Chapter 1: Stochastic approximation schemes of stochastic differential equations and their applications in mathematical finance

Noufel Frikha

Université Paris 1 Panthéon-Sorbonne

5th January 2026

Contents

Introduction

Example: pricing of vanilla options in the 2d Black-Scholes model

We consider the following 3-dimensional dynamics $(\mathbf{X}_t)_{t\geq 0} = (X_t^0, X_t^1, X_t^2)_{t\geq 0}$ with $(X_0^0, X_0^1, X_0^2) = (1, x_0^1, x_0^2)$, under the risk neutral probability:

$$\begin{aligned} dX_t^0 &= rX_t^0\, dt \\ dX_t^1 &= rX_t^1\, dt + \sigma_1 X_t^1\, dW_t^1 \\ dX_t^2 &= rX_t^2\, dt + \sigma_2 X_t^2\, dW_t^2 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} X_t^0 &= e^{rt}, \\ X_t^1 &= x_0^1 e^{(r-\sigma_1^2/2)t+\sigma_1 W_t^1}, \\ X_t^2 &= x_0^2 e^{(r-\sigma_2^2/2)t+\sigma_2 W_t^2} \end{aligned}$$

where

$$(\mathbf{W}_t)_{t\geq 0} = (W_t^1, W_t^2)_{t\geq 0}$$

is a 2d Brownian motion such that

$$d < W^1, W^2 >_t = \rho\, dt, \quad \rho \in [-1, 1].$$

We write $W_t^2 = \rho W_t^1 + \sqrt{1-\rho^2}\widetilde{W}_t^2$ where $(W^1, \widetilde{W}^2)$ is a standard Brownian motion.

Objective: Pricing at time t of a financial contract with vanilla payoff $h(X_T^1, X_T^2)$. Its price is given by the risk neutral rule:

$$p_t = e^{-r(T-t)}\mathbb{E}[h(X_T^1, X_T^2)|\mathcal{F}_t^W].$$

◦ Example: Best of Call with payoff function $h(x^1, x^2) = (\max(x^1, x^2) - K)_+$.

◦ Key idea: Observe the following relation for $i = 1, 2$

$$\frac{X_T^i}{X_t^i} = e^{(r-\sigma_i^2/2)(T-t)+\sigma_i(W_T - W_t)} \stackrel{d}{=} \frac{X_{T-t}^i}{x_0^i} \text{ and is independent of } \mathcal{F}_t^W$$

Objective: Pricing at time t of a financial contract with vanilla payoff $h(X_T^1, X_T^2)$. Its price is given by the risk neutral rule:

$$p_t = e^{-r(T-t)}\mathbb{E}[h(X_T^1, X_T^2)|\mathcal{F}_t^W].$$

◦ Example: Best of Call with payoff function $h(x^1, x^2) = (\max(x^1, x^2) - K)_+$.

◦ Key idea: Observe the following relation for $i = 1, 2$

$$\frac{X_T^i}{X_t^i} = e^{(r-\sigma_i^2/2)(T-t)+\sigma_i(W_T - W_t)} \overset{d}{=} \frac{X_{T-t}^i}{x_0^i} \text{ and is independent of } \mathcal{F}_t^W$$

Hence, one writes

$$\begin{aligned} p_t &= e^{-r(T-t)}\mathbb{E}\left[h\left(\frac{X_T^1}{X_t^1}X_t^1, \frac{X_T^2}{X_t^2}X_t^2\right)\Big|\mathcal{F}_t^W\right] \\ &= e^{-r(T-t)}\mathbb{E}\left[h\left(x_1\frac{X_{T-t}^1}{x_0^1}, x_2\frac{X_{T-t}^2}{x_0^2}\right)\right]_{(x_1,x_2)=(X_t^1,X_t^2)} \\ &= e^{-r(T-t)}\mathbb{E}[\varphi(Z_1, Z_2)] \end{aligned}$$

where $(Z_1, Z_2) \sim \mathcal{N}(0, I_2)$ and

$$\varphi(z_1, z_2) = h(x_1 e^{(r-\frac{\sigma_1^2}{2})(T-t)+\sigma_1\sqrt{T-t}z_1}, x_2 e^{(r-\frac{\sigma_2^2}{2})(T-t)+\sigma_2\sqrt{T-t}(\rho z_1+\sqrt{1-\rho^2}z_2)}).$$

The previous quantity writes as $\mathbb{E}[X]$, for some random variable X, which can be estimated using the Monte Carlo estimator

$$\bar{X}_M = \bar{X}_M := \frac{1}{M}\sum_{n=1}^{M} X_n,$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same law as X.

Almost sure convergence

- We assume that $\mathbb{E}[|X|] < \infty$.Then, according to the *Strong Law of Large Numbers* (SLLN), one has

$$\mathbb{P}\left(\bar{X}_M := \frac{1}{M}\sum_{n=1}^{M} X_n \to \mathbb{E}[X]\right) = 1.$$

- Convergence of distribution function at each point: for any $x \in \mathbb{R}$,

$$F_M(x) := \frac{1}{M}\sum_{n=1}^{M} \mathbf{1}_{\{X_n \leq x\}} \to \mathbb{P}(X \leq x), \quad a.s.$$

  and, according to Glivenko-Cantelli's theorem, the convergence is uniform

$$\sup_{x\in\mathbb{R}} |F_M(x) - F(x)| \to 0, \quad a.s.$$

- SLLN may hold uniformly on a class of functions. Let q be a positive integer, for two positive real numbers $\underline{a}, \overline{a}$, define the vector space

$$\mathcal{A} = \left\{ f : f(x) = -\underline{a} \vee \sum_{\ell=1}^{q} \alpha_\ell \Phi_\ell(x) \wedge \overline{a}, \text{ for some } \alpha_\ell \in \mathbb{R} \right\}.$$

spanned by the functions $(\Phi_\ell)_{1 \leq \ell \leq q}$, truncated by $\underline{a}, \overline{a}$. Then, one has

$$\sup_{f \in \mathcal{A}} \left| \frac{1}{M} \sum_{i=1}^{M} f(X_i) - \mathbb{E}[f(X)] \right| \to 0, \quad a.s.$$

For a proof and other extensions of this result, please refer to:
A distribution-free theory of nonparametric regression, Springer Series in Statistics, 2002, by L. Gyorfi, M. Kohler, A. Krzyzak and H. Walk.

Rate of convergence

Fluctuations of $\bar{X}_M$ around $\mathbb{E}[X]$

- If $\sigma^2 := \mathrm{Var}(X) < \infty$, in other words, if $X \in \mathbb{L}^2(\mathbb{P})$, then the Central Limit Theorem (CLT) also yields

  $\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]) \Rightarrow \mathcal{N}(0,1), \quad M \to \infty,$ ($\Rightarrow$ means convergence in law.)

  or equivalently for every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[h(\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]))] \to \int_{\mathbb{R}} h(x)\frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}dx$$

  or equivalently for every $z \in \mathbb{R}$

$$\mathbb{P}\left(\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]) \in (z_1, z_2)\right) \to \int_{z_1}^{z_2} \exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}dx.$$

  These results are still valid if the variance σ^2 is replaced by its *empirical estimator*

$$\sigma_M^2 := \frac{1}{M}\sum_{n=1}^{M}(X_n - \bar{X}_M)^2.$$

Asymptotic Confidence Interval for the M-C method

- In particular, for $J_M^\alpha := \left[\bar{X}_M - a_\alpha \frac{\sigma_M}{\sqrt{M}}; \bar{X}_M + a_\alpha \frac{\sigma_M}{\sqrt{M}}\right]$, one has

$$\mathbb{P}\left(\mathbb{E}[X] \in J_M^\alpha\right) = \mathbb{P}\left(\sqrt{M}\frac{|\bar{X}_M - \mathbb{E}[X]|}{\sigma_M} \leq a_\alpha\right) \to \mathbb{P}(|N| \leq a_\alpha) = 2\Phi(a_\alpha) - 1$$

with $N \sim \mathcal{N}(0,1)$ and Φ its distribution function. Then, one chooses $a_\alpha = \Phi^{-1}(\frac{1+\alpha}{2})$ so that $2\Phi(a_\alpha) - 1 = \alpha$.

α	90%	95%	99%
a_α	1.65	1.96	2.58

- For $I_M^\alpha := [\bar{X}_M - c_\alpha \frac{\sigma_M}{\sqrt{M}}, +\infty)$ and $c_\alpha = \Phi^{-1}(\alpha)$, one has

$$\mathbb{P}\left(\mathbb{E}[X] \in I_M^\alpha\right) = \mathbb{P}\left(\sqrt{M}\frac{\bar{X}_M - \mathbb{E}[X]}{\sigma_M} \leq c_\alpha\right) \to \Phi(c_\alpha) = \alpha.$$

α	90%	95%	99%
c_α	1.29	1.65	2.33

The Monte Carlo method is very robust, the convergence is independent of the dimension but slow. To reduce the error by 2, one has to multiply the number of simulations by 4!

What about the convergence in the CLT?

The CLT says that $\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M - \mathbb{E}[X]))}{\sigma} \leq x\right) \to \Phi(x)$ as $M \to +\infty$ for all $x \in \mathbb{R}$.
From the second Dini's theorem,

$$\sup_{x\in\mathbb{R}} |\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M - \mathbb{E}[X]))}{\sigma} \leq x\right) - \Phi(x)| \to 0, \quad M \to +\infty$$

Question: What is the convergence rate?

Theorem, Berry-Esseen (1941-1942)
There exists a positive (universal) constant C such that

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M - \mathbb{E}[X]))}{\sigma} \leq x\right) - \Phi(x)\right| \leq C\frac{\mathbb{E}[|X - \mathbb{E}[X]|^3]}{\sigma^3\sqrt{M}}.$$

Remark: The rate $1/\sqrt{M}$ is optimal and cannot be improved in general.

Theorem, Edgeworth expansion
Still if X has a density, there exist universal polynomials $(\mathbf{P}_\ell)_{1\leq \ell \leq L}$ such that

$$\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M - \mathbb{E}[X]))}{\sigma} \leq x\right) - \Phi(x) = \sum_{\ell=1}^{L} \frac{1}{M^{\frac{\ell}{2}}}\mathbf{P}_\ell(\partial)[\Phi](x) + \mathcal{O}(M^{-(L+1)/2}).$$

Coming back to the price of the contract with payoff $h(X_T^1, X_T^2)$, the Monte Carlo estimator is

$$\bar{p}_M(x_1,x_2) := \frac{1}{M}\sum_{n=1}^{M} e^{-r(T-t)}\varphi(Z_1^{(n)}, Z_2^{(n)}),$$

$$\bar{\sigma}_M^2 := \frac{1}{M-1}\sum_{n=1}^{M}(e^{-r(T-t)}\varphi(Z_1^{(n)}, Z_2^{(n)}) - \bar{p}_M(x_1,x_2))^2$$

where

$$\varphi(z_1,z_2) = h\left(x_1 e^{(r-\frac{\sigma_1^2}{2})(T-t)+\sigma_1\sqrt{T-t}z_1}, x_2 e^{(r-\frac{\sigma_2^2}{2})(T-t)+\sigma_2\sqrt{T-t}(\rho z_1+\sqrt{1-\rho^2}z_2)}\right)$$

and $(Z_1^{(n)}, Z_2^{(n)})_{1\leq n\leq M}$ are i.i.d. random variables with law $\mathcal{N}(0, I_2)$.

◦ Don't forget to compute the confidence interval: For M large enough, one has

$$\mathbb{P}(p_t(x_1,x_2) \in I_M^\alpha) = \alpha \quad \text{with } I_M^\alpha := \left[\bar{p}_M - a_\alpha \frac{\sigma_M}{\sqrt{M}}, \bar{p}_M + a_\alpha \frac{\sigma_M}{\sqrt{M}}\right]$$

Some results to compute Greeks

Our aim now is to provide a general framework to compute the derivatives of

$$x \mapsto \mathbb{E}[h(X_T^x)]$$

where $(X_t^x = x_0 \exp((r - \sigma^2/2)t + \sigma W_t))_{t \geq 0}$ follows the B.-S. dynamics.

Theorem: Differentiation under Expectation

Let $\varphi : I \times \Omega \to \mathbb{R}$ be a random function $\mathcal{B}(\mathbb{R}) \otimes \mathcal{A}$-measurable defined on a given probability space and $I \subset \mathbb{R}$ an interval.

a) Local version: Let $x_0 \in I$. If φ satisfies the following conditions:
- $\forall x \in I$, $\varphi(x, .) \in \mathbb{L}^1(\mathbb{P})$.
- $\partial_x \varphi(x_0, w)$ exists for *almost all* w.
- $\exists Y \in \mathbb{L}^1(\mathbb{P})$ s.t. $\forall x \in I$, $\frac{|\varphi(x,w) - \varphi(x_0,w)|}{|x - x_0|} \leq Y(w)$.

Then, $x \mapsto p(x) := \mathbb{E}[\varphi(x, w)]$ is well-defined, differentiable at $x_0 \in I$ and

$$\partial_x p(x_0) = \mathbb{E}[\partial_x \varphi(x_0, w)].$$

b) Global version:

- $\forall x \in I$, $\varphi(x, .) \in \mathbb{L}^1(\mathbb{P})$;
- $\partial_x \varphi(x, w)$ exists for all $x \in I$ and *almost all* w.
- $\exists Y \in \mathbb{L}^1(\mathbb{P})$, s.t. $\forall x \in I$, one has $|\partial_x \varphi(x, w)| \leq Y(w)$.

Then, $x \mapsto p(x) := \mathbb{E}[\varphi(x, w)]$ is well-defined, continuously differentiable on I and

$$\partial_x p(x) = \mathbb{E}[\partial_x \varphi(x, w)].$$

Application to Delta and Gamma hedging

Proposition:

1. If $h : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{C}^1$ and h' is of polynomial growth then p_0 is also $\mathcal{C}^1$:

$$p_0'(x) = \mathbb{E}\left[h'(X_T^x)\frac{X_T^x}{x}\right]$$

2. If h is Borel measurable and of polynomial growth then p_0 is also $\mathcal{C}^1$:

$$p_0'(x) = \mathbb{E}\left[h(X_T^x)\frac{W_T}{x\sigma T}\right].$$

Proof:

1. $\partial_x h(X_T^x) = h'(X_T^x)\partial_x X_T^x = h'(X_T^x)X_T^x/x$ and since h' is of polynomial growth for all $x \in [0, R]$, $R > 0$, one has

$$|\partial_x h(X_T^x)| \leq C(1 + |X_T^x|^p)e^{\sigma W_T} \leq C(1 + R^p e^{p\sigma W_T})e^{\sigma W_T} \in \mathbb{L}^1(\mathbb{P}).$$

2. For h smooth then extend by approximation

$$\begin{aligned}
\partial_x \mathbb{E}[h(X_T^x)] &= \mathbb{E}[h'(X_T^x)e^{(r-\frac{\sigma^2}{2})T+\sigma W_T}] \\
&= \int h'(xe^{(r-\frac{\sigma^2}{2})T+\sigma z})e^{(r-\frac{\sigma^2}{2})T+\sigma z}\frac{e^{-\frac{z^2}{2T}}}{\sqrt{2\pi T}}\,dz \\
&\overset{IBP}{=} \int h(xe^{(r-\frac{\sigma^2}{2})T+\sigma z})\frac{z}{x\sigma T}e^{-\frac{z^2}{2T}}\frac{dz}{\sqrt{2\pi T}} \\
&= \mathbb{E}[h(X_T^x)\frac{W_T}{x\sigma T}].
\end{aligned}$$

Let $Y(\theta)$ be a random variable depending on a parameter $\theta \in \mathbb{R}^q$ with a positive density function $y \mapsto p(\theta, y)$ w.r.t. a reference measure $\mu \in \mathbb{R}^d$, that is,

$$\mathbb{E}[h(Y(\theta))] = \int_{\mathbb{R}^d} h(y)p(\theta, y)\,\mu(dy).$$

Proposition: Assume that $\mathbb{R}^q \ni \theta \mapsto p(\theta, y)$ is continuously differentiable and satisfies for all $\theta \in \Theta \subset \mathbb{R}^q$

$$|\partial_\theta p(\theta, y)| \leq \bar{p}(y), \quad \text{with} \quad \int \bar{p}(y)\mu(dy) < \infty.$$

Then, for any bounded and measurable map $h : \mathbb{R}^d \to \mathbb{R}$ and any $\theta \in \Theta$, it holds

$$\partial_\theta \mathbb{E}[h(Y(\theta)] = \mathbb{E}[h(Y(\theta))\partial_\theta \log(p(\theta, y))_{y=Y(\theta)}].$$

Proof: Very simple. Write $\mathbb{E}[h(Y(\theta))] = \int h(y)p(\theta,y)\mu(dy)$ and use Lebesgue's differentiation theorem to get

$$\partial_\theta \mathbb{E}[h(Y(\theta))] = \int h(y)\frac{\partial_\theta p(\theta,y)}{p(\theta,y)}p(\theta,y)\mu(dy) = \mathbb{E}\Big[h(Y(\theta))\frac{\partial_\theta p(\theta,Y(\theta))}{p(\theta,Y(\theta))}\Big].$$

Examples:

- If $Y \sim \mathcal{P}(\lambda)$, then one has

$$\partial_\lambda \mathbb{E}[h(Y)] = \mathbb{E}[h(Y)(\log(Y) - 1)]$$

- If $Y \sim \mathcal{N}(m,\sigma^2)$, then one

$$\partial_m \mathbb{E}[h(Y)] = \mathbb{E}[h(Y)\frac{Y-m}{\sigma^2}], \quad \partial_\sigma \mathbb{E}[h(Y)] = \mathbb{E}\Big[h(Y)\frac{1}{\sigma}[\frac{(Y-m)^2}{\sigma^2} - 1]\Big].$$

We have seen some Monte-Carlo estimators with some explicit asymptotic control of the error. But they are valid only for M large enough.
Question: How to choose M in practice? Difficult question...

We here want to develop some non-asymptotic point of view.

Reference: *Concentration inequalities, a non-asymptotic theory of independence* by S. Boucheron, G. Lugosi and P. Massart.

A concentration inequality writes as follows: $\forall M \geq 0, \forall \varepsilon > 0$, one has

$$\mathbb{P}(\bar{X}_M - \mathbb{E}[X] > \varepsilon) \leq \exp(-M\phi(\varepsilon))$$

or

$$\mathbb{P}(|\bar{X}_M - \mathbb{E}[X]| > \varepsilon) \leq 2\exp(-M\phi(\varepsilon))$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing, convex function s.t. $\phi(0) = 0$.

Typical examples:

- $\phi(\varepsilon) = \beta^{-1}\varepsilon^2$ corresponds to the *Gaussian concentration regime*.
- $\phi(\varepsilon) = \beta^{-1}\varepsilon$ corresponds to the *Poisson/Exponential concentration regime*.

Some non-asymptotic control for bounded r.v. 1/2

- Proposition: Let X be a r.v. such that $\mathbb{P}(X \in [a, b]) = 1$ for some $-\infty < a < b < +\infty$. Then, one has

$$Var(X) \leq \frac{(b-a)^2}{4}.$$

Proof:
Notice that for any $p \geq 1$,
$\inf_{x\in\mathbb{R}} |X - x|^p \leq |X - (a + b)/2|^p \leq \max(|a - b|^p/2^p; |b - a|^p/2^p)$ so that for $p = 2$, one gets

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \inf_{x\in\mathbb{R}} \mathbb{E}[(X - x)^2] \leq \frac{(b-a)^2}{4}.$$

Some non-asymptotic control for bounded r.v. 2/2

- Proposition (Hoeffding's inequality): Let $(X_1, \cdots, X_M)$ be a sequence of i.i.d. random variables satisfying:

$$\forall m \in \{1, \cdots, M\}, \quad \mathbb{P}(X_m \in [a_m, b_m]) = 1.$$

Then, for every $\varepsilon \geq 0$, one has

$$\mathbb{P}\left(|\frac{1}{M}\sum_{m=1}^{M} X_m - \mathbb{E}[X_m]| \geq \varepsilon \right) \leq 2\exp\left(-2\frac{M\varepsilon^2}{(M^{-1}\sum_{m=1}^{M}|b_m - a_m|^2)} \right)$$

In particular, if $b_m = b$, $a_m = a$, we obtain

$$\mathbb{P}\left(|\frac{1}{M}\sum_{m=1}^{M} X_m - \mathbb{E}[X_m]| \geq \varepsilon \right) \leq 2\exp\left(-2\frac{M\varepsilon^2}{|b-a|^2} \right).$$

Proof: By convexity of $x \mapsto \exp(\lambda x)$, one has

$$e^{\lambda(\beta b+(1-\beta)a)} \leq \beta e^{\lambda b} + (1-\beta)e^{\lambda a}$$

so that taking

$$\beta = \frac{X_m - \mathbb{E}[X_m] - a}{b-a}, \quad \text{so that} \quad 1-\beta = \frac{b-(X_m - \mathbb{E}[X_m])}{b-a}$$

which satisfies $\beta b + (1-\beta)a = X_m - \mathbb{E}[X_m]$, we get

$$\exp(\lambda(X_m - \mathbb{E}[X_m])) \leq \frac{X_m - \mathbb{E}[X_m] - a}{b-a} e^{\lambda b} + \frac{b-(X_m - \mathbb{E}[X_m])}{b-a} e^{\lambda a}.$$

Taking expectation in both sides of the previous inequality yields

$$\mathbb{E}[\exp(\lambda(X_m - \mathbb{E}[X_m]))] \leq f(\lambda) := pe^{\lambda b} + (1-p)e^{\lambda a},$$

with $p := \frac{-a}{b-a}$.

Note that the function f satisfies:

$$f(\lambda) := e^{\lambda a}(pe^{\lambda(b-a)} + (1-p),$$
$$f'(\lambda) := e^{\lambda a}(pbe^{\lambda(b-a)} + a(1-p)),$$
$$f''(\lambda) := e^{\lambda a}(pb^2e^{\lambda(b-a)} + a^2(1-p)).$$

so that

$$\begin{aligned}\partial^2_\lambda \log(f(\lambda)) &= \frac{f''(\lambda)f(\lambda) - f'(\lambda)^2}{f(\lambda)^2} \\ &= \frac{(b-a)^2 p(1-p)e^{\lambda(b-a)}}{(pe^{\lambda(b-a)} + (1-p))^2} \\ &\leq \frac{(b-a)^2}{4}\end{aligned}$$

since $4p(1-p)e^{\lambda(b-a)} \leq (pe^{\lambda(b-a)} + (1-p))^2$. Hence, for some $\hat{\lambda} \in [0, \lambda]$, it holds

$$\log(f(\lambda)) = \log(f(0)) + \frac{f'(0)}{f(0)}\lambda + \frac{1}{2}\partial^2_\lambda \log(f(\lambda))_{|\lambda=\hat{\lambda}}\lambda^2 \leq \frac{(b-a)^2}{8}\lambda^2.$$

We thus conclude that

$$\mathbb{E}[\exp(\lambda(X_m - \mathbb{E}[X_m])] \leq \exp(\lambda^2(b-a)^2/8).$$

By the exponential Markov inequality, for any $\lambda \geq 0$, one gets

$$\begin{aligned}
\mathbb{P}(\frac{1}{M}\sum_{k=1}^{M}(X_k - \mathbb{E}[X_k]) \geq \varepsilon) &\leq e^{-\lambda M\varepsilon}\mathbb{E}[e^{\lambda\sum_{k=1}^{M}(X_k - \mathbb{E}[X_k])}] \\
&= e^{-\lambda M\varepsilon}\prod_{k=1}^{M}\mathbb{E}[e^{\lambda(X_k - \mathbb{E}[X_k])}] \\
&\leq e^{-\lambda M\varepsilon + \frac{\lambda^2(b-a)^2 M}{8}}.
\end{aligned}$$

Optimizing with respect to λ the right-hand side of the last inequality allows to conclude the proof.

Gaussian concentration inequality

References: Gross (1975), Ledoux (1999), Villani *Optimal transport: old and new* (2009).

Definition: Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. For a Borel function $f : \mathbb{R}^d \to \mathbb{R}_+$, we introduce the *Entropy* of f w.r.t. the measure μ:

$$Ent_\mu(f) = \mathbb{E}_\mu[f \log(f)] - \mathbb{E}_\mu[f] \log(\mathbb{E}_\mu[f])$$

Observe that $Ent_\mu(f) < \infty$ iif $f \log(f)_+$ is μ-integrable and since $x \mapsto x \log(x)$ is convex, $Ent_\mu(f) \geq 0$.

Definition: A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies the logarithmic-Sobolev inequality with a constant C_μ if for all smooth functions ($\mathcal{C}_b^1$ is enough), one has:

$$Ent_\mu(f) \leq C_\mu \mathbb{E}_\mu[|\nabla f|^2]$$

Proposition: If $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies the logarithmic-Sobolev inequality with a constant C_μ then for all 1-Lipschitz function f one has

$$\forall \lambda \geq 0, \quad \mathbb{E}_\mu(e^{\lambda f}) \leq e^{\lambda \mathbb{E}_\mu[f] + \frac{C_\mu \lambda^2}{4}}.$$

Proof: Herbst's argument.

Proof: Assume first that $f \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$ with $|\nabla f|_\infty \leq 1$. Set $F = \exp(\frac{1}{2}\lambda f)$, then $F^2 \in \mathcal{C}_b^1$. By the Sobolev inequality, one has

$$\begin{aligned} Ent_\mu(F^2) &= \mathbb{E}_\mu[\lambda f e^{\lambda f}] - \mathbb{E}_\mu[e^{\lambda f}]\log(\mathbb{E}_\mu[e^{\lambda f}]) \\ &= \lambda H'(\lambda) - H(\lambda)\log(H(\lambda)) \\ &\leq C_\mu \mathbb{E}_\mu[|\nabla F|^2] = C_\mu \mathbb{E}_\mu[\frac{\lambda^2}{4}|\nabla f|^2 e^{\lambda f}] \\ &\leq C_\mu \frac{\lambda^2}{4} H(\lambda) \end{aligned}$$

where $H(\lambda) := \mathbb{E}_\mu[e^{\lambda f}]$. Set $K(\lambda) := \frac{1}{\lambda}\log(H(\lambda))$ so that

$$K'(\lambda) = \frac{1}{\lambda^2}\left\{\lambda \frac{H'(\lambda)}{H(\lambda)} - \log(H(\lambda))\right\} \leq \frac{C_\mu}{4}$$

and $K(0) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda}\log(H(\lambda)) = \mathbb{E}_\mu[f]$ since $H(0) = 1$ and $H'(0) = \mathbb{E}_\mu[f]$. Hence, $K(\lambda) \leq \frac{C_\mu}{4}\lambda + \mathbb{E}_\mu[f]$ and $H(\lambda) \leq \exp(\frac{C_\mu}{4}\lambda^2 + \lambda\mathbb{E}_\mu[f])$. An approximation argument allows to extend the proof to f 1-Lipschitz.

Gaussian concentration inequality

An important example of probability measure that satisfies the log-Sobolev inequality is the Gaussian measure

$$\gamma_d(dx) = (2\pi)^{-\frac{d}{2}} \exp(-\frac{|x|^2}{2})dx$$

with constant $C_\mu = 2$.

Corrolary (Non-asymptotic concentration inequality): Let $(X_1, \cdots, X_M)$ be a sequence of i.i.d. random variables with common law μ satisfying the Log-Sobolev inequality with constant C_μ. Then, for all $M \geq 0$, for all $\varepsilon \geq 0$ and for all 1-Lipschitz function f, one has

$$\mathbb{P}\left(\left|\frac{1}{M}\sum_{m=1}^{M} f(X_m) - \mathbb{E}[f(X)]\right| \geq \varepsilon\right) \leq 2\exp(-M\frac{\varepsilon^2}{C_\mu}).$$

Proof: For all $\lambda \geq 0$, it holds

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} f(X_m) - \mathbb{E}[f(X)] \geq \varepsilon\right) \leq e^{-M\lambda\varepsilon}[\mathbb{E}[e^{\lambda(f(X_1)-\mathbb{E}[f(X_1)]}]]^M$$

$$\leq e^{-M\lambda\varepsilon + \frac{C_\mu \lambda^2 M}{4}}$$

Then, optimize the right hand side over $\lambda \geq 0$, i.e., take $\lambda = 2\varepsilon/C_\mu$, one gets

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} f(X_m) - \mathbb{E}[f(X)] \geq \varepsilon\right) \leq \exp(-M\varepsilon^2/C_\mu)$$

and conclude taking $f \equiv -f$.

Non-asymptotic confidence interval

The previous concentration inequality allows to device non-asymptotic confidence interval.

Proposition: With probability greater than $1-\alpha$, $\alpha \in (0,1)$, for all positive integer M, for all Lipschitz function f, one has

$$-[f]_1 \sqrt{\frac{C_\mu \log(2/\alpha)}{M}} \leq \frac{1}{M} \sum_{m=1}^{M} f(X_m) - \mathbb{E}[f(X)] \leq [f]_1 \sqrt{\frac{C_\mu \log(2/\alpha)}{M}}.$$

Option pricing of European option in Bachelier's model

◦ Consider the pricing problem of $h(S_T)$ in Bachelier's model with dynamics

$$dS_t = rS_t\,dt + \sigma dW_t, \quad S_0 = x, \Rightarrow S_T \sim S_0 e^{rT} + \sqrt{\frac{\sigma^2}{2r}(e^{2rT}-1)}Z,\ Z \sim \mathcal{N}(0,1).$$

◦ Assuming that h is Lipschitz continuous, we deduce that $f : z \mapsto e^{-rT} h(S_0 e^{rT} + \sqrt{\frac{\sigma^2}{2r}(e^{2rT}-1)}z)$ is Lipschitz continuous with

$$[f]_L := \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|} \leq [h]_L \sqrt{\frac{\sigma^2}{2r}(1-e^{-2rT})}.$$

◦ With probability greater than $1-\alpha$, $\alpha \in (0,1)$, for all positive integer M, for all Lipschitz payoff h, one has

$$\begin{aligned} -[h]_L \sqrt{\frac{\sigma^2}{2r}(1-e^{-2rT})}\sqrt{\frac{C_\mu \log(2/\alpha)}{M}} &\leq \frac{1}{M}\sum_{m=1}^{M} e^{-rT} h(S_T^{(m)}) - e^{-rT}\mathbb{E}[h(S_T)] \\ &\leq [h]_L \sqrt{\frac{\sigma^2}{2r}(1-e^{-2rT})}\sqrt{\frac{C_\mu \log(2/\alpha)}{M}}. \end{aligned}$$

Stochastic approximation: why?

◦ Stochastic approximation schemes appear in many fields of applied mathematics, especially in mathematical finance for the computation of option prices and sensivities and also in stochastic optimization/control.

$\rightsquigarrow$ axiomatics and stochastic calculus theory and numerical probability.

Stochastic approximation: why?

◦ Stochastic approximation schemes appear in many fields of applied mathematics, especially in mathematical finance for the computation of option prices and sensitivities and also in stochastic optimization/control.

$\rightsquigarrow$ axiomatics and stochastic calculus theory and numerical probability.

◦ Pricing and hedging of financial options: In mathematical finance, the price at time t of a financial contract with payoff $h(X_T)$ at maturity T is given by the risk-neutral rule

$$p_t = e^{-r(T-t)}\mathbb{E}[h(X_T)|\mathcal{F}_t].$$

When the underlying process $(X_t)_{t\geq 0}$ is Markov then

$$p_t = p_t(X_t) \quad \text{with} \quad p_t(x) = e^{-r(T-t)}\mathbb{E}[h(X_T)|X_t = x].$$

In this case, the delta hedge is based on the computation of the process

$$\delta_t = \nabla_x p_t(X_t), \quad t \geq 0.$$

Example: pricing of vanilla options in the 2d Black-Scholes model

We consider the following 3-dimensional dynamics $(\mathbf{X}_t)_{t\geq 0} = (X_t^0, X_t^1, X_t^2)_{t\geq 0}$ with $(X_0^0, X_0^1, X_0^2) = (1, x_0^1, x_0^2)$, under the risk neutral probability:

$$\begin{aligned} dX_t^0 &= rX_t^0\,dt \\ dX_t^1 &= rX_t^1\,dt + \sigma_1 X_t^1\,dW_t^1 \\ dX_t^2 &= rX_t^2\,dt + \sigma_2 X_t^2\,dW_t^2 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} X_t^0 &= e^{rt}, \\ X_t^1 &= x_0^1 e^{(r-\sigma_1^2/2)t+\sigma_1 W_t^1}, \\ X_t^2 &= x_0^2 e^{(r-\sigma_2^2/2)t+\sigma_2 W_t^2} \end{aligned}$$

where

$$(\mathbf{W}_t)_{t\geq 0} = (W_t^1, W_t^2)_{t\geq 0}$$

is a 2d Brownian motion such that

$$d < W^1, W^2 >_t = \rho\,dt, \quad \rho \in [-1, 1].$$

We write $W_t^2 = \rho W_t^1 + \sqrt{1-\rho^2}\widetilde{W}_t^2$ where $(W^1, \widetilde{W}^2)$ is a standard Brownian motion.

Objective: Pricing at time t of a financial contract with vanilla payoff $h(X_T^1, X_T^2)$. Its price is given by the risk neutral rule:

$$p_t = e^{-r(T-t)}\mathbb{E}[h(X_T^1, X_T^2)|\mathcal{F}_t^W].$$

- Example: Best of Call with payoff function $h(x^1, x^2) = (\max(x^1, x^2) - K)_+$.
- Key idea: Observe the following relation for $i = 1, 2$

$$\frac{X_T^i}{X_t^i} = e^{(r-\sigma_i^2/2)(T-t)+\sigma_i(W_T - W_t)} \overset{d}{=} \frac{X_{T-t}^i}{x_0^i} \text{ and is independent of } \mathcal{F}_t^W$$

Objective: Pricing at time t of a financial contract with vanilla payoff $h(X_T^1, X_T^2)$.
Its price is given by the risk neutral rule:

$$p_t = e^{-r(T-t)}\mathbb{E}[h(X_T^1, X_T^2)|\mathcal{F}_t^W].$$

◦ Example: Best of Call with payoff function $h(x^1, x^2) = (\max(x^1, x^2) - K)_+$.

◦ Key idea: Observe the following relation for $i = 1, 2$

$$\frac{X_T^i}{X_t^i} = e^{(r-\sigma_i^2/2)(T-t)+\sigma_i(W_T-W_t)} \overset{d}{=} \frac{X_{T-t}^i}{x_0^i}$$ and is independent of $\mathcal{F}_t^W$

Hence, one writes

$$\begin{aligned} p_t &= e^{-r(T-t)}\mathbb{E}\Big[h\Big(\frac{X_T^1}{X_t^1}X_t^1, \frac{X_T^2}{X_t^2}X_t^2\Big)\Big|\mathcal{F}_t^W\Big] \\ &= e^{-r(T-t)}\mathbb{E}\Big[h\Big(x_1\frac{X_{T-t}^1}{x_0^1}, x_2\frac{X_{T-t}^2}{x_0^2}\Big)\Big]_{(x_1,x_2)=(X_t^1,X_t^2)} \\ &= e^{-r(T-t)}\mathbb{E}[\varphi(Z_1, Z_2)] \end{aligned}$$

where $(Z_1, Z_2) \sim \mathcal{N}(0, I_2)$ and

$$\varphi(z_1, z_2) = h(x_1 e^{(r-\frac{\sigma_1^2}{2})(T-t)+\sigma_1\sqrt{T-t}z_1}, x_2 e^{(r-\frac{\sigma_2^2}{2})(T-t)+\sigma_2\sqrt{T-t}(\rho z_1+\sqrt{1-\rho^2}z_2)}).$$

The previous quantity writes as $\mathbb{E}[X]$, for some random variable X, which can be estimated using the Monte Carlo estimator

$$\bar{X}_M = \bar{X}_M := \frac{1}{M}\sum_{n=1}^{M} X_n,$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same law as X.

Almost sure convergence

- We assume that $\mathbb{E}[|X|] < \infty$.Then, according to the *Strong Law of Large Numbers* (SLLN), one has

$$\mathbb{P}\left(\bar{X}_M := \frac{1}{M}\sum_{n=1}^{M} X_n \to \mathbb{E}[X]\right) = 1.$$

- Convergence of distribution function at each point: for any $x \in \mathbb{R}$,

$$F_M(x) := \frac{1}{M}\sum_{n=1}^{M} \mathbf{1}_{\{X_n \leq x\}} \to \mathbb{P}(X \leq x), \quad a.s.$$

  and, according to Glivenko-Cantelli's theorem, the convergence is uniform

$$\sup_{x\in\mathbb{R}} |F_M(x) - F(x)| \to 0, \quad a.s.$$

- SLLN may hold uniformly on a class of functions. Let q be a positive integer, for two positive real numbers $\underline{a}, \overline{a}$, define the vector space

$$\mathcal{A} = \left\{ f : f(x) = -\underline{a} \vee \sum_{\ell=1}^{q} \alpha_\ell \Phi_\ell(x) \wedge \overline{a}, \text{ for some } \alpha_\ell \in \mathbb{R} \right\}.$$

spanned by the functions $(\Phi_\ell)_{1 \leq \ell \leq q}$, truncated by $\underline{a}, \overline{a}$. Then, one has

$$\sup_{f \in \mathcal{A}} \left| \frac{1}{M} \sum_{i=1}^{M} f(X_i) - \mathbb{E}[f(X)] \right| \to 0, \quad a.s.$$

For a proof and other extensions of this result, please refer to:
A distribution-free theory of nonparametric regression, Springer Series in Statistics, 2002, by L. Gyorfi, M. Kohler, A. Krzyzak and H. Walk.

Rate of convergence

Fluctuations of $\bar{X}_M$ around $\mathbb{E}[X]$

- If $\sigma^2 := \mathrm{Var}(X) < \infty$, in other words, if $X \in \mathbb{L}^2(\mathbb{P})$, then the Central Limit Theorem (CLT) also yields

  $\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]) \Rightarrow \mathcal{N}(0,1), \quad M \to \infty,$ ($\Rightarrow$ means convergence in law.)

  or equivalently for every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[h(\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]))] \to \int_{\mathbb{R}} h(x)\frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}dx$$

  or equivalently for every $z \in \mathbb{R}$

$$\mathbb{P}\left(\sqrt{M}\sigma^{-1}(\bar{X}_M - \mathbb{E}[X]) \in (z_1, z_2)\right) \to \int_{z_1}^{z_2} \exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}dx.$$

  These results are still valid if the variance σ^2 is replaced by its *empirical estimator*

$$\sigma_M^2 := \frac{1}{M}\sum_{n=1}^{M}(X_n^2 - \bar{X}_M)^2.$$

Asymptotic Confidence Interval for the M-C method

- In particular, for $J_M^\alpha := \left[\bar{X}_M - a_\alpha \frac{\sigma_M}{\sqrt{M}}; \bar{X}_M + a_\alpha \frac{\sigma_M}{\sqrt{M}}\right]$, one has

$$\mathbb{P}\left(\mathbb{E}[X] \in J_M^\alpha\right) = \mathbb{P}\left(\sqrt{M}\frac{|\bar{X}_M - \mathbb{E}[X]|}{\sigma_M} \leq a_\alpha\right) \to \mathbb{P}(|N| \leq a_\alpha) = 2\Phi(a_\alpha)-1$$

with $N \sim \mathcal{N}(0,1)$ and Φ its distribution function. Then, one chooses $a_\alpha = \Phi^{-1}(\frac{1+\alpha}{2})$ so that $2\Phi(a_\alpha) - 1 = \alpha$.

α	90%	95%	99%
a_α	1.65	1.96	2.58

- For $I_M^\alpha := [\bar{X}_M - c_\alpha \frac{\sigma_M}{\sqrt{M}}, +\infty)$ and $c_\alpha = \Phi^{-1}(\alpha)$, one has

$$\mathbb{P}\left(\mathbb{E}[X] \in I_M^\alpha\right) = \mathbb{P}\left(\sqrt{M}\frac{\bar{X}_M - \mathbb{E}[X]}{\sigma_M} \leq c_\alpha\right) \to \Phi(c_\alpha) = \alpha.$$

α	90%	95%	99%
c_α	1.29	1.65	2.33

The Monte Carlo method is very robust, the convergence is independent of the dimension but slow. To reduce the error by 2, one has to multiply the number of simulations by 4!

What about the convergence in the CLT?

The CLT says that $\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M-\mathbb{E}[X]))}{\sigma}\leq x\right)\to\Phi(x)$ as $M\to+\infty$ for all $x\in\mathbb{R}$.
From the second Dini's theorem,

$$\sup_{x\in\mathbb{R}}|\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M-\mathbb{E}[X]))}{\sigma}\leq x\right)-\Phi(x)|\to 0,\quad M\to+\infty$$

Question: What is the convergence rate?

Theorem, Berry-Esseen (1941-1942)
There exists a positive (universal) constant C such that

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M-\mathbb{E}[X]))}{\sigma}\leq x\right)-\Phi(x)\right|\leq C\frac{\mathbb{E}[|X-\mathbb{E}[X]|^3]}{\sigma^3\sqrt{M}}.$$

Remark: The rate $1/\sqrt{M}$ is optimal and cannot be improved in general.

Theorem, Edgeworth expansion
Still if X has a density, there exist universal polynomials $(\mathbf{P}_\ell)_{1\leq\ell\leq L}$ such that

$$\mathbb{P}\left(\sqrt{M}\frac{(\bar{X}_M-\mathbb{E}[X]))}{\sigma}\leq x\right)-\Phi(x)=\sum_{\ell=1}^{L}\frac{1}{M^{\frac{\ell}{2}}}\mathbf{P}_\ell(\partial)[\Phi](x)+\mathcal{O}(M^{-(L+1)/2}).$$

Coming back to the price of the contract with payoff $h(X_T^1, X_T^2)$, the Monte Carlo estimator is

$$\bar{p}_M(x_1, x_2) := \frac{1}{M}\sum_{n=1}^{M} e^{-r(T-t)}\varphi(Z_1^{(n)}, Z_2^{(n)}),$$

$$\bar{\sigma}_M^2 := \frac{1}{M-1}\sum_{n=1}^{M}(e^{-r(T-t)}\varphi(Z_1^{(n)}, Z_2^{(n)}) - \bar{p}_M(x_1, x_2))^2$$

where

$$\varphi(z_1, z_2) = h\left(x_1 e^{(r-\frac{\sigma_1^2}{2})(T-t)+\sigma_1\sqrt{T-t}z_1}, x_2 e^{(r-\frac{\sigma_2^2}{2})(T-t)+\sigma_2\sqrt{T-t}(\rho z_1+\sqrt{1-\rho^2}z_2)}\right)$$

and $(Z_1^{(n)}, Z_2^{(n)})_{1\leq n\leq M}$ are i.i.d. random variables with law $\mathcal{N}(0, I_2)$.

◦ Don't forget to compute the confidence interval: For M large enough, one has

$$\mathbb{P}(p_t(x_1, x_2) \in I_M^\alpha) = \alpha \quad \text{with } I_M^\alpha := \left[\bar{p}_M - a_\alpha \frac{\sigma_M}{\sqrt{M}}, \bar{p}_M + a_\alpha \frac{\sigma_M}{\sqrt{M}}\right]$$

General Framework

Mathematical framework

In mathematical finance, the price of an (vanilla) option writes

$$\mathbb{E}[F(X_T)]$$

where in general $X = (X_t)_{t\in[0,T]}$ is a stochastic process with dynamics

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s \tag{1}$$

where $W = (W_t)_{t\geq 0}$ is a Brownian motion of dimension q. We let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$. We work under the following assumption $(\mathbf{H}^{\sigma,b})$:

- $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$ are Lipschitz continuous in space uniformly in time

  $$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \leq K|x-y|$$

  and linear growth in space uniformly in time

  $$\|\sigma(t,x)\| + |b(t,x)| \leq K(1+|x|).$$

Mathematical framework

Theorem: Existence and uniqueness + Markov property

Under $(\mathbf{H}^{\sigma,b})$, the SDE (1) admits a unique $\mathcal{F}$-adapted solution $X = (X_t)_{t\geq 0}$ defined on any probability space $(\Omega, \mathcal{A}, \mathbb{P})$ equipped with the Brownian motion W.

Moreover, X satisfies the Markov property, namely, for any real-valued bounded and measurable map h defined on $\mathbb{R}^d$

$$\mathbb{E}[h(X_t)|\mathcal{F}_s] = \mathbb{E}[h(X_t)|X_s] := P_{s,t}h(X_s).$$

Two sources of error

In general, we don't know the exact form/law of the solution to (1) so that we don't have an analytical formula for the computation of the price $\mathbb{E}[F(X_T)]$.

In this case, one usually proceeds in two steps:

1. We introduce an approximation/discretization scheme $(\bar{X}_t)_{t\in[0,T]}$ of the SDE (1) and replace $\mathbb{E}[F(X_T)]$ by $\mathbb{E}[F(\bar{X}_T)]$. The corresponding error is

$$\mathcal{E}_W = \mathbb{E}[F(X_T)] - \mathbb{E}[F(\bar{X}_T)]. \quad \text{(Weak error)}$$

2. We simulate M i.i.d. copies $(\bar{X}_T^{(k)})_{1\leq k\leq M}$ of $\bar{X}_T$ and compute the quantity

$$E_M = \frac{1}{M}\sum_{k=1}^{M} F(\bar{X}_T^{(k)}).$$

The corresponding error is given by

$$\mathcal{E}_S = \mathbb{E}[F(\bar{X}_T)] - E_M. \text{ (Statistical error)}$$

Euler-Maruyama discretization scheme

Constant step Euler scheme

For a fixed $T > 0$, we choose the number of discretization dates $N \in \mathbb{N}\backslash\{0\}$ and consider the time mesh $\Delta = T/N$ and the discrete time grid points $t_i = i\Delta$, $i = 0, \cdots, N$ and define the constant step Euler scheme $(\bar{X}_{t_i})_{0 \leq i \leq N}$

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b(t_i, \bar{X}_{t_i})\Delta + \sigma(t_i, \bar{X}_{t_i})(W_{t_{i+1}} - W_{t_i}), \quad \bar{X}_0 = x \tag{2}$$

or equivalently

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b(t_i, \bar{X}_{t_i})\Delta + \sigma(t_i, \bar{X}_{t_i})\sqrt{\Delta}U_{i+1}, \quad \bar{X}_0 = x$$

where $(U_i)_{1 \leq i \leq N}$ is an i.i.d. sequence with common law $\mathcal{N}(0, I_q)$.

Constant step Euler scheme

For a fixed $T > 0$, we choose the number of discretization dates $N \in \mathbb{N}\backslash\{0\}$ and consider the time mesh $\Delta = T/N$ and the discrete time grid points $t_i = i\Delta$, $i = 0, \cdots, N$ and define the constant step Euler scheme $(\bar{X}_{t_i})_{0\leq i\leq N}$

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b(t_i, \bar{X}_{t_i})\Delta + \sigma(t_i, \bar{X}_{t_i})(W_{t_{i+1}} - W_{t_i}), \quad \bar{X}_0 = x \tag{2}$$

or equivalently

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b(t_i, \bar{X}_{t_i})\Delta + \sigma(t_i, \bar{X}_{t_i})\sqrt{\Delta}U_{i+1}, \quad \bar{X}_0 = x$$

where $(U_i)_{1\leq i\leq N}$ is an i.i.d. sequence with common law $\mathcal{N}(0, I_q)$. The constant step Euler scheme $(\bar{X}_{t_i})_{0\leq i\leq N}$ is a discrete time inhomogeneous Markov chain with transition $(P_{i,j})_{1\leq i\leq j\leq N}$ defined by

$$P_{i,j}h(x) = \mathbb{E}[h(\bar{X}_{t_j})|\bar{X}_{t_i} = x] = \mathbb{E}[h(\bar{X}^{t_i,x}_{t_j})] = \int_{\mathbb{R}^d} h(y)\, P_{i,j}(x, dy)$$

where for $i \leq j \leq N-1$

$$\bar{X}^{t_i,x}_{t_{j+1}} = \bar{X}^{t_i,x}_{t_j} + b(t_j, \bar{X}^{t_i,x}_{t_j})\Delta + \sigma(t_j, \bar{X}^{t_i,x}_{t_j})\sqrt{\Delta}U_{j+1}, \quad \bar{X}^{t_i,x}_{t_i} = x$$

Continuous time Euler scheme

The constant step Euler scheme is easy to implement on a computer but from a theoretical point of view it is not an Itô process. To circumvent this issue, we define a continuous version of the previous process, called the continuous time Euler scheme, $(\bar{X}^c_{t_i})_{0\leq i\leq N}$

$$\forall t \in [t_i, t_{i+1}],\ \bar{X}^c_t = \bar{X}^c_{t_i} + b(t_i, \bar{X}^c_{t_i})(t - t_i) + \sigma(t_i, \bar{X}^c_{t_i})(W_t - W_{t_i}), \quad \bar{X}_0 = x. \tag{3}$$

Note that the process $(\bar{X}^c_t)_{t\in[0,T]}$ coincides with the constant step Euler scheme $(\bar{X}_{t_i})_{0\leq i\leq N}$ on the time grid $(t_i)_{0\leq i\leq N}$.

Continuous time Euler scheme

The constant step Euler scheme is easy to implement on a computer but from a theoretical point of view it is not an Itô process. To circumvent this issue, we define a continuous version of the previous process, called the continuous time Euler scheme, $(\bar{X}^c_{t_i})_{0\leq i\leq N}$

$$\forall t \in [t_i, t_{i+1}],\ \bar{X}^c_t = \bar{X}^c_{t_i} + b(t_i, \bar{X}^c_{t_i})(t-t_i) + \sigma(t_i, \bar{X}^c_{t_i})(W_t - W_{t_i}), \quad \bar{X}_0 = x. \quad (3)$$

Note that the process $(\bar{X}^c_t)_{t\in[0,T]}$ coincides with the constant step Euler scheme $(\bar{X}_{t_i})_{0\leq i\leq N}$ on the time grid $(t_i)_{0\leq i\leq N}$. The continuous Euler scheme $(\bar{X}^c_t)_{t\in[0,T]}$ is an Itô process with dynamics

$$\bar{X}^c_t = x + \int_0^t b(\phi(s), \bar{X}^c_{\phi(s)})ds + \int_0^t \sigma(\phi(s), \bar{X}^c_{\phi(s)})dW_s$$

where $\phi(t) := \sup\{t_i : t_i \leq t\}$, i.e. $\phi(t) = t_i$ for $t \in [t_i, t_{i+1})$.

Goals

When $N \uparrow \infty$, one expects that both schemes $(\bar{X}_{t_i})_{0\leq i\leq N}$ and $(\bar{X}^c_{t\in[0,T]})$ converge to the true solution $(X_t)_{t\in[0,T]}$ of (1). In this chapter, we aim at studying the error of this approximation:

1. Strong error: is the error on the path space. It is quantified by

$$\mathbb{E}[\sup_{0\leq t\leq T}|X_t - \bar{X}^c_t|^p]^{\frac{1}{p}} \quad \text{and} \quad \mathbb{E}[\sup_{0\leq t\leq T}|X_t - \bar{X}_{\phi(t)}|^p]^{\frac{1}{p}}, \quad p \geq 1.$$

2. Weak error: For a payoff function f,

$$\mathcal{E}_W = \mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}^c_T)]$$

depending on the smoothness of f.

Control of L^p moments

Under **(H)** $\equiv$ **(H1)**$+$ **(H2)**, for any $p \geq 1$, there exists a constant κ such that

$$\mathbb{E}[\sup_{0\leq t\leq T} |X_t|^p]^{\frac{1}{p}} \leq \kappa(1+|x|).$$

Similarly, for any $p \geq 1$, there exists a positive constant such for any $N \in \backslash\{0\}$, the continuous Euler scheme satisfies

$$\mathbb{E}[\sup_{0\leq t\leq T} |\bar{X}^c_t|^p]^{\frac{1}{p}} \leq \kappa(1+|x|).$$

Useful inequalities

Grönwall's Lemma

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a measurable and locally bounded function satisfying for some non-decreasing $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$

$$\exists \alpha > 0, \forall t \geq 0, \quad f(t) \leq \alpha \int_0^t f(s)\, ds + \Psi(t)$$

then for all $t \geq 0$: $\sup_{0 \leq s \leq t} f(s) \leq e^{\alpha t} \Psi(t)$.

Useful inequalities

Grönwall's Lemma

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a measurable and locally bounded function satisfying for some non-decreasing $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$

$$\exists \alpha > 0, \forall t \geq 0, \quad f(t) \leq \alpha \int_0^t f(s)\, ds + \Psi(t)$$

then for all $t \geq 0$: $\sup_{0 \leq s \leq t} f(s) \leq e^{\alpha t} \Psi(t)$.

Doob's inequality

Let $M = (M_t)_{t \geq 0}$ be a continuous martingale with $M_0 = 0$. Then, for all $T > 0$,

$$\forall p > 1, \quad \mathbb{E}[\sup_{0 \leq t \leq T} |M_t|^p] \leq (p/(p-1))^p \mathbb{E}[|M_T|^p]$$

Burkholder-Davis-Gundy's inequality

$\forall p \geq 1$, $\exists c_p$, $C_p > 0$, s.t. for all continuous local martingale $(M_t)_{t \geq 0}$, $M_0 = 0$, for all stopping time τ,

$$c_p \mathbb{E}[\langle M \rangle_\tau^{\frac{p}{2}}] \leq \mathbb{E}[(\sup_{0 \leq t \leq \tau} |M_t|)^p] \leq C_p \mathbb{E}[\langle M \rangle_\tau^{\frac{p}{2}}]$$

Proof of the L^p controls

Proof: Assume for simplicity that $q = d = 1$. For $p \geq 1$,

$$|X_t|^{2p} \leq 3^{2p-1}\left\{|x|^{2p} + t^{2p-1}\int_0^t |b(s,X_s)|^{2p}\,ds + |M_t|^{2p}\right\}, \quad M_t = \int_0^t \sigma(s,X_s)dW_s,$$

local martingale. Introduce the sequence of stopping time $\tau_m := \inf\{t \geq 0 : |X_t - x| \geq m\}$, $m \in$, $\tau_m \uparrow \infty$, a.s. as $m \uparrow \infty$. Using **(H)**, one gets that M satisfies $\langle M \rangle_{t\wedge\tau_m} = \int_0^{t\wedge\tau_m} |\sigma(s,X_s)|^2\,ds \leq \kappa(t + \int_0^{t\wedge\tau_m} |X_s|^2 ds)$. Thus, M^{τ_m} is a continuous martingale and by the B-D-G inequality

$$\mathbb{E}[\sup_{0\leq s\leq t} |M_s^{\tau_m}|^{2p}] \leq C_p \mathbb{E}[\langle M^{\tau_m}\rangle_t^{p}] \leq \kappa\left(1 + \mathbb{E}[\int_0^{t\wedge\tau_m} |X_{s\wedge\tau_m}|^{2p} ds]\right)$$

so that

$$\begin{aligned}\mathbb{E}[\sup_{0\leq s\leq t} |X_s^{\tau_m}|^{2p}] &\leq \kappa\left\{|x|^{2p} + 1 + \mathbb{E}[\int_0^{t\wedge\tau_m} |X_s^{\tau_m}|^{2p} ds]\right\} \\ &\leq \kappa\left\{|x|^{2p} + 1 + \int_0^t \mathbb{E}[\sup_{0\leq r\leq s} |X_r^{\tau_m}|^{2p}]\,ds\right\}.\end{aligned}$$

Apply Grönwall's Lemma to $t \mapsto f(t) = \mathbb{E}[\sup_{0\leq s\leq t} |X_s^{\tau_m}|^{2p}]$, to get

$$\mathbb{E}[\sup_{0\leq s\leq t} |X_s^{\tau_m}|^{2p}] \leq \kappa(1 + |x|^{2p})$$

Strong approximation error

Main result

Assumption: $\exists \beta \in [0,1]$, $\exists C > 0$ s.t. $\forall (s,t,x,y) \in [0,T]^2 \times (\mathbb{R}^d)^2$,

$$|b(t,x) - b(s,y)| + \|\sigma(t,x) - \sigma(s,y)\| \leq C(|t-s|^\beta + |y-x|) \quad (\mathbf{H}_T^\beta)$$

Theorem

If b and σ satisfies $(\mathbf{H}_T^\beta)$ then the strong error between $\bar{X}^c$ and X in any $L^p(\mathbb{P})$, $p > 0$, is of order $N^{-(\frac{1}{2}\wedge\beta)}$, that is, $\forall p \geq 1$:

$$\| \sup_{0\leq t\leq T} |X_t - \bar{X}_t^c| \|_{L^p(\mathbb{P})} \leq C(1+|x|)N^{-(\frac{1}{2}\wedge\beta)} \textit{(Continuous time)}$$

$$\| \sup_{0\leq t\leq T} |X_t - \bar{X}_{\phi(t)}^c| \|_{L^p(\mathbb{P})} \leq C(1+|x|)\left(N^{-(\frac{1}{2}\wedge\beta)} + \sqrt{\frac{1+\log(N)}{N}}\right) \quad \textit{(Constant step)}$$

Theorem

If b and σ satisfies: $|b(t,x)| + \|\sigma(t,x)\| \leq C(1+|x|)$ then one has

$$\forall p \geq 1,\ \mathbb{E}[\sup_{0\leq t\leq T} |\bar{X}_t^c - \bar{X}_{\phi(t)}^c|^p]^{\frac{1}{p}} \leq C(1+|x|)\sqrt{\frac{1+\log(N)}{N}}$$

Proof

For the first point (the second point is left as an exercice), we write

$$X_u - X_u^c = \int_0^u (\sigma(s, X_s) - \sigma(\phi(s), \bar{X}^c_{\phi(s)})\, dW_s + \int_0^u (b(s, X_s) - b(\phi(s), \bar{X}^c_{\phi(s)}))\, ds$$

so that, by the B-D-G inequality

$$\begin{aligned}
\mathbb{E}[\sup_{0\leq u\leq t} |X_u - \bar{X}_u^c|^{2p}] &\leq \kappa \mathbb{E}[(\int_0^t \|\sigma(s, X_s) - \sigma(\phi(s), \bar{X}^c_{\phi(s)})\|^2\, ds)^p] \\
&\quad + \kappa \int_0^t \mathbb{E}[|b(s, X_s) - b(\phi(s), \bar{X}^c_{\phi(s)})|^{2p}]\, ds \\
&\leq \kappa \int_0^t \mathbb{E}[\|\sigma(s, X_s) - \sigma(\phi(s), \bar{X}^c_{\phi(s)})\|^{2p}]\, ds \\
&\quad + \kappa \int_0^t \mathbb{E}[|b(s, X_s) - b(\phi(s), \bar{X}^c_{\phi(s)})|^{2p}]\, ds
\end{aligned}$$

Note that $(\mathbf{H}_T^\beta)$ implies

$$\begin{aligned}
|b(s, X_s) - b(\phi(s), \bar{X}^c_{\phi(s)})|^{2p} + \|\sigma(s, X_s) - \sigma(\phi(s), \bar{X}^c_{\phi(s)})\|^{2p} \\
&\leq \kappa(|s - \phi(s)|^{2p\beta} + |X_s - \bar{X}^c_{\phi(s)}|^{2p}) \\
&\leq \kappa(|s - \phi(s)|^{2p\beta} + |X_s - \bar{X}^c_s|^{2p} + |\bar{X}^c_s - \bar{X}^c_{\phi(s)}|^{2p})
\end{aligned}$$

Hence, we get

$$\mathbb{E}[\sup_{0\leq u\leq t}|X_u-\bar{X}^c_u|^{2p}]\leq \kappa\left\{\Delta^{2p\beta}+\int_0^t \mathbb{E}[|\bar{X}^c_s-\bar{X}^c_{\phi(s)}|^{2p}]\,ds+\int_0^t \mathbb{E}[|X_s-\bar{X}^c_s|^{2p}]\,ds\right\}$$

so that, by Grönwall's Lemma, we get

$$\mathbb{E}[\sup_{0\leq u\leq t}|X_u-\bar{X}^c_u|^{2p}]\leq \kappa\left\{\Delta^{2p\beta}+\int_0^t \mathbb{E}[|\bar{X}^c_s-\bar{X}^c_{\phi(s)}|^{2p}]\,ds\right\}.$$

To conclude we remark that
$\bar{X}^c_s-\bar{X}^c_{\phi(s)}=\sigma(\phi(s),\bar{X}^c_{\phi(s)})(W_s-W_{\phi(s)})+b(\phi(s),\bar{X}^c_{\phi(s)})(s-\phi(s))$ implies

$$\begin{aligned}\mathbb{E}[|\bar{X}^c_s-\bar{X}^c_{\phi(s)}|^{2p}] &\leq \kappa\left\{\mathbb{E}[\|\sigma(\phi(s),\bar{X}^c_{\phi(s)})\|^{2p}|W_s-W_{\phi(s)}|^{2p}]+\Delta^{2p}\mathbb{E}[|\bar{X}^c_{\phi(s)}|^{2p}]\right\}\\ &\leq \kappa(1+|x|^{2p})(\Delta^p+\Delta^{2p})\end{aligned}$$

so that

$$\mathbb{E}[\sup_{0\leq u\leq t}|X_u-\bar{X}^c_u|^{2p}]\leq \kappa(1+|x|^{2p})\left\{\Delta^{2p\beta}+\Delta^p+\Delta^{2p}\right\}.$$

For the second theorem, we write:

$$\bar{X}^c_t - \bar{X}^c_{\phi(t)} = b(\phi(t), \bar{X}^c_{\phi(t)})(t-\phi(t)) + \sigma(\phi(t), \bar{X}^c_{\phi(t)})(W_t - W_{\phi(t)})$$

so that

$$\sup_{0\le t\le T} |\bar{X}^c_t - \bar{X}^c_{\phi(t)}| \le \kappa(1 + \sup_{0\le t\le T} |\bar{X}^c_t|)(\Delta + \sup_{0\le t\le T} |W_t - W_{\phi(t)}|)$$

and by the Cauchy-Schwarz inequality

$$\begin{aligned}\| \sup_{0\le t\le T} |\bar{X}^c_t - \bar{X}^c_{\phi(t)}| \|_{L^p(\mathbb{P})} \le \kappa(1 + \| \sup_{0\le t\le T} |\bar{X}^c_t| \|_{L^{2p}(\mathbb{P})}) \\ \times (\Delta + \| \sup_{0\le t\le T} |W_t - W_{\phi(t)}| \|_{L^{2p}(\mathbb{P})}).\end{aligned}$$

We now use the decomposition

$$\sup_{0\le t\le T} |W_t - W_{\phi(t)}| = \sqrt{\Delta} \max_{1\le k\le N} \sup_{t_{k-1}\le t\le t_k} \frac{|W_t - W_{t_{k-1}}|}{\sqrt{\Delta}} = \sqrt{\Delta} \max_{1\le k\le N} \zeta_k$$

where $\zeta_k := \sup_{t_{k-1}\le t\le t_k} \frac{W_t - W_{t_{k-1}}}{\sqrt{\Delta}} \overset{d}{=} \sup_{k-1\le s\le k} |W_s - W_{k-1}|$, $(\zeta_k)_{1\le k\le N}$ are i.i.d. r.v.

Observe that $\zeta_1 \overset{d}{=} \sup_{0\leq s\leq 1}|W_s| \overset{d}{=} |W_1|$ so that $\mathbb{E}[\exp(\lambda\zeta_1^2)] < \infty$ for $\lambda < \frac{1}{2}$. To conclude the proof, we use the following lemma

Lemma

Let $(X_k)_{k\geq 1}$ be an i.i.d. sequence of positive r.v. such that $\mathbb{E}[\exp(\lambda X_1)] < \infty$ for some $\lambda > 0$. Then, for all $p > 0$, $\|\max(X_1, \cdots, X_n)\|_{L^p(\mathbb{P})} \leq C(\log(n) + 1)$

The previous lemma implies

$\|\sup_{0\leq t\leq T}|W_t - W_{\phi(t)}|\|_{L^p(\mathbb{P})} \leq \sqrt{\Delta}\|\max_{1\leq k\leq N}\zeta_k^2\|_{L^p(\mathbb{P})}^{\frac{1}{2}} \leq C\sqrt{\Delta}\sqrt{1+\log(N)}$

Proof of the lemma: Set $f(x) = \exp(x^{1/p})$ and $f^{-1}(x) = \log(x)^p$, so that $(f^{-1})''(x) = p(\log(x))^{p-2}\frac{p-1-\log(x)}{x^2}$ and f^{-1} is concave on $[\kappa(p), \infty)$. For some $\lambda > 0$,

$$\begin{aligned}
\lambda\mathbb{E}[\max(X_1, \cdots, X_n)^p] &\leq \mathbb{E}[f^{-1}(f(\lambda\max(X_1, \cdots, X_n)^p))\mathbf{1}_{\{\lambda\max(X_1,\cdots,X_n)\geq\kappa(p)\}}] \\
&\quad + \lambda\kappa(p) \\
&\leq f^{-1}(n\mathbb{E}[\exp(\lambda^{\frac{1}{p}}X_1)]) + \lambda\kappa(p) \\
&\leq \kappa(\log(n)^p + 1).
\end{aligned}$$

Corollary

Corollary

Under the assumptions of the previous theorem, one has

$$\forall \alpha < \frac{1}{2} \wedge \beta, \quad N^{\alpha} \sup_{0 \le t \le T} |X_t - \bar{X}_t^c| \xrightarrow{a.s.} 0.$$

Proof: For all $\alpha < \frac{1}{2} \wedge \beta$, $\mathbb{E}[(N^{\alpha} \sup_{0 \le t \le T} |X_t - \bar{X}_t^c|)^{2p}] \le C_p N^{-2p(\frac{1}{2} \wedge \beta - \alpha)}$ so that taking p s.t. $2p(\frac{1}{2} \wedge \beta - \alpha) > 1$ one has

$$\mathbb{E}[\sum_{N \ge 1} (N^{\alpha} \sup_{0 \le t \le T} |X_t - \bar{X}_t^c|)^{2p}] < \infty$$

which implies

$$N^{\alpha} \sup_{0 \le t \le T} |X_t - \bar{X}_t^c| \xrightarrow{a.s.} 0, \quad N \to \infty.$$

Heuristics

Assume here $d = q = 1$ and that $b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$. We observe that when Δ is small one has

$$\mathbb{E}[(\int_{t_k}^{t_{k+1}} b(X_s)\,ds)^2] \leq \Delta \int_{t_k}^{t_{k+1}} \mathbb{E}[b^2(X_s)]\,ds = O(\Delta^2),$$

$$\mathbb{E}[(\int_{t_k}^{t_{k+1}} \sigma(X_s)\,dW_s)^2] = \int_{t_k}^{t_{k+1}} \sigma^2(X_s)\,ds = O(\Delta)$$

so that the global strong error is ruled by the discretization error of the stochastic integral. To improve the strong rate of convergence, one has to improve the approximation of the stochastic integral: for $s \in [t_k, t_{k+1}]$

$$\begin{aligned}\sigma(X_s) &= \sigma(X_{t_k}) + \sigma'(X_{t_k})(X_s - X_{t_k}) + O_{L^2(\mathbb{P})}(s - t_k)\\ &= \sigma(X_{t_k}) + (\sigma\sigma')(X_{t_k})(W_s - W_{t_k}) + O_{L^2(\mathbb{P})}(s - t_k)\end{aligned}$$

so that

$$\begin{aligned}\int_{t_k}^{t} \sigma(X_s)\,dW_s &= \sigma(X_{t_k})(W_t - W_{t_k}) + (\sigma\sigma')(X_{t_k}) \int_{t_k}^{t} (W_s - W_{t_k})\,dW_s + O_{L^2(\mathbb{P})}((t-t_k)\\ &= \sigma(X_{t_k})(W_t - W_{t_k}) + \frac{1}{2}(\sigma\sigma')(X_{t_k})\left\{(W_t - W_{t_k})^2 - (t - t_k)\right\}\\ &\quad + O_{L^2(\mathbb{P})}((t-t_k)^{\frac{3}{2}})\end{aligned}$$

This heuristics lead us to introduce the following constant step and continuous time *Milstein schemes:*

$$\begin{aligned}\tilde{X}_{t_{k+1}} &= \tilde{X}_{t_k} + b(\tilde{X}_{t_{k+1}})\Delta + \sigma(\tilde{X}_{t_{k+1}})(W_{t_{k+1}} - W_{t_k}) \\ &\quad + \frac{1}{2}(\sigma\sigma')(\tilde{X}_{t_k})\left\{(W_{t_k+1} - W_{t_k})^2 - (t_{k+1} - t_k)\right\}\end{aligned}$$

and

$$\begin{aligned}\tilde{X}^c_t &= \tilde{X}^c_{t_k} + b(\tilde{X}^c_{t_k})(t - t_k) + \sigma(\tilde{X}^c_{t_{k+1}})(W_t - W_{t_k}) \\ &\quad + \frac{1}{2}(\sigma\sigma')(\tilde{X}^c_{t_k})\left\{(W_t - W_{t_k})^2 - (t - t_k)\right\}\end{aligned}$$

Theorem

If $b, \sigma \in \mathcal{C}^{1+\alpha}(\mathbb{R})$ *for some* $\alpha \in [0,1]$ *with bounded derivatives then: for all* $p > 0$*, there exists a positive constant* C *s.t.:*

$$\| \max |X_{t_k} - \tilde{X}_{t_k}| \|_{L^p(\mathbb{P})} \leq \| \sup_{0 \leq t \leq T} |X_t - \tilde{X}^c_t| \|_{L^p(\mathbb{P})} \leq C N^{-\frac{1+\alpha}{2}}(1 + |x|)$$

and

$$\| \sup_{0 \leq t \leq T} |X_t - \tilde{X}^c_{\phi(t)}| \|_{L^p(\mathbb{P})} \leq C(1 + |x|)\sqrt{(1 + \log(N))} N^{-1/2}$$

Remark: If b', σ' are Hölder-continuous, Milstein scheme converges faster than the E-M scheme

The multi-dimensional generalization of the previous result is far from being obvious and relies on the possibility of simulating the vector

$$\left(\int_{t_k}^{t_{k+1}} (W_s^i - W_{t_k}^i) dW_s^j,\, i,j = 1, \cdots, q,\, i \neq j\right).$$

For the moment, there is no reasonable methods to do that! An important simplification is when the following commuting condition is satisfied

$$\forall (j,m) \in \{1,\cdots,q\},\, (\partial\sigma_j\sigma_m)(x) = (\partial\sigma_m\sigma_j)(x) \Leftrightarrow \sum_{k=1}^{q} \left(\frac{\partial\sigma_{i,j}}{\partial x_k}\sigma_{k,m}\right)(x) = \sum_{k=1}^{q} \left(\frac{\partial\sigma_{i,m}}{\partial x_k}\sigma_{k,j}\right)$$

In this case, the Milstein scheme writes

$$\tilde{X}_{t_{k+1}} = \tilde{X}_{t_k} + b(\tilde{X}_{t_k})\Delta - \frac{1}{2}\sum_{i=1}^{q} (\partial\sigma_i\sigma_i)(\tilde{X}_{t_k}) + \sigma(\tilde{X}_{t_k})(W_{t_{k+1}} - W_{t_k})$$
$$+ \frac{1}{2}\sum_{1\leq i,j\leq q} (\partial\sigma_i\sigma_j)(\tilde{X}_{t_k})(W_{t_k+1} - W_{t_k})_i(W_{t_k+1} - W_{t_k})_j$$

Corollary

If $b \in \mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R}^d)$ and if $\sigma(x) = \Sigma$ is a constant matrix then the Euler-Maruyama and Milstein schemes coincide. Moreover, for any $p \geq 1$, it holds

$$\| \max_{0\leq k\leq N} |X_{t_k} - \tilde{X}_{t_k}\|_{L^p(\mathbb{P})} \leq CN^{-1}(1 + |x|)$$

The weak approximation error

We now study the weak approximation error which writes

$$\mathcal{E}_W := \mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]$$

Notice that as soon as f, b, σ are Lipschitz continuous, from the previous results on the strong error, one has

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]| \leq [f]_{Lip}\mathbb{E}[|X_T^1 - \bar{X}_T|] \leq C[f]_{Lip}N^{-\frac{1}{2}}(1+|x|)$$

so that $|\mathcal{E}_W| = O(N^{-\frac{1}{2}})$.

However, this inequality is far from being optimal! Numerical experiments suggest

$$\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)] \approx \frac{C}{N}.$$

Standard results on the weak error

Theorem (Talay-Tubaro, 1990; Bally Talay, 1996)

1. *Assume that b, $\sigma \in \mathcal{C}^4(\mathbb{R}^d)$ with bounded derivatives. Assume that $f \in \mathcal{C}^4(\mathbb{R}^d)$ with derivatives of polynomial growth. Then, for every $x \in \mathbb{R}^d$, one has*
$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]| \leq \frac{C}{N}.$$
2. *Assume that b, σ are infinitely differentiable with bounded derivatives:*
   1. *If f is infinitely differentiable with derivatives of polynomial growth, then, for every $x \in \mathbb{R}^d$, for every $R \in \mathbb{N}$, one has*
   $$\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)] = \sum_{k=1}^{N} \frac{C_k}{N^k} + O(\frac{1}{N^{R+1}}), \quad as \quad N \uparrow \infty \tag{4}$$
   where the constants $(C_k)_{1 \leq k \leq R}$ depend only on f, b, σ, T.
   2. *If σ is uniformly elliptic, that is, for all $x \in \mathbb{R}^d$, $a(x) = (\sigma\sigma^T)(x) \geq \lambda I_d$, for some $\lambda > 0$, then the expansion (4) is still valid for any Borel function f of polynomial growth.*

Proof of (1)**:** The proof relies on the smoothness of the solution to the following parabolic PDE

$$(\partial_t + \mathcal{L}_t)u(t,x) = 0, \forall (t,x) \in [0,T] \times \mathbb{R}^d, \quad u(T,x) = f(x).$$

with $\mathcal{L}_t u(t,x) = \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(t,x)\partial^2_{x_i,x_j} + \sum_{i=1}^d b_i(t,x)\partial_{x_i}$. Under our assumptions, $u \in \mathcal{C}^{2,4}([0,T] \times \mathbb{R}^d)$ with derivatives of polynomial growth.

Proof of (1)**:** The proof relies on the smoothness of the solution to the following parabolic PDE

$$(\partial_t + \mathcal{L}_t)u(t,x) = 0, \forall (t,x) \in [0,T] \times \mathbb{R}^d, \quad u(T,x) = f(x).$$

with $\mathcal{L}_t u(t,x) = \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(t,x)\partial^2_{x_i,x_j} + \sum_{i=1}^d b_i(t,x)\partial_{x_i}$. Under our assumptions, $u \in \mathcal{C}^{2,4}([0,T] \times \mathbb{R}^d)$ with derivatives of polynomial growth. Moreover, from the Feynman-Kac formula, it follows that: $u(t,x) = \mathbb{E}[f(X_T^{t,x})]$ where

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x})\,dr + \int_s^t \sigma(r, X_r^{s,x})\,dW_r$$

Proof of (1)**:** The proof relies on the smoothness of the solution to the following parabolic PDE

$$(\partial_t + \mathcal{L}_t)u(t,x) = 0, \forall (t,x) \in [0,T] \times \mathbb{R}^d, \quad u(T,x) = f(x).$$

with $\mathcal{L}_t u(t,x) = \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(t,x)\partial^2_{x_i,x_j} + \sum_{i=1}^d b_i(t,x)\partial_{x_i}$. Under our assumptions, $u \in \mathcal{C}^{2,4}([0,T]\times\mathbb{R}^d)$ with derivatives of polynomial growth. Moreover, from the Feynman-Kac formula, it follows that: $u(t,x) = \mathbb{E}[f(X_T^{t,x})]$ where

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x})\,dr + \int_s^t \sigma(r, X_r^{s,x})\,dW_r$$

Next, we observe the relation $\mathbb{E}[f(\bar{X}_T^c)] = \mathbb{E}[u(T,\bar{X}_T^c)]$ and $\mathbb{E}[f(X_T)] = u(0,x)$ so that

$$\mathbb{E}[f(\bar{X}_T^c)] - \mathbb{E}[f(X_T)] = \mathbb{E}[u(T,\bar{X}_T^c)] - u(0,x) = \sum_{k=1}^N \mathbb{E}[u(t_k,\bar{X}_{t_k}^c) - u(t_{k-1},\bar{X}_{k-1}^c)]$$

Then, apply Itô's rule, on the time interval $[t_{k-1}, t_k]$,

$$u(t_k,\bar{X}_{t_k}^c) - u(t_{k-1},\bar{X}_{t_{k-1}}^c) = \int_{t_{k-1}}^{t_k} (\partial_t + \bar{\mathcal{L}})u(s,\bar{X}_s^c)ds + \int_{t_{k-1}}^{t_k} \sigma(\phi(s),\bar{X}_{\phi(s)}^c)\partial_x u(s,\bar{X}_s^c)dW_s$$

with $\bar{\mathcal{L}}h(s,x) = \sum_{i=1}^d b_i(\phi(s),\bar{X}_{\phi(s)}^c)\partial_{x_i}h(s,x) + \sum_{i,j=1}^d a_{i,j}(\phi(s),\bar{X}_{\phi(s)}^c)\partial^2_{x_i,x_j}h(s,x)$.

The bracket of the local martingale $M_t := \int_{t_{k-1}}^t \sigma(\phi(s), \bar{X}^c_{\phi(s)}) \partial_x u(s, \bar{X}^c_s) dW_s$, for $t \in [t_{k-1}, t_k]$ is bounded by

$$< M >_t \leq \int_0^T |\partial_x u(s, \bar{X}^c_s)|^2 \|\sigma(\phi(s), \bar{X}^c_{\phi(s)}\|^2 \, ds \leq C \int_0^T (1+|X^c_s|^{2p}+|\bar{X}^c_{\phi(s)}|^{2p}) \, ds \in L^1(\mathbb{P})$$

so that M is a true martingale. Hence, we have

$$\mathbb{E}[u(t_k, \bar{X}^c_{t_k}) - u(t_{k-1}, \bar{X}^c_{k-1})] = \int_{t_{k-1}}^{t_k} (\partial_t + \bar{\mathcal{L}}) u(s, \bar{X}^c_s) \, ds = \int_{t_{k-1}}^{t_k} f_k(s, \bar{X}^c_s) \, ds$$

with $f_k(t,x) :=$
$\partial_t u(t,x) + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(t_{k-1}, \bar{X}^c_{t_{k-1}}) \partial^2_{x_i, x_j} u(t,x) + \sum_{i=1}^d b_i(t_{k-1}, \bar{X}^c_{t_{k-1}}) \partial_{x_i} u(t,x)$
satisfying

- $f_k(t_{k-1}, \bar{X}^c_{t_{k-1}}) = 0$ easily seen from the PDE satisfied by $u(t,x)$.
- $(t,x) \mapsto f_k(t,x) \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d)$ with derivatives of polynomial growth.

From Itô's rule, one gets

$$f_k(t, \bar{X}^c_t) = \int_{t_{k-1}}^t \partial_x f_k(s, \bar{X}^c_s) \sigma(t_{k-1}, \bar{X}^c_{t_{k-1}}) \, dW_s + \int_{t_{k-1}}^t (\partial_t + \bar{\mathcal{L}}) f_k(s, \bar{X}^c_s) \, ds$$

The local martingale $N_t = \int_{t_{k-1}}^t \partial_x f_k(s, \bar{X}^c_s) \sigma(t_{k-1}, \bar{X}^c_{t_{k-1}}) \, dW_s$ is a true martingale so that

$$\mathbb{E}[f_k(t, \bar{X}^c_t)] = \int_{t_{k-1}}^t \mathbb{E}[(\partial_t + \bar{\mathcal{L}}) f_k(s, \bar{X}^c_s)] \, ds \quad \text{and} \quad \mathbb{E}[\sup_{0 \leq s \leq T} |(\partial_t + \bar{\mathcal{L}}) f_k(s, \bar{X}^c_s)|] < \infty.$$

Conclusion of the proof

It follows that

$$\begin{aligned}\mathbb{E}[u(t_k,\bar{X}^c_{t_k})-u(t_{k-1},\bar{X}^c_{k-1})] &= \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}\mathbb{E}[(\partial_t+\bar{\mathcal{L}})f_k(s,\bar{X}^c_s)]\,ds\,dt \\ &\leq \mathbb{E}[\sup_{0\leq s\leq T}|(\partial_t+\bar{\mathcal{L}})f_k(s,\bar{X}^c_s)|]\frac{1}{2}\Delta^2\end{aligned}$$

so that

$$\begin{aligned}|\mathbb{E}[f(\bar{X}^c_T)]-\mathbb{E}[f(X_T)]| &= |\sum_{k=1}^{N}\mathbb{E}[u(t_k,\bar{X}^c_{t_k})-u(t_{k-1},\bar{X}^c_{k-1})]| \\ &\leq \mathbb{E}[\sup_{0\leq s\leq T}|(\partial_t+\bar{\mathcal{L}})f_k(s,\bar{X}^c_s)|]\frac{1}{2}\frac{T^2}{N}.\end{aligned}$$

This concludes the first part of the theorem. In order to obtain the full expansion, one relies on the smoothness of $(t,x)\mapsto u(t,x)$ and repeated use of Itô's Lemma in order to get rid of stochastic integrals.
To get rid of the smoothness assumption on the payoff function f, one relies on additional tools that go beyond this course: Malliavin calculus or the parametrix method.

Improving the Monte Carlo path simulation method for SDEs: old and new

Bias-variance decomposition

◦ Bias-Variance decomposition of the quadratic error:

$$\|\mathbb{E}[f(X_T)] - \frac{1}{M}\sum_{k=1}^{M} f(\bar{X}_T^{(k)})\|_{L^2(\mathbb{P})}^2 = (\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)])^2$$

$$+ \|\mathbb{E}[f(\bar{X}_T)] - \frac{1}{M}\sum_{k=1}^{M} f(\bar{X}_T^{(k)})\|_{L^2(\mathbb{P})}^2$$

$$= \left(\frac{C_1}{N}\right)^2 + \frac{\mathrm{Var}(f(\bar{X}_T))}{M} + O(N^{-3}).$$

◦ Interpretation: In order to achieve a global $L^2(\mathbb{P})$-error

$$\|\mathbb{E}[f(X_T)] - \frac{1}{M}\sum_{k=1}^{M} f(\bar{X}_T^{(k)}\|_{L^2(\mathbb{P})} = O(\varepsilon)$$

one has to choose

$$N \approx \varepsilon^{-1}, \quad M \approx \varepsilon^{-2}$$

so that the optimal complexity of the Monte Carlo method is $C_{MC} = \kappa\varepsilon^{-3}$.

Richardson-Romberg extrapolation method

◦ Idea: take advantage of the expansion of the weak error.
Take $(\bar{X}_t^N)_{t\in[0,T]}$ and $(\bar{X}_t^{2N})_{t\in[0,T]}$ two Euler schemes with two Brownian motions W^1 and W^2 and two mesh sizes $\Delta = \frac{T}{N}$ and $\frac{\Delta}{2}$.

$$\mathbb{E}[f(\bar{X}_T^N)] - \mathbb{E}[f(X_T)] = \frac{C_1}{N} + \frac{C_2}{N^2} + O(N^{-3}),$$
$$\mathbb{E}[f(\bar{X}_T^{2N})] - \mathbb{E}[f(X_T)] = \frac{C_1}{2N} + \frac{C_2}{4N^2} + O(N^{-3})$$

so that

$$2\mathbb{E}[f(\bar{X}_T^{2N})] - \mathbb{E}[f(\bar{X}_T^N)] - \mathbb{E}[f(X_T)] = -\frac{C_2}{2N^2} + O(N^{-3}).$$

Then, the new global quadratic error becomes:

$$\begin{aligned}\|\mathbb{E}[f(X_T)] - \frac{1}{M}\sum_{k=1}^{M}(2f(\bar{X}_T^{2N,(k)}) - f(\bar{X}_T^{N,(k)}))\|_{L^2(\mathbb{P})}^2 \\ = \left(\frac{C_2}{2N^2}\right)^2 + \frac{\mathrm{Var}(2f(\bar{X}_T^{2N}) - f(\bar{X}_T^N))}{M} + O(N^{-5})\end{aligned}$$

1. Reduce the complexity of the Monte Carlo estimator?
2. Is it possible to control the variance $\mathrm{Var}(2f(\bar{X}_T^{2N}) - f(\bar{X}_T^N))$ as $N \uparrow \infty$?

◦ Lazy simulation: If one takes W^1 and W^2 independent then, one has

$$\mathsf{Var}(2f(\bar{X}_T^{2N}) - f(\bar{X}_T^N)) = 4\mathsf{Var}(f(\bar{X}_T^N)) + \mathsf{Var}(f(\bar{X}_T^N)) \to 5\mathsf{Var}(f(X_T)),\ N \uparrow \infty$$

so that the variance increases by a factor of 5.
◦ Consistent simulation of Brownian paths: If $W^2 = W^1 = W$ then

$$\mathsf{Var}(2f(\bar{X}_T^{2N}) - f(\bar{X}_T^N)) \to \mathsf{Var}(2f(X_T) - f(X_T)) = \mathsf{Var}(f(X_T)),\ N \uparrow \infty.$$

◦ Practical aspects of implementation: From a practical point of view, one first simulates an Euler scheme with time step $\Delta/2$ using Gaussian i.i.d. r.v. $(U_k^2)_{1\leq k\leq 2N}$, then one simulates the r.v. $(U_k^1)_{1\leq k\leq N}$:

$$U_k^1 := \frac{U_{2k}^2 + U_{2k-1}^2}{\sqrt{2}},\ k = 1, \cdots, N.$$

◦ Complexity consideration: In order to achieve a global $L^2(\mathbb{P})$-error

$$\|\mathbb{E}[f(X_T)] - \frac{1}{M}\sum_{k=1}^{M}(2f(\bar{X}_T^{2N,(k)}) - f(\bar{X}_T^{N,(k)}))\|_{L^2(\P)} \approx \varepsilon$$

one has to choose $N \approx \varepsilon^{-\frac{1}{2}}$, $M \approx \varepsilon^{-2}$ and the complexity of the estimator is

$$C_{R-R} = C\varepsilon^{-2-\frac{1}{2}} << C_{MC} = C\varepsilon^{-3}, \quad \text{as } \varepsilon \downarrow 0.$$

◦ To take advantage of the full expansion at any order R, see G. Pagès (2007): *Multistep Richardson-Romberg extrapolation: controlling variance and complexity.*

Multi-level Monte Carlo simulation

○ Introduction of the method and its application in Finance is attributed to M. Giles (2008) generalizing an idea of A. Kebaier (2005).
○ The basis of the MLMC consists in using the following decomposition:

$$\mathbb{E}[f(\bar{X}_T^{h_L})] = \mathbb{E}[f(\bar{X}_T^{h_0})] + \sum_{\ell=1}^{L} \mathbb{E}[f(\bar{X}_T^{h_\ell}) - f(\bar{X}_T^{h_{\ell-1}})]$$

where $(\bar{X}_t^{h_\ell})_{0 \leq t \leq T}$ is an Euler scheme with time mesh h_ℓ, such that $h_\ell \to 0$, $\ell \uparrow \infty$. A common choice is $h_\ell = T/m^\ell$ for some positive integer m.
○ Then, one implements a standard Monte Carlo method on each level $\ell = 0, \cdots, L$:

$$E_{MC} = \frac{1}{M_0} \sum_{k=1}^{M_0} f(\bar{X}_T^{h_0,(k)}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{k=1}^{M_\ell} [f(\bar{X}_T^{h_\ell,(k)}) - f(\bar{X}_T^{h_{\ell-1},(k)})]$$

This requires

- L independent Monte Carlo estimators
- on each level ℓ, $(X_T^{h_\ell,(k)}, \bar{X}_T^{h_{\ell-1},(k)})_{1 \leq k \leq M_\ell}$ are i.i.d. with the same law as $(\bar{X}_T^{h_\ell}, \bar{X}_T^{h_{\ell-1}})$.

$\rightsquigarrow$Main idea: Balance the simulation effort on the different levels keeping in mind that most samples will be taken at level 0 and few samples will be required at the finest level L.

Assuming that f, b and σ are Lipschitz then the variance $V_\ell := \mathrm{Var}(f(\bar{X}_T^{h_\ell}) - f(\bar{X}_T^{h_{\ell-1}}))$ satisfies

$$V_\ell \leq 2(\mathbb{E}[(f(\bar{X}_T^{h_\ell}) - f(X_T))^2] + \mathbb{E}[(f(\bar{X}_T^{h_{\ell-1}}) - f(X_T))^2]) \leq 2C(1+|x|^2)h_\ell \to 0,$$

as $,\ell \uparrow \infty$. Hence, one needs fewer and fewer samples to accurately estimate the expectation $\mathbb{E}[f(\bar{X}_T^{h_\ell}) - f(\bar{X}_T^{h_{\ell-1}})]$.

◦ The variance of the multi-level Monte Carlo method is

$$\mathrm{Var}(E_{MLMC}) = \frac{1}{M_0}V_0 + \sum_{\ell=1}^{L} \frac{1}{M_\ell} V_\ell$$

◦ The computational cost of the multi-level Monte Carlo method is

$$C_{MLMC} = CM_0 h_0^{-1} + \sum_{\ell=1}^{L} CM_\ell(h_\ell^{-1} + h_{\ell-1}^{-1}) = C\sum_{\ell=0}^{L} M_\ell h_\ell^{-1}$$

For a fixed computational cost, the variance is minimized by solving

$$\min_{M_0,\cdots,M_L,\lambda}\left\{g(M_0,\cdots,M_L,\lambda):=\sum_{\ell=0}^{L}M_\ell^{-1}V_\ell+\lambda\left\{\sum_{\ell=0}^{L}M_\ell h_\ell^{-1}-C\right\}\right\}$$

One has $\nabla g(M_0,\cdots,M_L)=0$ iif $M_\ell^2=\lambda^{-1}h_\ell V_\ell\approx h_\ell^2$ so that M_ℓ is proportional to h_ℓ.
For a fixed ε, to get Var$(E_{MLMC})=C^{-1}\varepsilon^2$, one can choose

$$M_\ell=C\varepsilon^{-2}(L+1)h_\ell.$$

With this choice, the computational cost of the multi-level method is

$$C_{MLMC}=C\sum_{\ell=0}^{L}M_\ell h_\ell^{-1}=C\varepsilon^{-2}(L+1)^2$$

The bias is given by

$$|\mathbb{E}[f(X_T)-f(\bar{X}_T)]|\leq Ch_L=CTm^{-L}\approx\varepsilon\Leftrightarrow L=\frac{\log(\varepsilon^{-1})}{\log(m)}$$

To sum up, in order to get a precision of order ε, one has to choose the mesh size $h_\ell = T/m^\ell$, $M_\ell = C\varepsilon^{-2}(L+1)h_\ell$ and the total number of level $L = \log(\varepsilon^{-1})/\log(m)$ so that the total computational cost of the MLMC is

$$C_{MLMC} = C\varepsilon^{-2}\log(\varepsilon)^2 << C_{MC} = C\varepsilon^{-3} \text{ as } \varepsilon \downarrow 0.$$

More generally, one has the following result:

Theorem

If there exist positive constants $\alpha \geq 1/2$, β, c_1, c_2, c_3 such that

1. $|\mathbb{E}[f(X_T) - f(\bar{X}_T^{h_\ell})]| \leq c_1 h_\ell^\alpha$,
2. *Var*$(f(\bar{X}_T^{h_\ell}) - f(\bar{X}_T^{h_{\ell-1}})) \leq c_2 h_\ell^\beta$ *and Var*$(f(\bar{X}_T^{h_0})) \leq c_2 h_0^\beta$,
3. *The computational cost of level ℓ is: $C_\ell \leq c_3 M_\ell h_\ell^{-1}$.*

Then, for ε sufficiently small, there exist $M_\ell, \ell = 0, \cdots, L$; L and a constant c_4 such that $\mathbb{E}[(f(X_T) - E_{MLMC})^2] \leq \varepsilon^2$ with a computational cost satisfying:

$$C_{MLMC} \leq \begin{cases} c_4\varepsilon^{-2}, \ \beta > 1 \\ c_4\varepsilon^{-2}(\log(\varepsilon))^2, \ \beta = 1 \\ c_4\varepsilon^{-2-\frac{1-\beta}{2}}, \ \beta \in (0,1). \end{cases} \qquad M_\ell = \begin{cases} 2c_2\varepsilon^{-2}(1 - m^{-\frac{1-\beta}{2}})^{-1} h_\ell^{\frac{1+\beta}{2}}, \ \beta > 1 \\ 2c_2\varepsilon^{-2}(L+1)h_\ell, \ \beta = 1 \\ 2c_2\varepsilon^{-2}(1 - m^{\frac{1-\beta}{2}})^{-1} h_L^{-\frac{(1-\beta)}{2}} h_\ell^{\frac{1+\beta}{2}}, \ \beta \in (\end{cases}$$

with $L = \lfloor \log(\sqrt{2c_1 T^\alpha \varepsilon^{-1}})/\alpha \log(m) \rfloor$

Unbiased simulation of SDEs

Originally developed by

- Bally & Kohatsu-Higa (2015): *A probabilistic interpretation of the parametrix method*
- Anderson & Kohatsu-Higa (2017): *Unbiased simulation of stochastic differential equations using parametrix expansions*

For sake of simplicity, one dimensional case:

$$d = q = 1, \quad \sigma(t,x) = \sigma > 0, \quad b \in \mathcal{C}^1_b(\mathbb{R}).$$

The map $[0,T] \times \mathbb{R} \ni (s,x) \mapsto u(s,x) = \mathbb{E}[f(X^x_s)]$ is a solution to the PDE:

$$(\partial_s - \mathcal{L})u(s,x) = 0, \quad u(0,x) = f(x).$$

Let $\bar{X}^{s,x}_t = x + \sigma(W_t - W_s)$. Note that $\mathbb{E}[u(0, \bar{X}^x_T)] = \mathbb{E}[f(\bar{X}^{0,x}_T)]$. Apply Itô's rule to $u(T-t, \bar{X}^x_t)$:

$$\begin{aligned}
\mathbb{E}[f(\bar{X}^{0,x}_T)] &= u(T,x) + \int_0^T \mathbb{E}[(-\partial_t u(T-r, \bar{X}^{0,x}_r) + \frac{1}{2}\sigma^2 \partial^2_x u(T-r, \bar{X}^{0,x}_r)]dr \\
&= u(T,x) + \int_0^T \mathbb{E}[(-\mathcal{L} + \frac{1}{2}\sigma^2\partial^2_x) u(T-r, \bar{X}^{0,x}_r)]\, dr \\
&= u(T,x) - \int_0^T \mathbb{E}[b(\bar{X}^{0,x}_r)\partial_x u(T-r, \bar{X}^{0,x}_r)]\, dr
\end{aligned}$$

First step of the expansion

Then, remark that one has

$$\begin{aligned}\mathbb{E}[b(\bar{X}_r^{0,x})\partial_x u(T-r,\bar{X}_r^{0,x})] &= \int_{\mathbb{R}} b(y)\partial_x u(T-r,y)\frac{e^{-\frac{(y-x)^2}{2\sigma^2 r}}}{\sqrt{2\pi\sigma^2 r}}\,dy \\ &= -\int_{\mathbb{R}} \left\{b'(y)+b(y)H_1(\sigma^2 r, y-x)\right\} u(T-r,y)\frac{e^{-\frac{(y-x)^2}{2\sigma^2 r}}}{\sqrt{2\pi\sigma^2 r}}\,dy \\ &=: \mathbb{E}[\bar{\theta}_r(x,\bar{X}_r^{0,x})u(T-r,\bar{X}_r^{0,x})]\end{aligned}$$

with the notation $H_1(a,y) := -y/a$, $\bar{\theta}_r(x,y) := -b'(y) - b(y)H_1(\sigma^2 r, y-x)$.
Hence we get

$$u(T,x) = \mathbb{E}[f(\bar{X}_T^{0,x})] + \int_0^T \mathbb{E}[\bar{\theta}_r(x,\bar{X}_r^{0,x})u(T-r,\bar{X}_r^{0,x})]\,dr \qquad (5)$$

Remark that for $r \in [0,T]$ one has

$$\begin{aligned}|\mathbb{E}[\bar{\theta}_r(x,\bar{X}_r^{0,x})u(T-r,\bar{X}_r^{0,x})]| &\leq |f|_\infty \mathbb{E}[|\bar{\theta}_r(x,\bar{X}_r^{0,x})|] \\ &\leq \kappa \int_{\mathbb{R}} (1+\frac{|y-x|}{r})\frac{e^{-\frac{(y-x)^2}{2\sigma^2 r}}}{\sqrt{2\pi\sigma^2 r}}\,dy \leq \kappa r^{-\frac{1}{2}} \in L^1([0,T])\end{aligned}$$

Two steps of the expansion

Then, one just has to iterate using the same argument

$$\begin{aligned} u(T-r,x) &= \mathbb{E}[f(\bar{X}^{0,x}_{T-r})] + \int_0^{T-r} \mathbb{E}[\bar{\theta}_{r_1}(x,\bar{X}^{0,x}_{r_1})u(T-r-r_1,\bar{X}^{0,x}_{r_1})]\,dr_1 \\ &= \mathbb{E}[f(\bar{X}^{r,x}_{T})] + \int_r^{T} \mathbb{E}[\bar{\theta}_{r_1-r}(x,\bar{X}^{0,x}_{r_1-r})u(T-r_1,\bar{X}^{0,x}_{r_1-r})]\,dr_1 \\ &= \mathbb{E}[f(\bar{X}^{r,x}_{T})] + \int_r^{T} \mathbb{E}[\bar{\theta}_{r_1-r}(x,\bar{X}^{r,x}_{r_1})u(T-r_1,\bar{X}^{r,x}_{r_1})]\,dr_1 \end{aligned}$$

so that

$$\begin{aligned} \mathbb{E}[\bar{\theta}_r(x,\bar{X}^{0,x}_r)u(T-r,\bar{X}^{0,x}_r)] &= \mathbb{E}[f(\bar{X}^{0,x}_T)\bar{\theta}_r(x,\bar{X}^{0,x}_r)] \\ &\quad + \int_r^T \mathbb{E}[\bar{\theta}_r(x,\bar{X}^{0,x}_r)\bar{\theta}_{r_1-r}(\bar{X}^{0,x}_r,\bar{X}^{0,x}_{r_1})u(T-r_1,\bar{X}^{0,x}_{r_1})]\,dr_1 \end{aligned}$$

and plug the previous equality into (5) to get

$$\begin{aligned} u(T,x) &= \mathbb{E}[f(\bar{X}^{0,x}_T)] + \int_0^T \mathbb{E}[f(\bar{X}^{0,x}_T)\bar{\theta}_{r_1}(x,\bar{X}^{0,x}_{r_1})]\,dr_1 \\ &\quad + \int_0^T\int_{r_1}^T \mathbb{E}[\bar{\theta}_{r_1}(x,\bar{X}^{0,x}_{r_1})\bar{\theta}_{r_2-r_1}(\bar{X}^{0,x}_{r_1},\bar{X}^{0,x}_{r_2})u(T-r_2,\bar{X}^{0,x}_{r_2})]\,dr_1\,dr_2 \end{aligned}$$

N-step of the expansion and convergence of the series

Repeating N times the same arguments, we can prove by induction (left as an exercice)

$$u(T,x) = \sum_{n=0}^{N} \int_{\Delta_n(T)} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})]\, dr_1 \cdots dr_n \tag{6}$$
$$+ \int_{\Delta_{N+1}(T)} \mathbb{E}[u(T - r_{N+1}, \bar{X}_T^{0,x}) \prod_{k=1}^{N} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})]\, dr_1 \cdots dr_{N+1}$$

where $\Delta_n(T) := \{(r_1, \cdots, r_n) \in [0,T]^n : 0 \leq r_1 \leq r_2 \leq \cdots \leq r_n \leq T\}$ and the convention $r_0 = 0$ and $\prod_\emptyset \cdots = 1$.

○ The idea now is to let $N \uparrow \infty$ in the previous identity. To do this, we need the following estimate

$$\forall p \geq 0,\ \mathbb{E}[|\bar{\theta}_{t-s}(\bar{X}_s^{0,x}, \bar{X}_t^{0,x})|^p | \mathcal{F}_s] \leq \kappa \sup_{x \in \mathbb{R}} \int_{\mathbb{R}} (1 + \frac{|y-x|^p}{r^p}) \frac{e^{-\frac{(y-x)^2}{2\sigma^2 r}}}{\sqrt{2\pi\sigma^2 r}}\, dy \leq \kappa r^{-\frac{p}{2}} \tag{7}$$

Convergence of the series

The estimate (7) yields

$$
\begin{aligned}
\left|\int_{\Delta_n(T)} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})]\, dr_1 \cdots dr_n\right| &\leq \kappa^n \int_{\Delta_n(T)} \prod_{k=1}^{n} (r_k - r_{k-1})^{-\frac{1}{2}} \\
&= \kappa^n T^{\frac{n}{2}} \frac{\Gamma^n(\frac{1}{2})}{[1 + \frac{n}{2}]!}
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\left|\int_{\Delta_{N+1}(T)} \mathbb{E}[u(T - r_{N+1}, \bar{X}_T^{0,x}) \prod_{k=1}^{N} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})]\, dr_1 \cdots dr_{N+1}\right| \\
&\qquad \leq \kappa^{N+1} T^{\frac{N+1}{2}} \frac{\Gamma^{N+1}(\frac{1}{2})}{[1 + \frac{N+1}{2}]!}
\end{aligned}
$$

so that one may let N goes to infinity in (6) to get

$$
u(T, x) = \sum_{n \geq 0} \int_{\Delta_n(T)} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})]\, dr_1 \cdots dr_n
$$

Probabilistic representation of the series

In order to get a probabilistic representation of the series, one remarks that for a Poisson process N with intensity λ, independent of W, when $N_T = n$, its jump times $\zeta_1, \cdots, \zeta_n$ are distributed as the order statistics of n i.i.d. uniform r.v on $[0, T]$, that is

$$\mathbb{P}(N_T = n, \zeta_1 \in dr_1, \cdots, \zeta_n \in dr_n) = \lambda^n e^{-\lambda T} \, dr_1 \cdots dr_n \text{ on } \Delta_n(T).$$

As a consequence,

$$\begin{aligned}\int_{\Delta_n(T)} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \bar{\theta}_{r_k - r_{k-1}}(\bar{X}_{r_{k-1}}^{0,x}, \bar{X}_{r_k}^{0,x})] \, dr_1 \cdots dr_n \\ = e^{\lambda T} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \lambda^{-1} \bar{\theta}_{\zeta_k - \zeta_{k-1}}(\bar{X}_{\zeta_{k-1}}^{0,x}, \bar{X}_{\zeta_k}^{0,x}) \mathbf{1}_{\{N_T = n\}}]\end{aligned}$$

and

$$\begin{aligned}u(T, x) &= \sum_{n \geq 0} e^{\lambda T} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{n} \lambda^{-1} \bar{\theta}_{\zeta_k - \zeta_{k-1}}(\bar{X}_{\zeta_{k-1}}^{0,x}, \bar{X}_{\zeta_k}^{0,x}) \mathbf{1}_{\{N_T = n\}}] \\ &= e^{\lambda T} \mathbb{E}[f(\bar{X}_T^{0,x}) \prod_{k=1}^{N_T} \lambda^{-1} \bar{\theta}_{\zeta_k - \zeta_{k-1}}(\bar{X}_{\zeta_{k-1}}^{0,x}, \bar{X}_{\zeta_k}^{0,x})].\end{aligned}$$

Pros:

- The previous probabilistic representation allows to compute $u(T,x) = \mathbb{E}[f(X_T^{0,x})]$ without any discretization error but only a statistical error! The complexity of the method is thus $\kappa\varepsilon^{-2}$ for a given error of order ε.
- The previous probabilistic representation can be generalized to the multi-dimensional case and to the case $\sigma(t,x) \neq \sigma$ without any difficulty under the assumptions that $x \mapsto a(x) = \sigma(x)\sigma(x)^T \in \mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R}^{d\times d})$ and $a(x) \geq \lambda I_d$. (uniform ellipticity)

Cons:

- The variance of the Monte Carlo estimator associated to the previous probabilistic representation is infinite! Actually, the estimate (7) shows that $L^p(\mathbb{P})$-moments for $p \in [0,2)$ are finite but in general the variance of the estimator is too large.
  - See the importance sampling procedure of Anderson and Kohatsu-Higa (2017) to partially circumvent this issue.
- Numerical experiments show a good behavior for b, σ and T not too large but bad behavior otherwise.

Thank you!