_0 = 100, K = 110.$

Plot shows weak error versus analytic expectation when using 10^8 paths, and also Monte Carlo error (3 standard deviations)

Weak convergence



Weak convergence

Previous plot showed difference between exact expectation and numerical approximation.

What if the exact solution is unknown? Compare approximations with timesteps h and $2h$.

If

$$\mathbb{E}[f(S_T)] - \mathbb{E}[f(\widehat{S}^h_{T/h})] \approx a\ h$$

then

$$\mathbb{E}[f(S_T)] - \mathbb{E}[f(\widehat{S}^{2h}_{T/2h})] \approx 2\,a\,h$$

and so

$$\mathbb{E}[f(\widehat{S}^h_{T/h})] - \mathbb{E}[f(\widehat{S}^{2h}_{T/2h})] \approx a\,h$$

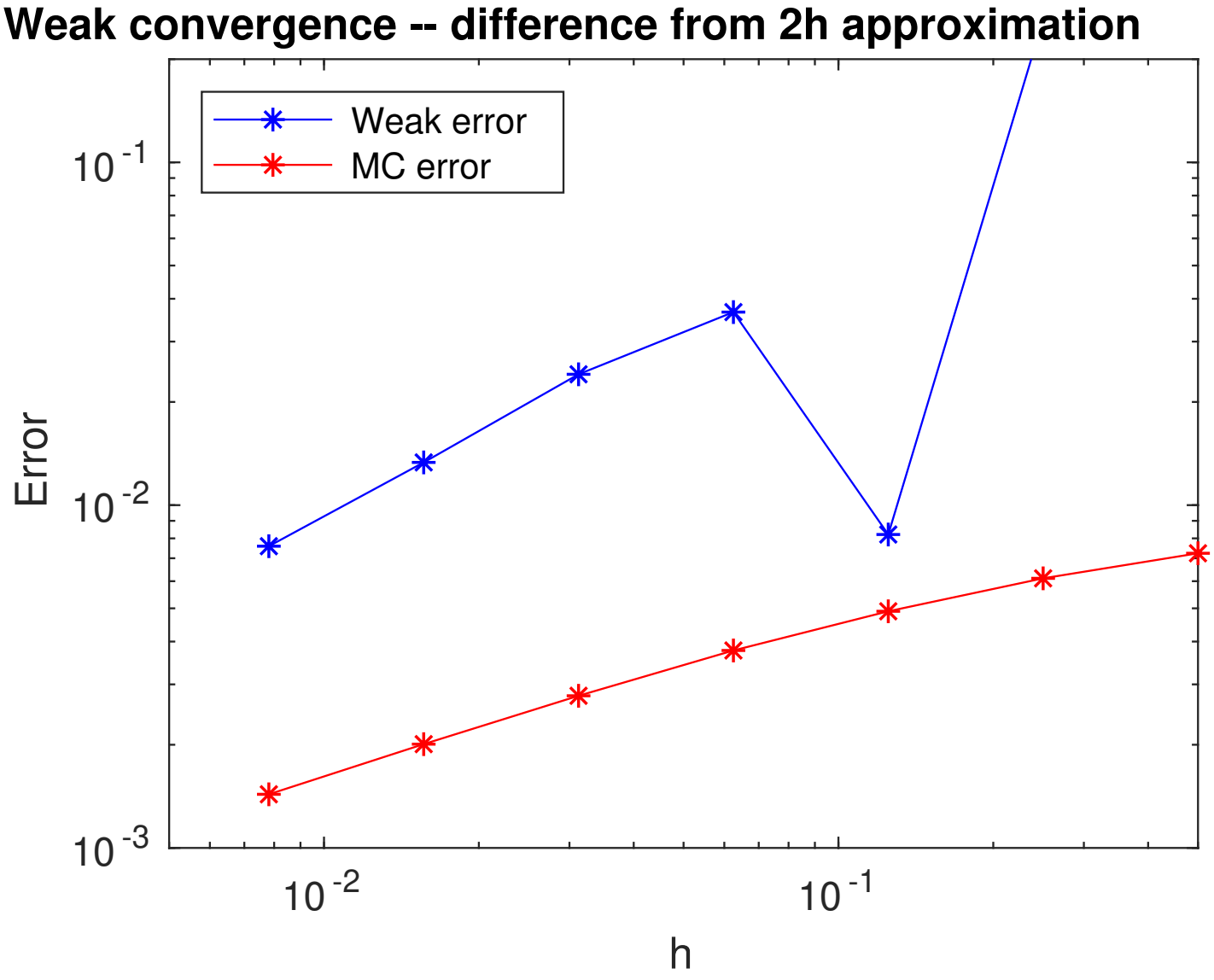
Weak convergence

To minimise the number of paths that need to be simulated, best to use **same** driving Brownian path when doing $2h$ and h approximations – i.e. take Brownian increments for h simulation and sum in pairs to get Brownian increments for $2h$ simulation.

The variance is lower because the h and $2h$ paths are close to each other (**strong** convergence).

This forms the basis for the **Multilevel Monte Carlo** method

Weak convergence



Strong convergence

Strong convergence looks instead at the average error in each individual path:

$$\left(\mathbb{E}\left[\left(S_T - \widehat{S}_{T/h}\right)^2\right]\right)^{1/2} \quad \text{or} \quad \left(\mathbb{E}\left[\sup_{[0,T]}\left(S_t - \widehat{S}_{\lfloor t/h \rfloor}\right)^2\right]\right)^{1/2}$$

It is of order β if it is $O(h^\beta)$ as $h \to 0$.

The main theoretical result (Kloeden & Platen 1992) is that for the Euler-Maruyama method these are both $O(\sqrt{h})$.

Strong convergence

Thus, each approximate path deviates by $O(\sqrt{h})$ from its true path.

How can the weak error be $O(h)$? Because the error

$$S_T - \widehat{S}_{T/h}$$

has mean $O(h)$ even though the r.m.s. is $O(\sqrt{h})$.

(In fact to leading order it is normally distributed with zero mean and variance $O(h)$.)

Strong convergence

Numerical demonstration for same Geometric Brownian Motion.

Plot shows two curves, one showing the difference from the true solution

$$S_T = S_0 \exp\left(\left(r - \tfrac{1}{2}\sigma^2\right)T + \sigma\, W(T)\right)$$

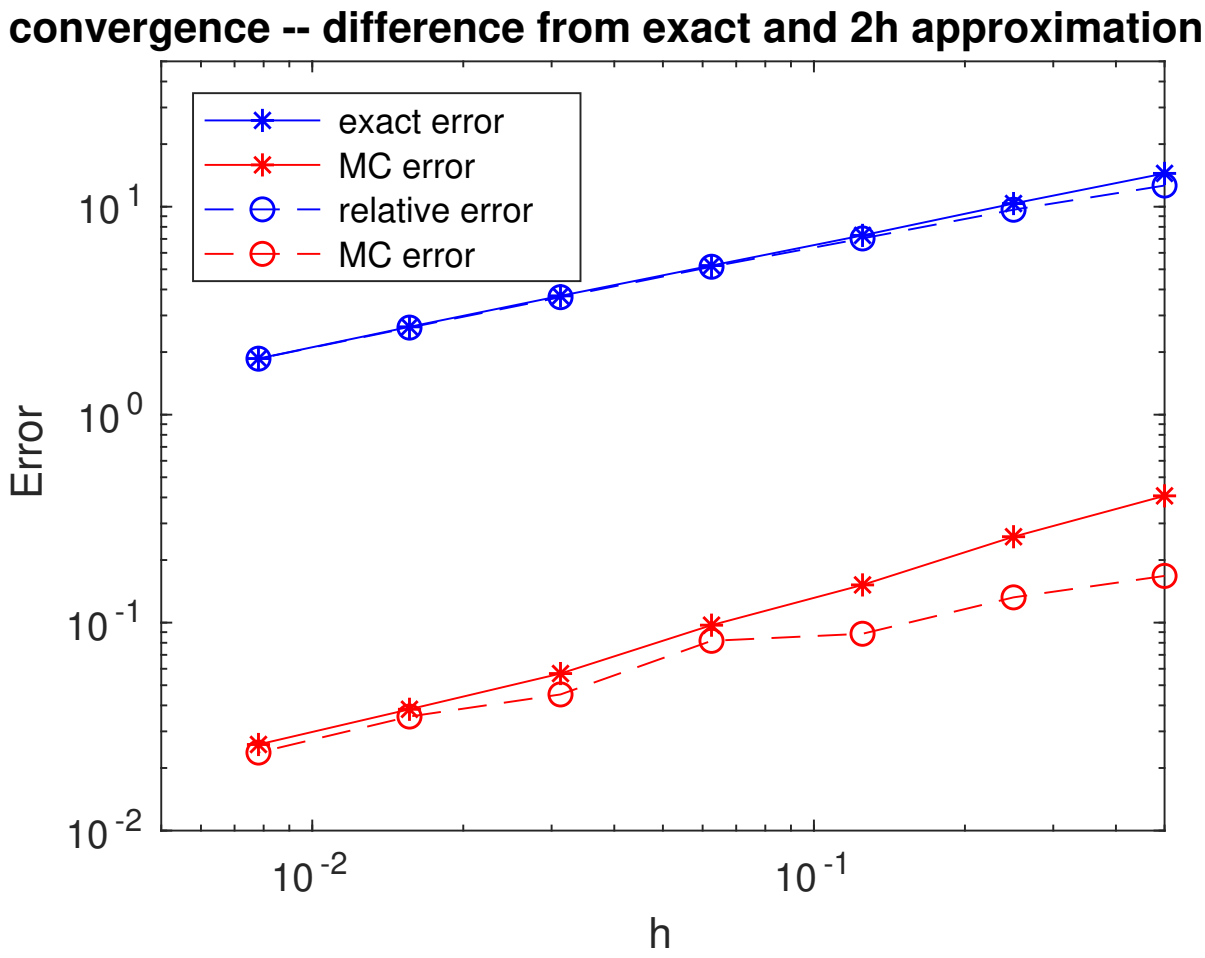
and the other showing the difference from the $2h$ approximation

Note that:

$$g(E+\delta) \approx g(E) + g'(E)\,\delta$$

so a confidence interval of $\delta = \pm 3\sigma/\sqrt{N}$ for $E \equiv \mathbb{E}\left[\Delta S^2\right]$, where $\sigma^2 \equiv \mathbb{V}\left[\Delta S^2\right]$, becomes a confidence interval of $\pm\frac{3}{2}\sigma/\sqrt{N\,E}$ for $\sqrt{E}$.

Strong convergence



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

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

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

Mean Square Error

If there are M timesteps, the computational cost is proportional to $C = NM$ and the MSE is approximately

$$a\,N^{-1} + b\,M^{-2} = a\,N^{-1} + b\,C^{-2}N^2.$$

For a fixed computational cost, this is a minimum when

$$N = \left(\frac{a\,C^2}{2\,b}\right)^{1/3}, \quad M = \left(\frac{2\,b\,C}{a}\right)^{1/3},$$

and hence

$$a\,N^{-1} = \left(\frac{2\,a^2 b}{C^2}\right)^{1/3}, \quad b\,M^{-2} = \left(\frac{a^2 b}{4\,C^2}\right)^{1/3},$$

so the MC term is twice as big as the bias term.

Extra bits – path-dependent functionals

The Bally & Talay result is for functions of the terminal state S_T.

For path-dependent output quantities of interest such as

$$f(S) = \max\left(S_T - \inf_{[0,T]} S_t,\ 0\right)$$

or

$$f(S) = \max(S_T - K, 0)\ \mathbf{1}_{\inf_{[0,T]} S_t > B}$$

then the most obvious approximation gives only $O(h^{1/2})$ weak convergence.

Fortunately first order convergence can be recovered (at least for scalar SDEs) through a Brownian Bridge treatment.

Extra bits – Milstein

There is a key result by Clark & Cameron (1980) that in general for multi-dimensional SDEs it is not possible to achieve better than $O(h^{1/2})$ strong convergence using just Brownian increments.

However, for scalar SDEs there is the first order Milstein approximation

$$\begin{aligned} \widehat{S}_{n+1} &= \widehat{S}_n + a(\widehat{S}_n, t_n)\, h + b(\widehat{S}_n, t_n)\, \Delta W_n \\ &\quad + \tfrac{1}{2}\, b(\widehat{S}_n, t_n)\, b'(\widehat{S}_n, t_n)\, (\Delta W_n^2 - h) \end{aligned}$$

which has a multi-dimensional generalisation under certain conditions.

I use this whenever I can, and for more general SDEs it is also the basis for a very effective MLMC treatment (antithetic Milstein).

Extra bits – adaptive time-stepping

For simplicity, I have presented only the simplest Euler-Maruyama approximation with uniform timesteps.

There are various circumstances under which it is good to adaptively choose the current timestep based on the current $\widehat{S}_n$ to maximise the overall accuracy for a given average cost.

This can also be important to ensure stability for SDEs with a drift which grows faster than linear

$$\mathrm{d}S_t = -S_t^3 \, \mathrm{d}t + \mathrm{d}W_t$$

Extra bits – discontinuous drift

Recent research by Müller-Gronbach and Yaroslavtseva (arXiv, 2018) proves that Euler-Maruyama also has $O(h^{1/2})$ strong convergence for SDEs with a discontinuous drift, such as

$$\mathrm{d}S_t = -\operatorname{sign}(S_t)\,\mathrm{d}t + \mathrm{d}W_t$$

There are also people researching what happens with discontinuous diffusion coefficients.

Extra bits – jump-diffusion

Some applications (particularly in finance) use jump-diffusion models

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

where $J(t)$ is the number of jumps which have taken place before time t, and the jump times are typically modelled as a simple Poisson process.

The Merton model is

$$\mathrm{d}S_t = r\,S_t\,\mathrm{d}t + \sigma\,S_t\,\mathrm{d}W_t + (k-1)S_t^-\,\mathrm{d}J_t$$

with jump times exponentially distributed with rate λ, and the jump magnitude k Normally distributed.

The numerical treatment of these is usually straightforward – you simulate the jumps at the jump times, and in between use Euler-Maruyama for the SDE.

Extra bits – Lévy processes

Jump-diffusion models have a finite rate of jumping.

A further generalisation is to Lévy processes in which there can be an infinite number of jumps in each time interval, but most are extremely small.

Exponential Lévy models have the form

$$S_t = \exp(L_t)$$

where L_t is a Lévy process, while Lévy driven processes have the form

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

Numerical methods often simulate the big jumps and approximate the small ones as a Brownian diffusion.

Extra bits – reflected diffusions

In 1D, the simplest reflected diffusion starting from $S_0 = x \geq 0$ is

$$S_t = W_t + x + L_t, \qquad L_t \equiv \max(-\inf_{[0,t]} W_s - x, 0)$$

which keeps $S_t \geq 0$.

The multi-dimensional generalisation of this gets more complex, and there is a distinction between normal and oblique reflections.

This class of problems is important in queueing theory, and there seems to be good potential for further research in this area.

Final Words

- simple Euler-Maruyama method is basis for most Monte Carlo simulation – $O(h)$ weak convergence and $O(\sqrt{h})$ strong convergence
- weak convergence is very important when estimating expectations
- strong convergence is usually not important – but is key for multilevel Monte Carlo method to be discussed next
- Mean-Square-Error is minimised by balancing bias due to weak error and Monte Carlo sampling error

Key references

P. Kloeden and E. Platen. "Numerical Solution of Stochastic Differential Equations". Springer, 1992.

D. Higham. "An algorithmic introduction to numerical simulation of stochastic differential equations". SIAM Review, 43(3):525-546, 2001.

P. Glasserman. "Monte Carlo Methods in Financial Engineering". Springer, 2003.

S. Asmussen, P.W. Glynn. "Stochastic Simulation: Algorithms and Analysis". Springer, 2007.