Strong convergence of the Euler–Maruyama method

In this section, we present a simple proof of strong convergence for the Euler–Maruyama method applied to the autonomous equation

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s, \qquad X_0 = x_0. \tag{1}$$

We will assume throughout that the following conditions, which guarantee the existence of a unique strong solution to (1), are satisfied:

Assumption 1. *The coefficients of* (1) *are globally Lipschitz and that they satisfy a linear growth condition: there exists K such that for all $x, x' \in \mathbb{R}$*

$$|b(x) - b(x')| + |\sigma(x) - \sigma(x')| \leq K|x - x'| \qquad \textit{and} \qquad |b(x)| + |\sigma(x)| \leq K(1 + |x|).$$

In addition, x_0 is independent of the Brownian motion W and $\mathbb{E}[x_0^2] < \infty$.

We mentioned in class that it was often convenient to define a continuous-time process $\{\hat{X}_t^{\Delta t}\}_{t\in[0,T]}$ from a discrete-time approximation $\{X_n^{\Delta t}\}_{n=0}^N$ obtained by the Euler–Maruyama method. (Here, by continuous-time process, we mean a process indexed by $t \in [0, T]$, not a process with continuous sample paths.) Some standard ways of defining a continuous-time approximation are the following:

Piecewise constant solution. For $t_i \leq t < t_{i+1}$, set $\hat{X}_t^{\Delta t} = X_i^{\Delta t}$. With this definition, notice that $\hat{X}_t^{\Delta t}$ satisfies

$$\mathrm{d}\hat{X}_t^{\Delta t} = x_0 + \int_0^{t_{n_t}} b(\hat{X}_s^{\Delta t})\,\mathrm{d}s + \int_0^{t_{n_t}} \sigma(\hat{X}_s^{\Delta t})\,\mathrm{d}W_s, \qquad n_t = \left\lfloor \frac{t}{\Delta t} \right\rfloor, \tag{2}$$

which has a structure similar to that of (1).

Piecewise constant drift and diffusion. For $t_i \leq t < t_{i+1}$, set

$$\tilde{X}_t^{\Delta t} = X_i^{\Delta t} + b(X_i^{\Delta t})\,(t - t_i) + \sigma(X_i^{\Delta t})\,(W_t - W_{t_i}).$$

With this definition, $\tilde{X}_t^{\Delta t}$ satisfies

$$\mathrm{d}\tilde{X}_t^{\Delta t} = x_0 + \int_0^t b(\tilde{X}_{n_s\Delta t}^{\Delta t})\,\mathrm{d}s + \int_0^t \sigma(\tilde{X}_{n_s\Delta t}^{\Delta t})\,\mathrm{d}W_s. \tag{3}$$

Recall also (if you have had a course on stochastic differential equations that the proof of existence and uniqueness of a strong solution to (1) relies on a fixed-point iteration in the Banach space of stochastic processes Y satisfying

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |Y_t|^2\right] < \infty.$$

It is therefore not surprising that the proof yields, as a byproduct, an estimate of the form

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |X_t|^2\right] < C(K, T, \mathbb{E}|x_0|^2), \tag{4}$$

which we will take for granted below. We are now ready to prove the strong convergence of the Euler–Maruyama scheme.

Theorem 1 (Strong convergence of the Euler–Maruyama method)**.** *Let $\hat{X}_t^{\Delta t}$ be the piecewise constant solution, as defined in* (2) *above, associated with the Euler–Maruyama method with time step Δt. Under Assumption 1, there exists $C = C(T, K, \mathbb{E}|X_0|^2)$ such that*

$$\sup_{0\le t\le T} \mathbb{E}|X_t - \hat{X}_t^{\Delta t}| \le C\sqrt{\Delta t}.$$

Proof of Theorem 1. To simplify the notations, we will denote the continuous-time Euler–Maruyama solution $\hat{X}_t^{\Delta t}$ by just $\hat{X}_t$ (without the Δt superscript). We will also denote by n_t the index of the discretization point less than t closest to t, that is $n_t = \lfloor t/\Delta t \rfloor$ (and similarly for n_s).

For $0 \le t \le T$, let us define

$$Z(t) = \sup_{0\le s\le t} \mathbb{E}\left[|X_s - \hat{X}_s|^2\right]$$

Substituting the exact and approximate solutions in the expression $Z(t)$, we obtain

$$\begin{aligned}
Z(t) &= \sup_{0\le s\le t} \mathbb{E}\left|\int_0^{t_{n_s}} b(X_u) - b(\hat{X}_u)\,\mathrm{d}u + \int_0^{t_{n_s}} \sigma(X_u) - \sigma(\hat{X}_u)\,\mathrm{d}W_u \right.\\
&\qquad \left. + \int_{t_{n_s}}^s b(X_u)\,\mathrm{d}u + \int_{t_{n_s}}^s \sigma(X_u)\,\mathrm{d}W_u\right|^2\\
&\le 4 \sup_{0\le s\le t} \mathbb{E}\left[\left|\int_0^{t_{n_s}} b(X_u) - b(\hat{X}_u)\,\mathrm{d}u\right|^2 + \left|\int_0^{t_{n_s}} \sigma(X_u) - \sigma(\hat{X}_u)\,\mathrm{d}W_u\right|^2\right.\\
&\qquad \left. + \left|\int_{t_{n_s}}^s b(X_u)\,\mathrm{d}u\right|^2 + \left|\int_{t_{n_s}}^s \sigma(X_u)\,\mathrm{d}W_u\right|^2\right].
\end{aligned}$$

Using the linearity of the expectation, Cauchy–Schwarz's inequality and Itô's isometry, we obtain

$$\begin{aligned}
Z(t) \le 4 \sup_{0\le s\le t} &\left(T\,\mathbb{E}\left[\int_0^{t_{n_s}} |b(X_u) - b(\hat{X}_u)|^2\,\mathrm{d}u\right] + \mathbb{E}\left[\int_0^{t_{n_s}} |\sigma(X_u) - \sigma(\hat{X}_u)|^2\,\mathrm{d}u\right]\right.\\
&\left. +\ \Delta t\,\mathbb{E}\left[\int_{t_{n_s}}^s |b(X_u)|^2\,\mathrm{d}u\right] + \mathbb{E}\left[\int_{t_{n_s}}^s |\sigma(X_u)|^2\,\mathrm{d}s\right]\right).
\end{aligned}$$

Using the global Lipschitz and linear growth assumptions on the coefficients, we obtain

$$\begin{aligned}
Z(t) &\le 4 \sup_{0\le s\le t} \left(K^2\,(T+1)\,\mathbb{E}\left[\int_0^{t_{n_s}} |X_u - \hat{X}_u|^2\,\mathrm{d}u\right] + 2\,K^2(\Delta t + 1)\,\mathbb{E}\left[\int_{t_{n_s}}^s (1+|X_u|)^2\,\mathrm{d}u\right]\right)\\
&\le 4 \sup_{0\le s\le t} \left(K^2\,(T+1) \int_0^{t_{n_s}} Z(u)\,\mathrm{d}u + 2\,K^2(\Delta t+1)\left(\Delta t + \int_{t_{n_s}}^s \mathbb{E}[|X_u|^2]\,\mathrm{d}u\right)\right)\\
&\le 4\left(K^2\,(T+1)\int_0^t Z(u)\,\mathrm{d}u + 2\,K^2\,\Delta t\,(\Delta t + 1)\left(1 + \sup_{0\le t\le T} \mathbb{E}[|X_t|^2]\right)\right).
\end{aligned}$$

Employing (4), and denoting by C constants depending only on K, T and $\mathbb{E}|x_0|^2$, we deduce

$$Z(t) \leq C \left(\int_0^t Z(u) \, \mathrm{d}u + \Delta t \right),$$

and so, by Grönwall's inequality, it follows that

$$Z(T) = \sup_{0 \leq s \leq T} \mathbb{E} \left[|X_s - \hat{X}_s|^2 \right] \leq C \, \Delta t,$$

which leads to the conclusion by using Jensen's (or just the Cauchy–Schwarz) inequality. □

Remark 1 (Not covered). In this result we worked with the following metric of the strong error,

$$\varepsilon_{\Delta t} = \sup_{0 \leq t \leq T} \mathbb{E}|X_t - \hat{X}_t^{\Delta t}|,$$

but note that this is one of several possibilities. In particular,

1. Some results provide bounds for the strong error only at the discretization points. This is the case, in particular, for many error estimates associated with higher-order schemes. The piecewise constant interpolated solution we use here, for example, cannot be expected to converge to the exact solution with strong order more than 1/2 (or 1 if the diffusion coefficient is equal to 0). To see this, consider the equation

$$\mathrm{d}X_t = \mathrm{d}W_t.$$

   This equation is integrated exactly (but only at the discrete times $t_0, \ldots, t_N$) by numerical schemes for SDEs, and the piecewise constant interpolated solution (2) is simply $X_t = W_{n_t \Delta t}$. Consequently $\mathbb{E}|X_t - \hat{X}_t^{\Delta t}| = \mathbb{E}|W_t - W_{n_t \Delta t}| = (\sqrt{2}/\pi) \sqrt{t - n_t \Delta t}$, by the properties of the folded normal distribution, and the supremum in the strong error employed in Theorem 1 is therefore $(\sqrt{2}/\pi)\sqrt{\Delta t}$.

2. Some results employ a different interpolated solution, such as the one defined in (3), corresponding to constant drift and diffusion coefficients over each interval $[t_i, t_{i+1})$,

3. Some results employ an even stronger metric for the strong error than the one we use. In particular, it is possible to show that the order of convergence is 1/2 also when the strong error $\varepsilon_{\Delta t}$ is defined by

$$\varepsilon_{\Delta t}^2 = \mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t - \tilde{X}_t^{\Delta t}|^2 \right].$$

   Here $\tilde{X}_t^{\Delta t}$ is the interpolated solution given by (3). To see that this is a stronger metric, notice that

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t - \tilde{X}_t^{\Delta t}|^2 \right] \geq \sup_{0 \leq t \leq T} \mathbb{E}|X_t - \tilde{X}_t^{\Delta t}|^2$$

   and, by Jensen's inequality, $\mathbb{E}|X_t - \tilde{X}_t^{\Delta t}|^2 \geq (\mathbb{E}|X_t - \tilde{X}_t^{\Delta t}|)^2$.

⊘

Weak convergence of the Euler–Maruyama method

In order to prove the weak convergence, we will rely on the fundamental relation, established by the Feynman–Kac formula below, between (1) and the following parabolic partial differential equation (PDE), known as the *backward Kolmogorov equation* associated to (1):

$$\begin{cases} \partial_t u(t,x) + \mathcal{L} u(t,x) = 0, & (t,x) \in [0,T) \times \mathbb{R} \\ u(T,x) = f(x), & x \in \mathbb{R}. \end{cases} \tag{5}$$

Here the operator $\mathcal{L}$ is the generator of the Markov semigroup associated with (1):

$$\mathcal{L} = b(x)\, \partial_x + \frac{1}{2}\, \sigma(x)^2\, \partial_x^2.$$

If you are not familiar with Markov semigroups, do not worry: here it is sufficient to view $\mathcal{L}$ as just an operator that is useful for our purposes. Note that (5), unlike most parabolic PDEs in physics, is paired with a condition at the final time, a *terminal condition.*

Theorem 2 (Feyman–Kac formula)**.** *Assume that f, b and σ are such that the solution $u(t,x)$ to the backward Kolmogorov equation* (5) *exists and satisfies the assumption of Itô's formula, i.e. that $u \in C^{1,2}$. Then u admits the representation*

$$u(t,x) = \mathbb{E}[f(X_T^{t,x})],$$

where $X_s^{t,x}$ denotes the solution of

$$X_s^{t,x} = x + \int_t^s b(X_u)\, \mathrm{d}u + \int_t^s \sigma(X_u)\, \mathrm{d}W_u, \qquad t \leq s \leq T.$$

Proof. Employing Itô's formula,

$$u(T, X_T^{t,x}) = u(t,x) + \int_t^T \partial_t u(s, X_s^{t,x}) + \mathcal{L} u(s, X_s^{t,x})\, \mathrm{d}s + \int_t^T \sigma(X_s^{t,x})\, \partial_x u(s, X_s^{t,x})\, \mathrm{d}W_s.$$

Using the fact that u solves (5) and that the Itô integral is a martingale, we obtain after taking expectations

$$\mathbb{E}[f(X_T^{t,x})] = u(t,x),$$

which concludes the proof. □

Remark 2. In the case of (1), where the drift and diffusion coefficients do not depend explicitly on time, it is often more convenient to consider the following initial-value problem instead of (5):

$$\begin{cases} \partial_t v(t,x) - \mathcal{L} v(t,x) = 0, & (t,x) \in (0,T] \times \mathbb{R} \\ v(0,x) = f(x), & x \in \mathbb{R}. \end{cases} \tag{6}$$

Notice that the solution v to this equation is related to that of (5) by $v(t,x) = u(T-t,x)$, as can be shown by an application of the chain rule. Consequently, v admits the representation

$$v(t,x) = \mathbb{E}[f(X_T^{x,T-t})] = \mathbb{E}[f(X_t^{x,0})]$$

because, in view of the fact that the coefficients $b(\cdot)$ and $\sigma(\cdot)$ do not depend explicitly on time, the processes $u \mapsto X^{x,s}_{s+u}$ and $u \mapsto X^{x,0}_u$ have the same law regardless of the value of s. $\oslash$

Much of the difficulty in proving a general weak convergence result for the Euler–Maruyama scheme lies in showing that the solution to (5) (or to (6) in the autonomous case) has good regularity properties. In order to focus on the part of the proof that is most interesting for our purposes in this course, we will make very strong additional assumptions on the coefficients to ensure that this is the case with as little work as possible.

Proposition 3 (Cauchy problem on the torus – not examinable)**.** *Assume that $b(\cdot)$, $\sigma(\cdot)$ and $f(\cdot)$ are smooth functions on the torus (i.e. smooth periodic functions) and that $\sigma(\cdot)$ is bounded from below uniformly by a positive constant. Then* (5) *admits a unique smooth classical solution.*

Proof. This follows from standard PDE theory (Lax–Milgram theorem, Fredholm alternative, spectral theorem for compact self-adjoint operators, etc.). $\square$

Theorem 4 (Weak convergence of the Euler–Maruyama method)**.** *Under the assumptions of Proposition 3, there exist $K > 0$ independent of Δt such that*

$$|\mathbb{E}[f(X^{\Delta t}_N)] - \mathbb{E}[f(X_T)]| \leq K\,\Delta t.$$

Proof. Let us denote by u the solution to (5). By the Feynman–Kac formula,

$$\begin{aligned}\mathbb{E}[f(X^{\Delta t}_N)] - \mathbb{E}[f(X_T)] &= \mathbb{E}\left[u(T, X^{\Delta t}_N) - u(0, x_0)\right] \\ &= \mathbb{E}\left[\sum_{i=0}^{N-1}\left(u(t_{i+1}, X^{\Delta t}_{i+1}) - u(t_i, X^{\Delta t}_i)\right)\right] =: \sum_{i=0}^{N-1}\mathbb{E}[e_i]. \qquad (7)\end{aligned}$$

Let $\tilde{X}_t$ be the interpolated Euler–Maruyama solution defined by (3). Since $\tilde{X}_t$ coincides with $X^{\Delta t}_n$ at the discretization points, we deduce from Itô's formula that

$$\mathbb{E}[e_i] = \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \partial_t u(t, \tilde{X}_t) + \left(b(\tilde{X}_{t_i})\,\partial_x + \frac{1}{2}\sigma(\tilde{X}_{t_i})^2\,\partial_x^2\right) u(t, \tilde{X}_t)\,\mathrm{d}t\right].$$

Notice that the operator in the second term of the integrand is the generator of (3). Since u is the solution to (5), $(\partial_t + \mathcal{L})u(t_i, \tilde{X}_{t_i}) = 0$, and so

$$\begin{aligned}\mathbb{E}[e_i] = \mathbb{E}\Bigg[\int_{t_i}^{t_{i+1}} &\left(\partial_t u(t, \tilde{X}_t) - \partial_t u(t_i, \tilde{X}_{t_i})\right) + b(\tilde{X}_{t_i})\left(\partial_x u(t, \tilde{X}_t) - \partial_x u(t_i, \tilde{X}_{t_i})\right) && (8a)\\ &+ \frac{1}{2}\sigma(\tilde{X}_{t_i})^2\left(\partial_x^2 u(t, \tilde{X}_t) - \partial_x^2 u(t_i, \tilde{X}_{t_i})\right)\,\mathrm{d}t\Bigg] && (8b)\\ =: \mathbb{E}&\left[\sum_{j=1}^{3} h_j(\tilde{X}_{t_i})\int_{t_i}^{t_{i+1}} g_j(t, \tilde{X}_t) - g_i(t_i, \tilde{X}_{t_i})\,\mathrm{d}t\right] =: \sum_{i=1}^{3}\mathbb{E}[e_{ij}]. && (8c)\end{aligned}$$

where we introduced

$$h_1(x) = 1, \qquad h_2(x) = b(x), \qquad h_3(x) = \frac{1}{2}\,\sigma(x)^2, \qquad g_1 = \partial_t u, \qquad g_2 = \partial_x u, \qquad g_3 = \partial_x^2 u.$$

By Itô's formula, it holds for $j = 1, 2, 3$ that

$$g_j(t, \tilde{X}_t) - g_j(t_i, \tilde{X}_{t_i}) = \left[\int_{t_i}^{t} \partial_t g_j(s, \tilde{X}_s) + \left(b(\tilde{X}_{t_i})\,\partial_x + \frac{1}{2}\sigma(\tilde{X}_{t_i})^2\,\partial_x^2\right) g_j(s, \tilde{X}_s)\,\mathrm{d}s \right.$$
$$\left. + \int_{t_i}^{t} \sigma(\tilde{X}_{t_i})\,\partial_x g_j(s, \tilde{X}_s)\,\mathrm{d}W_s\right].$$

Using the law of total expectation and the fact that the Itô integral is a martingale, we observe that the Itô integral does not contribute to the expectation:

$$\begin{aligned}
\mathbb{E}\left[e_{ij}\right] &= \mathbb{E}\left[\mathbb{E}\left[e_{ij}|\tilde{X}_{t_i}\right]\right] = \mathbb{E}\left[h_j(\tilde{X}_{t_i})\,\mathbb{E}\left[\int_{t_i}^{t_{i+1}} g_j(t, \tilde{X}_t) - g_j(t_i, \tilde{X}_{t_i})\,\mathrm{d}t \,|\, \tilde{X}_{t_i}\right]\right] \\
&= \mathbb{E}\left[h_j(\tilde{X}_{t_i})\,\mathbb{E}\left[\int_{t_i}^{t_{i+1}}\int_{t_i}^{t} \partial_t g_j(s, \tilde{X}_s) + \left(b(\tilde{X}_{t_i})\,\partial_x + \frac{1}{2}\sigma(\tilde{X}_{t_i})^2\,\partial_x^2\right) g_j(s, \tilde{X}_s)\,\mathrm{d}s\,\mathrm{d}t \,|\, \tilde{X}_{t_i}\right]\right], \\
&= \mathbb{E}\left[h_j(\tilde{X}_{t_i})\,\int_{t_i}^{t_{i+1}}\int_{t_i}^{t} \partial_t g_j(s, \tilde{X}_s) + \left(b(\tilde{X}_{t_i})\,\partial_x + \frac{1}{2}\sigma(\tilde{X}_{t_i})^2\,\partial_x^2\right) g_j(s, \tilde{X}_s)\,\mathrm{d}s\,\mathrm{d}t\right].
\end{aligned}$$

By Proposition 3, $u(\cdot)$ is a smooth function on the compact set $[0, T] \times \mathbb{T}$, where $\mathbb{T}$ denotes the torus. Therefore, together with all their derivatives in space and time, the functions g_j are uniformly bounded from below and from above. The functions h_j are also uniformly bounded from below and from above independently of Δt, because they are periodic by assumption. It follows from these considerations that

$$|\mathbb{E}\left[e_{ij}\right]| \leq \int_{t_i}^{t_{i+1}}\int_{t_i}^{t} C\mathrm{d}s\,\mathrm{d}t = \frac{1}{2}\,C\Delta t^2, \qquad i = 1, 2, 3,$$

where C is a constant independent of Δt. We deduce, going back to (7),

$$|\mathbb{E}[f(X_N^{\Delta t})] - \mathbb{E}[f(X_T)]| = \left|\sum_{i=0}^{N-1}\sum_{j=1}^{3}\mathbb{E}[e_{ij}]\right| \leq \sum_{i=0}^{N-1}\sum_{j=1}^{3}|\mathbb{E}[e_{ij}]| \leq \frac{3}{2}\,C\,\Delta t,$$

which concludes the proof. □

Remark 3 (Not covered). The assumptions of Theorem 4 are very restrictive: as stated, the theorem applies only to SDEs with state space $\mathbb{T}$. When the state space of the SDE is $\mathbb{R}$, it can be shown at the cost of substantial additional work that, if $b, \sigma \in C_b^4$ and $f \in C_p^4$, then the solution to the backward Kolmogorov equation is sufficiently regular for the proof outlined above to go through, although more advanced arguments need to be employed to bound the terms in (8). Here C_b^ℓ (resp. C_p^ℓ) denotes the subspace of C^ℓ consisting of functions which, together with their derivatives of order up to ℓ, are bounded (resp. grow at most polynomially). This result is an improvement upon the one we proved but it is still not completely satisfactory: indeed, simple SDEs such as the Ornstein–Uhlenbeck process, for which the drift coefficient is unbounded, are still not covered. For more general results, the interested reader can refer to [1] or [2]. ⊘

Exercise 1. In the case of geometric Brownian and for $f(x) = x$, show using a more direct method that the weak error for the Euler–Maruyama scheme scales as $\mathcal{O}(\Delta t)$, i.e. that there exists C independent of Δt such that, for any sufficiently small Δt,

$$\left|\mathbb{E}\left[X_{N\Delta t} - X_N^{\Delta t}\right]\right| \leq C\Delta t.$$

Solution. We use the following parametrization of the geometric Brownian motion:

$$\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}W_t, \qquad X_0 = x_0 > 0, \tag{9}$$

where μ and σ are constants and x_0 is deterministic. The solution to this SDE (which we found earlier in the course by applying Itô's formula to the function $\ln X_t$) is

$$X_t = x_0 \, \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right).$$

The Euler-Maruyama discretisation of (9), denoted by $X_n^{\Delta t}$, is obtained from the iteration

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + \mu \, X_n^{\Delta t} \, \Delta t + \sigma \, X_n^{\Delta t} \, \Delta W_n = (1 + \mu \, \Delta t + \sigma \, \Delta W_n) \, X_n^{\Delta t},$$

where $\Delta W_n = W_{(n+1)\Delta t} - W_{n\Delta t}$. It follows that

$$X_N^{\Delta t} = x_0 \prod_{n=0}^{N-1} \left(1 + \mu \Delta t + \sigma \Delta W_n\right).$$

Since all the factors are independent,

$$\mathbb{E}[X_N^{\Delta t}] = x_0 \prod_{n=0}^{N-1} \mathbb{E}[1 + \mu \, \Delta t + \sigma \Delta W_n] = x_0 \prod_{n=0}^{N-1} (1 + \mu \, \Delta t).$$

On the other hand, by the properties of the lognormal distribution,

$$\mathbb{E}[X_{N\Delta t}] = x_0 \, \mathrm{e}^{\mu \, T}.$$

Therefore

$$\left|\mathbb{E}\left[X_{N\Delta t} - X_N^{\Delta t}\right]\right| = \left|\mathrm{e}^{\mu T} - \left(1 + \frac{\mu \, T}{N}\right)^N\right|. \tag{10}$$

Employing the first of the well-known characterizations of the exponential function,

$$\lim_{N \to \infty} \left(1 + \frac{\mu \, T}{N}\right)^N = \mathrm{e}^{\mu T},$$

we deduce convergence of the weak error to zero. To obtain the rate of convergence to zero, let introduce $x := \mu T$ and rewrite, by Taylor's formula with remainder,

$$\mathrm{e}^x = \left(\mathrm{e}^{\frac{x}{N}}\right)^N = \left(1 + \frac{x}{N} + \frac{1}{2} \, \mathrm{e}^{\xi_N} \, \frac{x^2}{N^2}\right)^N, \qquad 0 \leq \xi_N \leq \frac{x}{N}.$$

By the binomial theorem,

$$\mathrm{e}^x = \sum_{k=0}^{N} \binom{N}{k} \left(1 + \frac{x}{N}\right)^{N-k} \left(\frac{\mathrm{e}^{\xi_N} \, x^2}{2 \, N^2}\right)^k.$$

Going back to (10), denoting by C a constant independent of N (and thus of Δt) changing form line to line, and noticing that

$$\frac{1}{2}\,\mathrm{e}^{\xi_N}\,x^2 \le \frac{1}{2}\,\mathrm{e}^{|x|}\,x^2 =: M_x,$$

we deduce

$$\begin{aligned}
\left|\mathbb{E}\left[X_{N\Delta t} - X_N^{\Delta t}\right]\right| &\le \sum_{k=1}^{N} \binom{N}{k}\left(1+\frac{|x|}{N}\right)^{N-k}\left(\frac{M_x}{N^2}\right)^k \\
&\le \sum_{k=1}^{N} \binom{N}{k}\left(\mathrm{e}^{\frac{|x|}{N}}\right)^{N-k}\left(\frac{M_x}{N^2}\right)^k \le \sum_{k=1}^{N} N^k\,\mathrm{e}^{|x|}\left(\frac{M_x}{N^2}\right)^k \le C\sum_{k=1}^{N}|M_x\,\Delta t|^k .
\end{aligned}$$

By the formula for geometric series,

$$\sum_{k=1}^{N}|M_x\,\Delta t|^k \le M_x\,\Delta t\left(\sum_{k=0}^{\infty}|M_x\,\Delta t|^k\right) = M_x\,\Delta t\left(\frac{1}{1-M_x\,\Delta t}\right) \le C\Delta t,$$

for Δt sufficiently small, which concludes the exercise. □

Exercise 2. Repeat the previous exercise for $f(x) = x^2$.

Solution. From the properties of the lognormal distribution,

$$\mathbb{E}[|X_T|^2] = |x_0|^2\,\exp((2\mu+\sigma^2)T).$$

For the Euler–Maruyama discretization, we calculate

$$\mathbb{E}[|X_N^{\Delta t}|^2] = |x_0|^2\prod_{n=0}^{N-1}\left(|1+\mu\Delta t|^2+\sigma^2\Delta t\right) = |x_0|^2\prod_{n=0}^{N-1}\left(1+(2\mu+\sigma^2)\Delta t+\mu^2\Delta t^2\right).$$

From the calculations in the previous exercise with $x = 2\mu+\sigma^2$, we know

$$\left|\exp((2\mu+\sigma^2)T) - \prod_{n=0}^{N-1}\left(1+(2\mu+\sigma^2)\Delta t\right)\right| \le C\Delta t,$$

where, as before, C denotes a constant independent of Δt (possibly changing from line to line). It is thus sufficient to show

$$\left|\prod_{n=0}^{N-1}\left(1+(2\mu+\sigma^2)\Delta t+\mu^2\Delta t^2\right) - \prod_{n=0}^{N-1}\left(1+(2\mu+\sigma^2)\Delta t\right)\right| \le C\Delta t.$$

Letting now $x = 1+(2\mu+\sigma^2)\Delta t$ and employing the binomial theorem, we obtain

$$\begin{aligned}
\left|\prod_{n=0}^{N-1}\left(x+\mu^2\Delta t^2\right) - \prod_{n=0}^{N-1}x\right| &\le \sum_{k=1}^{N}\left|\binom{N}{k}x^{N-k}\,\mu^{2k}\,\Delta t^{2k}\right| \\
&\le C\sum_{k=1}^{N}\left|x^{N-k}\,\mu^{2k}\,\Delta t^{k}\right| \le C\,\mathrm{e}^{|2\mu+\sigma^2|T}\sum_{k=1}^{N}\left|\mu^2\,\Delta t\right|^k,
\end{aligned}$$

after which it is easy to conclude. □

Exercise 3. Let us modify the Euler–Maruyama update as follows

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + b(X_n^{\Delta t})\,\Delta t + \sigma(X_n^{\Delta t})\,\sqrt{\Delta t}\,\xi_n,$$

where $\{\xi_n\}_{n=0}^{N-1}$ are i.i.d. discrete-valued random variables taking values 1 and -1 with equal probability. Show that the corresponding weak error, for geometric Brownian motion and for the observables $f(x) = x$ and $f(x) = x^2$, also scales as Δt.

4.7. Monte Carlo Estimates of SDEs

An important application of having accurate numerical approximations of solutions to SDEs is that they allow us to generate Monte Carlo estimates of statistical quantities that depend on that solution (e.g., when we computed $\mathbb{E} f(\hat{X}_{n\Delta t})$ for the weak error!)

For example, given an SDE $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, many applications involve computing averages of observables of X_t such as

- $\mathbb{E}(f(X_T))$
- $\mathbb{E}\left(\int_{T_0}^{T} g(X_s)\,ds\right)$
- $\mathbb{E}\left(\sup_{t\in[0,T]} |X_t|\right)$

all these cases can be expressed as $I = \mathbb{E}(f(X))$

where f is a function of the path X_t over $t \in [0,T]$.

In order to estimate I we can generate N independent realisations of X_t: $\hat{X}_n^{(1)}, \hat{X}_n^{(2)}, \ldots, \hat{X}_n^{(N)}$. Each realisation is a time series with $\lfloor T/\Delta t \rfloor = n$ values. Then we can use the MC estimator:

$$\hat{I}_N^n = \frac{1}{N}\sum_{j=1}^{N} F(\{\hat{X}_k^{(j)}\}_{k=1,\ldots,n}).$$

The difference between this and the previous MC estimator we studied is that this estimator is biased because of the numerical error that comes from the numerical solution of the SDE.

We can compute the mean-square error of the estimator: example, $f(X_T)$.

$$MSE(\hat{I}_N^n) = \mathbb{E}\left(\left(\hat{I}_N^n - \mathbb{E}(f(X_T))\right)^2\right) = Var(\hat{I}_n^N) + Bias(I_n^N)^2$$

Since the N realisations are iid, we have that

$$Var(\hat{I}_n^N) = \frac{1}{N} Var\, f(\hat{X}_m).$$

Furthermore, the bias is

$$|bias(\hat{I}_N^n)| = |\mathbb{E}(\hat{I}_N^n) - \mathbb{E}(f(X_T))| = e_{weak}(f),$$

i.e., it is the weak error of the numerical approximation to the solution to the SDE! (for the observable f).

$$\Rightarrow \boxed{MSE(\hat{I}_N^n) = \frac{Var\, f(\hat{X}_m)}{N} + e_{weak}(f)^2.}$$

MC error — goes to 0 as N increases

discretisation error — goes to 0 as $\Delta t \to 0$.

The total cost of computing this estimator is $O(nN)$ ⇒ there is a tradeoff between the two errors: increasing n means decreasing N and vice versa, if we have a fixed computational cost K.

<u>Question</u>: Can we compute the optimal n and N given a computational cost W?

The answer is Yes!

Suppose we are using an EM discretisation to approximate X_t. Since it has weak error of order 1, we can write

$$e_{weak}(f) \approx C\Delta t = K_f\, T/n$$

for a constant $K_f > 0$ and the MSE can be written as

$$MSE(\hat{I}_N^n) \approx \frac{\sigma^2}{N} + \left(\frac{K_f T}{n}\right)^2 \qquad \sigma^2 = Var\, f(\hat{X}_n).$$

If we assume the computational cost is

$$C(N, n) \approx C\, N n$$

and we have a fixed computational budget W, then we wish to minimise MSE subject to this cost, i.e.

$$\begin{cases} \text{minimise} & \sigma^2/N + (K_f T/n)^2 \\ \text{subject to} & CnN = W. \end{cases}$$

We can solve this problem to find the optimal n by introducing a Lagrange multiplier λ. We need to solve

$$\partial/\partial N \left(\sigma^2/N + (K_f T/n)^2 - \lambda(CnN - W)\right) = 0$$

$$\partial/\partial n \left(\sigma^2/N + (K_f T/n)^2 - \lambda(CnN - W)\right) = 0$$

and this gives

$$\lambda C n = -\frac{\sigma^2}{N^2}, \qquad \lambda C N = -\frac{2(K_f T)^2}{n^3}$$

at the optimal value, we have $\lambda CnN = \lambda W$ and we obtain

$$\lambda W = \frac{\sigma^2}{N^2 n} = \frac{2(K_f T)^2}{n^3 N} \iff n^2 = \frac{2(K_f T)^2 N}{\sigma^2}$$

Since $CnN = W$, we conclude that

$$n \sim W^{1/3}, \quad N \sim W^{2/3}.$$

<u>Note that</u> this assumes everything is continuous. But we expect that the integer part will be close enough. This approximation should only be seen as a rule of thumb.

Suppose now that the quantity we want to estimate is a rare event. Using this MC estimator can be prohibitively expensive:

<u>example</u>: Consider $X_t = W_t$, i.e., a standard Brownian motion and suppose we want to compute

$$I = \mathbb{P}\left(\sup_{t \in [0,1]} X_t > c \right)$$

where $c > 0$. Then we have

$$\mathbb{P}\left(\sup_{t \in [0,1]} X_t > c \right) = 2\,\mathbb{P}(W_1 \geq c) = 2(1 - \phi(c))$$

reflection principle.

$$\leq 2e^{-c^2/2}$$

For $c = 5$, this probability is of the order 10^{-6}.

However, if we consider a new process
$Y_t = a\,t + W_t$
where $a > 0$ is constant, we obtain instead

$$I_a = \mathbb{P}\left(\sup_{t \in [0,1]} Y_t > c \right) = \int_0^t \frac{c}{\sqrt{2\pi s^3}} \exp\left(-\frac{(c - as)^2}{2s} \right) ds$$

So we have, for $a = 1$ $\quad I_a = 5 \cdot 10^{-5}$

$a = 2$ $\quad I_a = 2 \cdot 10^{-3}$

$a = 5$ $\quad I_a = 0.539$

So it is much easier to compute this probability for Y_t.

$\Rightarrow$ We would like to implement some sort of importance sampling to compute I using trajectories of Y_t instead of W_t. Girsanov's theorem provides us with a way to do this.

Girsanov's Theorem (Oksendal Thm 8.6.8):

Consider the SDEs

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

$$dY_t = (b(Y_t) + \gamma(t,\omega))\,dt + \sigma(Y_t)\,dW_t$$

on the time interval $t \in [0,T]$, where W_t is a standard Brownian motion and $X_0 = Y_0 = x$.

Suppose that there exists a process $u(t,\omega)$ such that

$$\sigma(Y_t)\,u(t,\omega) = \gamma(t,\omega).$$

and $u(t,\omega)$ satisfies Novikov's condition

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\int_0^T u^2(s,\omega)\,ds\right)\right) < \infty.$$

Then, define for $t \leq T$

$$M_t = \exp\left(-\int_0^t u(s,\omega)\,dW_s - \frac{1}{2}\int_0^t u^2(s,\omega)\,ds\right),$$

$$d\mathbb{Q}(\omega) = M_T\,d\mathbb{P}(\omega),$$

$$\tilde{W}_t = \int_0^t u(s,\omega)\,ds + W_t.$$

$\Rightarrow$ $\mathbb{Q}$ is a probability measure on $\mathcal{F}_T$, and

$$dY_t = b(Y_t)\,dt + \sigma(Y_t)\,d\tilde{W}_t$$

$\Rightarrow$ the $\mathbb{Q}$-law of Y_t is the same as the $\mathbb{P}$-law of X_t,

i.e., $\underbrace{\mathbb{E}(F(X))}_{\text{expected value of } F(X) \text{ in } \mathbb{P}} = \underbrace{\mathbb{E}(F(Y)M_T)}_{\text{expected value of } Y \text{ in } \mathbb{Q} = M_T\mathbb{P}.}$

<u>Note that</u>: for the previous example, we would have $M_T = \exp\left(-aW_T - \frac{a^2 T}{2}\right)$.

References

[1] C. Graham and D. Talay. *Stochastic simulation and Monte Carlo methods*. Springer, 2013. Mathematical foundations of stochastic simulation.

[2] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.