

IEOR E4703: Monte-Carlo Simulation

Simulating Stochastic Differential Equations

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Outline

The Euler Scheme

Weak and Strong Convergence

Other Discretization Schemes

Richardson Extrapolation

Some Examples From Finance

Improvements and Extensions

Change of Variables

Simulating Jump-Diffusion Processes

Variance Reduction Techniques

Allocation of Computational Resources

Extremes and Barrier Crossings

Multilevel Monte-Carlo

The Euler Scheme for Diffusions

Have an SDE of the form

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t. \quad (1)$$

Wish to simulate values of X_T but we don't know its distribution.

So simulate a **discretized** version of the SDE $\{\hat{X}_0, \hat{X}_h, \hat{X}_{2h}, \dots, \hat{X}_{mh}\}$ where:

- m is the number of time steps
- h is a constant step-size
- and $m = \lfloor T/h \rfloor$.

The simplest and most commonly used scheme is the **Euler** scheme:

$$\hat{X}_{kh} = \hat{X}_{(k-1)h} + \mu\left((k-1)h, \hat{X}_{(k-1)h}\right) h + \sigma\left((k-1)h, \hat{X}_{(k-1)h}\right) \sqrt{h} Z_k \quad (2)$$

where the Z_k 's are IID $N(0, 1)$.

The Euler Scheme for Diffusions

Note that even though we only care about X_T , we still need to generate intermediate values, X_{ih} , if we are to minimize the **discretization error**

- so simulating SDEs is computationally intensive
- because of discretization error, $\hat{\theta}_n$ is no longer an unbiased estimator of θ .

If we wished to estimate $\theta = \mathbb{E}[f(X_{t_1}, \dots, X_{t_p})]$ then in general we would need to keep track of $(X_{t_1}, \dots, X_{t_p})$.

Question: Can you think of a derivative where the payoff depends on $(X_{t_1}, \dots, X_{t_p})$, but where it would not be necessary to keep track of $(X_{t_1}, \dots, X_{t_p})$ on each sample path?

The Euler Scheme for Multidimensional Diffusions

In the multidimensional case, $\mathbf{X}_t \in \mathbb{R}^d$, $\mathbf{W}_t \in \mathbb{R}^p$ and $\boldsymbol{\mu}(t, \mathbf{X}_t) \in \mathbb{R}^d$ are now vectors, and $\boldsymbol{\sigma}(t, \mathbf{X}_t) \in \mathbb{R}^{d \times p}$ is a matrix.

Multidimensional case often occurs in applications:

1. Modeling the evolution of multiple stocks.
2. Modeling the evolution of a single stock in a **stochastic volatility** model.
3. Modeling the evolution of interest rates in **short rate**, **HJM** and **LIBOR market models**.

If the Brownian motions, \mathbf{W}_t , are correlated then can use the Cholesky decomposition.

But often the case that \mathbf{W}_t is standard (and therefore has independent components)

- any correlations between components of \mathbf{X}_t then induced through $\boldsymbol{\sigma}(t, \mathbf{X}_t)$.

Weak and Strong Convergence of Discretization Schemes

Two approaches for measuring the error in a discretization scheme:

1. A **strong** error criterion might take the form

$$\begin{aligned} & \mathbb{E} \left[\|\hat{X}_{mh} - X_T\| \right] \\ & \mathbb{E} \left[\sup_{0 \leq t \leq T} \|\hat{X}_{[t/h]h} - X_t\| \right] \end{aligned} \quad (3)$$

2. A **weak** error criterion takes the form

$$\left| \mathbb{E}[f(\hat{X}_{mh})] - \mathbb{E}[f(X_T)] \right| \quad (4)$$

where f ranges over “smooth” functions from \mathbb{R}^d to \mathbb{R} .

With a weak error criterion, only the **distribution** of \hat{X}_{mh} matters.

In finance applications we generally care about derivatives prices and so the weak criterion of (4) is more appropriate.

Given an error criterion, can assess the performance of the Euler scheme (and others) via its **order of convergence**.

Weak and Strong Convergence of Discretization Schemes

Definition. We say the discretization \hat{X} has a **strong order of convergence** of $\beta > 0$ if

$$\mathbb{E} \left[\|\hat{X}_{mh} - X_T\| \right] \leq ch^\beta \quad (5)$$

for some constant c and all sufficiently small h .

Definition. We say the discretization \hat{X} has a **weak order of convergence** of $\beta > 0$ if

$$\left| \mathbb{E}[f(\hat{X}_{mh})] - \mathbb{E}[f(X_T)] \right| \leq ch^\beta \quad (6)$$

for some constant c (possibly depending on f), all sufficiently small h , and all sufficiently smooth f .

Weak and Strong Convergence of Discretization Schemes

Note that a larger value of β in (5) and (6) is better.

In practice, often the case that a given discretization scheme will have a smaller strong order of convergence than its weak order of convergence.

e.g. The Euler scheme has a strong order of $\beta = 1/2$ but its weak order is $\beta = 1$
- but these orders of convergence require additional smoothness conditions on $\mu(t, \mathbf{X}_t)$ and σ .

Also worth noting that the conditions on f in weak order definition often not met in practice.

e.g. If f represents the payoff of a simple European call option, then f will not be differentiable and so f not sufficiently smooth.

Technical conditions on $\mu(t, \mathbf{X}_t)$ and σ are also sometimes violated in practice.

As a result, experimentation is often required to understand which schemes perform better for a given payoff f and / or SDE X_t .

The Milstein Scheme

A **scalar** SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

has Euler scheme

$$\hat{X}_{kh} = \hat{X}_{(k-1)h} + \mu(\hat{X}_{(k-1)h})h + \sigma(\hat{X}_{(k-1)h})\sqrt{h}Z_k.$$

Can apply Itô's Lemma to $\sigma(X_t)$ to construct a superior approximation for the diffusion term over the interval $[(k-1)h, kh]$.

This leads to “the” **Milstein** scheme

$$\hat{X}_{kh} = \hat{X}_{(k-1)h} + \mu(\hat{X}_{(k-1)h})h + \sigma(\hat{X}_{(k-1)h})\sqrt{h}Z_k + \frac{1}{2}\sigma'(\hat{X}_{(k-1)h})\sigma(\hat{X}_{(k-1)h})h(Z_k^2 - 1). \quad (7)$$

Approximation in (7) means that both drift and diffusion terms have both been expanded to $O(h)$.

Under various smoothness conditions (which again often do not hold in practice) it can be shown that the Milstein scheme has a weak and strong order of convergence of $\beta = 1$.

The Euler Scheme With Richardson Extrapolation

An alternative to **second order schemes** is the Euler scheme with **Richardson extrapolation**:

- it is easy to implement
- and often has superior performance to second order schemes, especially in high dimensions.

Euler scheme with Richardson extrapolation therefore often considered a benchmark scheme for reducing discretization error.

To simplify notation, we write \hat{X}_T^h for $\hat{X}_{\lfloor T/h \rfloor h}$.

First recall the Euler (often) has weak order 1 so that

$$\left| \mathbb{E}[f(\hat{X}_T^h)] - \mathbb{E}[f(X_T)] \right| \leq Ch. \quad (8)$$

Can sometimes strengthen (8) so that

$$\mathbb{E}[f(\hat{X}_T^h)] = \mathbb{E}[f(X_T)] + ch + o(h) \quad (9)$$

where c depends on f .

The Euler Scheme With Richardson Extrapolation

Then can apply (9) with discretization step $2h$ to obtain

$$\mathbb{E}[f(\hat{X}_T^{2h})] = \mathbb{E}[f(X_T)] + 2ch + o(h). \quad (10)$$

Can then combine estimators in (9) and (10) to eliminate the leading $O(h)$ term:

$$2\mathbb{E}[f(\hat{X}_T^h)] - \mathbb{E}[f(\hat{X}_T^{2h})] = \mathbb{E}[f(X_T)] + o(h). \quad (11)$$

Suggests an obvious improvement to the basic Euler scheme:

1. Simulate with time step h to estimate $\mathbb{E}[f(\hat{X}_T^h)]$
2. Simulate with time step $2h$ to estimate $\mathbb{E}[f(\hat{X}_T^{2h})]$
3. Double first estimate and subtract second to obtain an estimate of $\mathbb{E}[f(X_T)]$

The Euler Scheme With Richardson Extrapolation

Should use **consistent** Brownian increments in simulating paths of \hat{X}^h and \hat{X}^{2h}
- will typically result in an often substantial reduction in variance.

So if we use $\sqrt{h}Z_1, \sqrt{h}Z_2, \dots$ as Brownian increments for \hat{X}^h then can use $\sqrt{h}(Z_1 + Z_2), \sqrt{h}(Z_3 + Z_4), \dots$ as Brownian increments for \hat{X}^{2h} .

Using such a construction amounts to rewriting (11) as

$$\mathbb{E}[2f(\hat{X}_T^h) - f(\hat{X}_T^{2h})] = \mathbb{E}[f(X_T)] + o(h). \quad (12)$$

and then computing $2f(\hat{X}_T^h) - f(\hat{X}_T^{2h})$ along each sample path.

Variance of this estimator is

$$\text{Var}\left(2f(\hat{X}_T^h) - f(\hat{X}_T^{2h})\right) = 4\text{Var}\left(f(\hat{X}_T^h)\right) + \text{Var}\left(f(\hat{X}_T^{2h})\right) - 4\text{Cov}\left(f(\hat{X}_T^h), f(\hat{X}_T^{2h})\right).$$

A variance reduction will therefore be obtained if the covariance term is positive
- not always the case but can be guaranteed under monotonicity conditions.

Example: Option Pricing Under GBM

Consider pricing of a European call option in the Black-Scholes framework by simulating the SDE

$$dS_t = rS_t dt + \sigma S_t dW_t$$

with parameters $S_0 = K = 100$, $T = .5$ years, $r = .01$ and $\sigma = 0.4$.

Of course can price such an option using the Black-Scholes formula but it is of interest to see how well our discretization schemes perform.

Results were obtained by simulating 16 million sample paths.

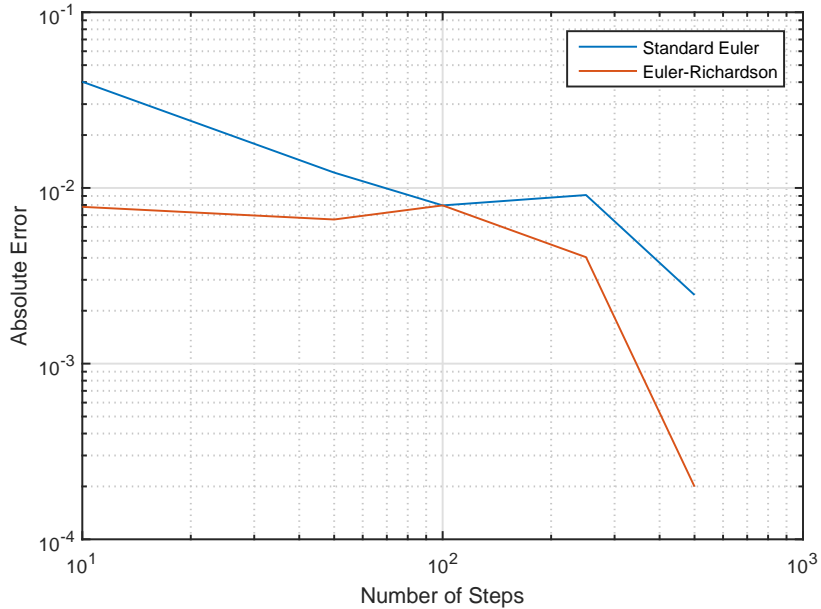
See that the absolute pricing error generally decreases as h decreases.

But occasionally see the error increase and this can largely be explained by the (unreported) **statistical error**

- even with 16m samples, approx. 95% CIs had a width of approx. 2 cents.

Also see superior performance of the Euler scheme with Richardson extrapolation kick in at about 500 time steps

- but due to the statistical noise this superiority may not be so clear on a different set of simulated paths.



Example: Option Pricing Under Heston

Consider Heston's stochastic volatility model:

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^{(1)} \quad (13)$$

$$dV_t = \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^{(2)}. \quad (14)$$

with $dW_t^{(1)} \times dW_t^{(2)} = \rho dt$.

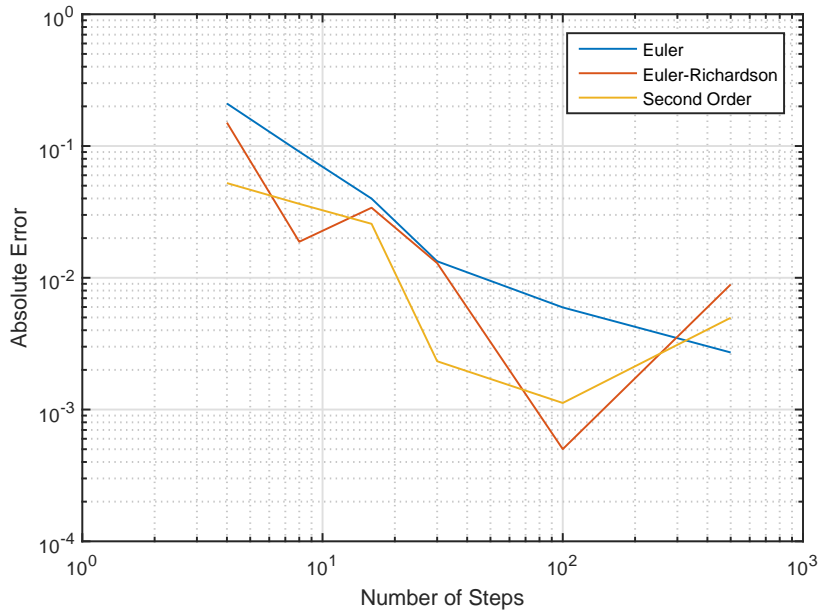
Again wish to price a European call option on the stock and we use the same parameters as those in Example 6.2.2 of Glasserman.

An explicit formula for the call option price is not available but can price it very accurately using **Fourier inversion methods**

- we find it's price to be 10.3009.

Can use this price to compare the absolute error of various discretization schemes as a function of the number of time steps.

Results are plotted in next figure with each point based on 8 million sample paths.



Example: Option Pricing Under Heston

Again see the general decrease in the mean absolute error of all three schemes as the number of time steps increases.

The various conditions (on both the option payoff and the SDE) that are required to guarantee a given order of convergence of the schemes are not satisfied here.

Moreover, even if the conditions were satisfied it may be the case that a very small value of the time-step h would be necessary before the stated order of convergence actually became apparent.

These observations and the (unreported) statistical error help explain the somewhat erratic convergence of the schemes and the apparently superior performance of the Euler scheme when 500 time-steps are employed.

This apparent “superior” performance can easily switch to an “inferior” performance with an alternative set of simulated sample paths.

Example: Option Pricing Under Heston

Also worth noting that the Euler scheme can perform **extremely poorly** in practice with Heston's stochastic volatility model.

e.g. Andersen considered pricing an ATM 10-year call option with $r = q = 0$, $S_0 = K = 100$, $\kappa = .5$, $V_0 = \theta = .04$, $\sigma = .1$ and $\rho = -0.9$
- the true option price is 23.69.

He used 1m sample paths and ia **“sticky zero”** or **“reflection”** assumption.

Time Steps	Sticky Zero	Reflection
100	28.3	45.1
200	27.1	41.3
500	25.6	37.1
1000	24.8	34.6

Euler scheme with reflection assumption converges very slowly!

Therefore need to be very careful when applying an Euler scheme to this SDE.

But common sense and some care should alert you to these problems and help resolve them!

Change of Variables

Given a discretization scheme, have considerable flexibility in choosing what process we apply it to.

More specifically, we can apply our scheme to $\mathbf{X}_t \in \mathbb{R}^d$ or to $\mathbf{Y}_t := g(\mathbf{X}_t)$ where $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a **smooth invertible** function.

If we apply it to \mathbf{Y}_t then $\hat{\mathbf{X}}_{kh} := g^{-1}(\hat{\mathbf{Y}}_{kh})$ is the corresponding discretized scheme for \mathbf{X}_t .

Often the case that a particular transformation seems intuitively appealing.

e.g If X_t represent a stock price then it makes sense (why?) to apply the scheme to $Y_t := \log(X_t)$ with $g^{-1}(\hat{Y}_{kh}) = \exp(\hat{Y}_{kh})$.

Change of Variables

Question: Characterize the discretization error that results from applying an Euler scheme to $\log(S_t)$ when $S_t \sim \text{GBM}$.

Question: Suppose we wish to simulate the known dynamics of a zero-coupon bond. How would you ensure that the simulated process satisfies $0 < Z_t^T < 1$?

An important advantage of this flexibility is that we can seek a g with a view to minimizing discretization error.

A common strategy is to choose a g (if possible) so that the dynamics of $\mathbf{Y}_t := g(\mathbf{X}_t)$ have a **constant volatility** coefficient.

Simulating Jump-Diffusion Processes

Consider a **jump-diffusion** process of the form

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t + c(X_{t-}, Y_{N_{t-}+1})dN_t \quad (15)$$

where:

- N_t is a Poisson process (independent of W_t) with parameter λ
- The Y_i 's are IID random variables independent of the Brownian motion W_t .

Note $X_{t-} := \lim_{u \uparrow t} X_u$ so if t is a jump time then X_{t-} is the value of the process **immediately before** t .

If the n^{th} jump in the Poisson process occurs at time t , then

$$X_t - X_{t-} = c(X_{t-}, Y_n)$$

If a jump does not occur at time t then $X_{t-} = X_t$.

Simulating Jump-Diffusion Processes

An obvious approach to simulating a discretized version of (15) on the interval $[0, T]$ is:

1. First simulate the arrival times in the Poisson process up to time T .
2. Use a pure diffusion discretization between the jump times.
3. At the n^{th} jump time τ_n , simulate the jump size $c(\hat{X}_{\tau_n-}, Y_n)$ conditional on the value of the discretized process, \hat{X}_{τ_n-} , immediately before τ_n .

Question: Suppose the process N_t in (15) is a more general **jump process** with stochastic intensity $\lambda(X_t)$.

If the intensity is bounded above by some constant $\bar{\lambda}$, how would you extend the scheme outlined above to this new process?

The Brownian Bridge and Stratified Sampling

Consider a short rate model of the form

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t.$$

When pricing a derivative that matures at time T using an Euler scheme it is necessary to generate the path $(W_h, W_{2h}, \dots, W_{mh} = W_T)$.

Will often be the case, however, that value of W_T will be particularly significant in determining the payoff.

Might want to **stratify** using W_T . This is easy since:

1. $W_T \sim N(0, T)$ so can easily generate a sample of W_T
and
2. Can easily generate $(W_h, W_{2h}, \dots, W_{T-h} \mid W_T)$.

The Brownian Bridge and Stratified Sampling

To see this note that for $s < t < v$

$$(W_t \mid W_s = x, W_v = y) \sim N\left(\frac{(v-t)x + (t-s)y}{v-s}, \frac{(v-t)(t-s)}{v-s}\right) \quad (16)$$

and can use this result to generate $(W_h \mid W_0, W_T)$.

More generally, can use (16) to successively simulate

$(W_h \mid W_0, W_T)$, $(W_{2h} \mid W_h, W_T)$, \dots , $(W_{T-h} \mid W_{T-2h}, W_T)$.

Can in fact simulate the points on the sample path in any order we like.

In particular, to simulate W_v we use (16) and condition on the two closest sample points before and after v , respectively, that have already been sampled.

This method of pinning the beginning and end points of the Brownian motion is known as a **Brownian bridge** construction.

Allocation of Computational Resources

Question: How should we choose $n = \#$ of sample paths and $m = \#$ of discretization points given a fixed computational budget?

- A smaller value of m will result in greater bias and numerical error
- A smaller value of n will result in greater statistical error.

Suppose then

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

and we wish to estimate $\theta := E[f(X_T)]$ using a scheme with weak order β .

Bias then satisfies

$$\text{Bias} \approx am^{-\beta}.$$

Suppose we have a fixed computational budget, C , and that each simulation step costs c – must therefore have $n = C/mc$.

Would like to choose optimal values of m (and therefore n) as a function of C .

Do this by minimizing the **mean squared error** (MSE), i.e. sum of the bias squared and the variance v/n .

Allocation of Computational Resources

We have

$$\text{MSE} \approx \frac{a^2}{m^{2\beta}} + \frac{v}{n} \quad (17)$$

for sufficiently large m .

Substituting for n in (17), easy to see that's optimal to take

$$m \propto C^{1/(2\beta+1)} \quad (18)$$

$$n \propto C^{2\beta/(2\beta+1)} \quad (19)$$

with the optimal $\text{MSE} \approx C^{-\frac{2\beta}{2\beta+1}}$.

Note that the $\text{RMSE} \approx C^{-\frac{\beta}{2\beta+1}} \rightarrow C^{-1/2}$ as $\beta \rightarrow \infty$, which is (why?) as expected.

Allocation of Computational Resources

When it comes to estimating θ , (18) and (19) provide guidance as follows.

Suppose we are using an Euler scheme with $\beta = 1$.

Begin by using n_0 paths and m_0 points per path to compute initial estimate, $\hat{\theta}_0$.

If we then compute a new estimate, $\hat{\theta}_1$, by setting $m_1 = 2m_0$, then (18) and (19) suggest we should set $n_1 = 4n_0$.

May then continue to compute new estimates, $\hat{\theta}_i$, in this manner until the estimates and their associated CI's converge.

In general, if we increase m by a factor of 2 then should increase n by a factor of 4.

Although estimating θ in this way requires additional computational resources, it is not usually necessary to perform more than two or three iterations, provided we begin with sufficiently large values of m_0 and n_0 .

Extremes and Barrier Crossings

Suppose X_t is a standard Brownian motion and let $M_t := \max_{0 \leq u \leq t} X_u$ denote the **running maximum**.

Then maximum of corresponding Euler process given by

$$\hat{M}_m^h := \max \left\{ X_0, \hat{X}_h, \hat{X}_{2h}, \dots, \hat{X}_{mh} \right\}. \quad (20)$$

Can be shown that the weak order of convergence of this discretization scheme for M_t cannot be better than $1/2$

- even though the Euler scheme for X_t is exact
- and has a weak order of convergence equal to 1.

Can resolve this by simulating M_T directly for any value of T . We do this by:

1. Simulating $X_T \sim N(0, T)$
2. Simulating $M_T | X_T$. This can be done because it is known that

$$M_T | X_T \sim \frac{X_T + \sqrt{X_T^2 - 2T \log U}}{2} \quad (21)$$

where $U \sim U(0, 1)$ (independent of X_T).

Extremes and Barrier Crossings

This procedure can easily be adapted to handle more general processes.

Let \hat{X}_{kh} for $k = 0, 1, \dots$ be a discretization scheme for X_t satisfying

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t \quad (22)$$

We interpolate over each interval $[kh, (k+1)h]$ by using a **Brownian bridge** with fixed parameters $\mu(kh, \hat{X}_{kh})$ and $\sigma_k := \sigma(kh, \hat{X}_{kh})$.

So given endpoints \hat{X}_{kh} and $\hat{X}_{(k+1)h}$, the max of the process on $[kh, (k+1)h]$ can be simulated as

$$\hat{M}_k = \frac{\hat{X}_{(k+1)h} + \hat{X}_{kh} + \sqrt{\left(\hat{X}_{(k+1)h} - \hat{X}_{kh}\right)^2 - 2h\sigma_k^2 \log U_i}}{2} \quad (23)$$

where the U_i 's are IID $U(0, 1)$ random variables.

The maximum of X over $[0, T]$ can then be approximated using

$$\max \left\{ \hat{M}_0, \hat{M}_h, \hat{M}_2, \dots, \hat{M}_{m-1} \right\}.$$

Extremes and Barrier Crossings

Same ideas can be immediately applied to the pricing of barrier options.

e.g. Suppose we wish to price a knock-out put option with time T payoff

$$(K - X_T)^+ 1_{\{\tau > T\}}$$

where $\tau = \inf\{t \geq 0 : X_t > B\}$ and with $X_0 < B$.

Simplest approach would be to approximate τ with $\hat{\tau}$ where

$$\hat{\tau} := \inf\{k : \hat{X}_{kh} > B\}.$$

But can do much better by using the construction in (23).

We note that barrier is crossed in the interval $[kh, (k+1)h]$ if the maximum of the process in that interval exceeds B .

Can thus approximate the option payoff with

$$(K - \hat{X}_{mh})^+ \prod_{k=0}^{m-1} 1_{\{\hat{M}_k \leq B\}} \quad (24)$$

with \hat{M}_k generated as in (23) and $mh = T$.

Extremes and Barrier Crossings

Question: Can simplify approximation of survival indicator $1_{\{\tau > T\}}$ in (24) with

$$\prod_{k=0}^{m-1} 1_{\{U_k \leq \hat{p}_k\}}. \quad (25)$$

Provide an expression for \hat{p}_k in terms of B , \hat{X}_{kh} and $\hat{X}_{(k+1)h}$.

Question: Explain how this leads to a superior estimator of the form

$$(K - \hat{X}_{mh})^+ \prod_{k=0}^{m-1} \hat{p}_k.$$

In what sense is this estimator superior to the estimator in (24)?

Is there any sense in which the estimator might be inferior?

Multilevel Monte-Carlo

Multilevel Monte-Carlo is a recently developed approach that optimizes the allocation of computational resources to minimize the estimator's MSE.

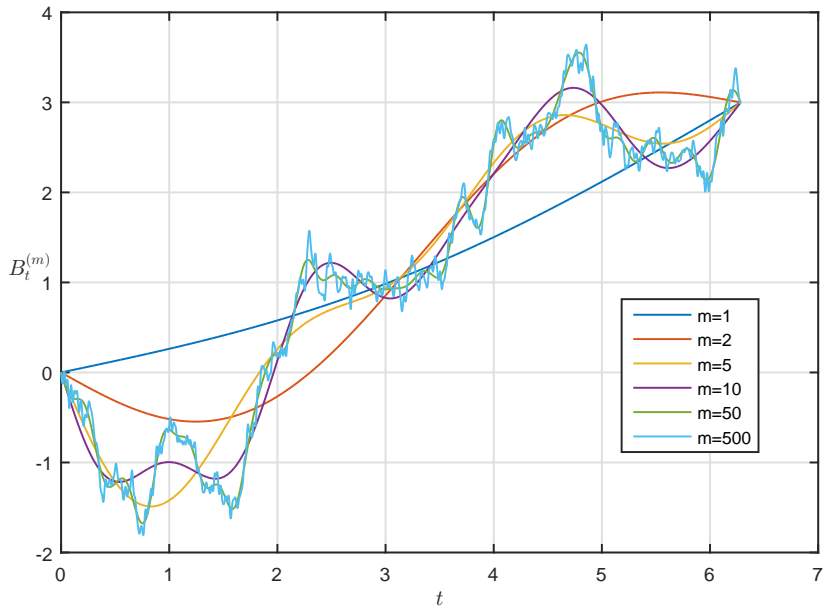
Can motivate the technique by considering the **Paley-Wiener representation** of Brownian motion on the interval $[0, 2\pi]$:

$$W_t = Z_0 \frac{t}{\sqrt{2\pi}} + \frac{2}{\sqrt{\pi}} \sum_{n=1}^{\infty} Z_n \frac{\sin\left(\frac{nt}{2}\right)}{n} \quad (26)$$

where the Z_i 's are IID $N(0, 1)$.

This representation suggests an obvious approximation to W_t :

$$W_t^{(m)} = Z_0 \frac{t}{\sqrt{2\pi}} + \frac{2}{\sqrt{\pi}} \sum_{n=1}^m Z_n \frac{\sin\left(\frac{nt}{2}\right)}{n} \quad (27)$$



Multilevel Monte-Carlo

Should be clear that earlier terms in the series determine the overall shape of the Brownian path while the later terms add the finer details and improve the resolution of the approximation.

Typically the case that earlier terms are more important for determining the quantity of interest, e.g. the payoff of an option.

Hence makes sense that we could construct a superior estimator by focusing more effort on simulating the Z_i 's for small values of i rather than large values of i

- essentially the insight used by the multilevel method.