Fiche
Numerical Probability for Finance

Antoine Falck

December 11, 2017

Contents

1 Simulation of random variables

1.1 Pseudo-random numbers

With the Monte Carlo methods and other types of application we have to generate random numbers in order to make simulations. So how can we generate random numbers from nothing ?

A lot of mathematicians tried very different methods but the congruential induction is one of the fundamental methods to simulate pseudo-random numbers. One considers the sequence (x_n) of so called pseudo-random numbers defined by:

$$x_n \quad = \quad \frac{y_n}{N}\ , \qquad y_n \in \{0, \dots, N-1\}. \tag{1.1}$$

And we generate the integers y_n by congruential induction

$$y_{n+1} \quad \equiv \quad a y_n \ + \ b \quad \mathrm{mod}\ N, \tag{1.2}$$

where $\gcd(a, N) = 1$. But there are other requirements that the parameters N, a and b must satisfy to properly use this method.

This introductory just aims to give the flavour of complexity to generate random numbers. One recent development is the family of Mersenne twister generators. The first level denoted `MT-p` are congruential generators whose period N_p is a prime Mersenne number, *i.e.* $N_p = 2^p - 1$ where p is a prime number. The most popular[1] is `MT-19937`.

1.2 Fundamental principle of simulation

Theorem 1.1 (Fundamental principle of simulation)**.** *Let (E, d) be a Polish space (complete and separable) and $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{B}(E))$ a r.v. with distribution $\mathbb{P}_X$. There exists $\varphi : ([0,1], \mathcal{B}([0,1]), \lambda) \to (E, \mathcal{B}(E))$ measurable s.t.*

$$\mathbb{P}_x \quad = \quad \lambda \circ \varphi^{-1}\ ,$$

where λ is the Lebesgue measure.

This theorem is practically of little use, it just says that any r.v. can be simulated juste with the uniform distribution on $[0,1]$, *i.e.*

$$X \sim \varphi(U)\ ,$$

where $U \sim \mathcal{U}([0,1])$.

1.3 The (inverse) distribution function method

Let μ be a probability distribution with distribution function F, $\forall x \in \mathbb{R}$,

$$F(x) \quad = \quad \mu([-\infty, x]).$$

[1] See `http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/SFMT/`.

One can associate its canonical left inverse F_l^{-1} on the open $]0,1[$,

$$F_l^{-1}(u) \quad = \quad \inf\{x : F(x) \geq u\},$$

for all $u \in]0,1[$.

Proposition 1.2. *If $U \sim \mathcal{U}([0,1])$, then $X := F_l^{-1}(U) \sim \mu$.*

Proof. Let $x \in \mathbb{R}$, $\{X \leq x\} = \{F_l^{-1}(U) \leq x\} = \{U \leq F(x)\}$. So that,

$$\mathbb{P}\{X \leq x\} \quad = \quad \mathbb{P}\{U \leq F(x)\} \ = \ F(x).$$

□

Example 1.1 (Exponential distribution)**.** Let $X \sim \mathcal{E}(\lambda)$, $\lambda > 0$, then

$$F_X(x) \quad = \quad \lambda \int_0^x e^{-\lambda t}\, \mathrm{d}t \ = \ 1 - e^{-\lambda x}\ , \quad \forall x \in]0, \infty].$$

Then, for all $y \in]0,1[$, $F_X^{-1} = -\frac{\ln(1-u)}{\lambda}$. Now with $U \sim \mathcal{U}([0,1])$, $1 - U \sim \mathcal{U}([0,1])$ and

$$X \quad := \quad -\frac{\ln U}{\lambda} \ \sim \ \mathcal{E}(\lambda).$$

Example 1.2 (Cauchy distribution)**.** We know that for $\mathcal{C}(c)$, $c > 0$, $\mathbb{P}_X(\mathrm{d}x) = \frac{c}{\pi(x^2+c^2)}\, \mathrm{d}x$. So,

$$F_X(x) \quad = \quad \frac{1}{\pi}\left(\arctan\left(\frac{\pi}{2}\right) + \frac{\pi}{2}\right)\ , \quad \forall x \in \mathbb{R}.$$

Hence, $F_X^{-1} = c \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$ and,

$$X \quad := \quad c \tan\left(\pi\left(U - \frac{1}{2}\right)\right) \ \sim \ \mathcal{C}(c).$$

Example 1.3 (Bernouilli r.v.)**.** Let $p \in]0,1[$, then,

$$X \quad := \quad \mathbb{1}_{\{U \leq p\}} \ \sim \ \mathcal{B}(p).$$

1.4 The acceptance-rejection method

Let μ be a non-negative measure on $(E, \mathcal{E})$ and let $f, g : (E, \mathcal{E}) \to \mathbb{R}_+$. Assume that $f \in L^1$ with $\int_E f \mathrm{d}\mu > 0$ and that g is probability density on μ satisfying $g > 0$ μ-a.s. and there exists a constant $c > 0$ s.t.

$$f(x) \quad \leq \quad c\, g(x)\ , \quad \mu\text{-a.e.}$$

Note that this implies that $c \geq \int_E f \mathrm{d}\mu$.

From a practical point of view we have to:

- know the value of the constant c ;
- $Y \sim g\mu$ can be simulated with a reasonable cost ;
- we can compute $\frac{f}{g}$ again at a reasonable cost.

Let h be a test function, X and Y r.v. with distributions respectively ν and $g.\mu$.

$$\begin{aligned}\mathbb{E}[h(X)] &= \frac{1}{\int_E f\mathrm{d}\mu}\int_E h(x)f(x)\mu(\mathrm{d}x)\\ &= \frac{1}{\int_E f\mathrm{d}\mu}\int_E h(y)\frac{f}{g}(y)g(y)\mu(\mathrm{d}y)\\ &= \mathbb{E}\left[h(Y)\frac{f}{g}(Y)\right].\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}[h(X)] &= \frac{c}{\int_E f\mathrm{d}\mu}\int_E h(y)\left(\int_0^1 \mathbb{1}_{\{u\leq \frac{1}{c}\frac{f}{g}(y)\}}\mathrm{d}u\right)g(y)\mu(\mathrm{d}y)\\ &= \frac{c}{\int_E f\mathrm{d}\mu}\int_E\int_0^1 h(y)\mathbb{1}_{\{u\leq \frac{1}{c}\frac{f}{g}(y)\}}g(y)\mu(\mathrm{d}y)\mathrm{d}u\\ &= \frac{c}{\int_E f\mathrm{d}\mu}\mathbb{E}\left[h(Y)\mathbb{1}_{\{U\leq \frac{1}{c}\frac{f}{g}(Y)\}}\right],\end{aligned}$$

where $U\sim\mathcal{B}([0,1])$ and independent from Y.

By considering $h\equiv 1$ we derive $\mathbb{P}\{cU\leq \frac{f}{g}(Y)\} = \frac{\int_E f\mathrm{d}\mu}{c}$. Finally we show that,

$$\mathbb{E}[h(X)] = \mathbb{E}\left[h(Y)\mid\left\{cU\leq\frac{f}{g}(Y)\right\}\right],$$

where $U\sim\mathcal{U}([0,1])$.

Proposition 1.3 (Acceptance-rejection simulation method)**.** *Let $(U_n,Y_n)_{n\geq 1}$ be a sequence of i.i.d. r.v. with distribution $\mathcal{U}([0,1])\otimes\mathbb{P}_Y$ defined on $(\Omega,\mathcal{A},\mathbb{P})$ where $\mathbb{P}(\mathrm{d}y)=g(y)\mu(\mathrm{d}y)$ is the distribution of Y. Set*

$$\tau := \min\{k\geq 1 : cU_k g(Y_k)\leq f(Y_k)\}.$$

Then τ has a geometric distribution $\mathcal{G}^*(p)$ with parameter $p=\mathbb{P}\{cU_1 g(Y_1)\leq f(Y_1)\}$ and

$$X := Y_\tau \sim \nu.$$

1.5 Simulation of Gaussian r.v.

1.5.1 Box-Muller method

Proposition 1.4. *Let R^2 and Θ be two r.v. with distributions respectively $\mathcal{E}(\frac{1}{2})$ and $\mathcal{B}([0,2\pi])$. Then,*

$$X := (R\cos\Theta, R\sin\Theta) \sim \mathcal{N}(0,I_2).$$

Proof. Let φ be a test function (bounded).

$$\begin{aligned}A &:= \iint_{\mathbb{R}^2}\varphi(x_1,x_2)\exp\left(-\frac{x_1^2+x_2^2}{2}\right)\frac{\mathrm{d}x_1\mathrm{d}x_2}{2\pi}\\ &= \iint\varphi(\rho\cos\theta,\rho\sin\theta)e^{-\frac{\rho^2}{2}}\mathbb{1}_{\mathbb{R}_+^*}(\rho)\mathbb{1}_{[0,2\pi]}(\theta)\rho\frac{\mathrm{d}\rho\mathrm{d}\theta}{2\pi}\\ &= \iint\varphi(\sqrt{r}\cos\theta,\sqrt{r}\sin\theta)\frac{e^{-\frac{r}{2}}}{2}\mathbb{1}_{\mathbb{R}_+^*}(r)\mathbb{1}_{[0,2\pi]}(\theta)\frac{\mathrm{d}\rho\mathrm{d}\theta}{2\pi}\\ &= \mathbb{E}\left[f\left(\sqrt{R^2}\cos\Theta;\sqrt{R^2}\sin\Theta\right)\right].\end{aligned}$$

Or *de facto* $A=\mathbb{E}[f(X_1,X_2)]$. □

On a avec $U_1,U_2\sim\mathcal{U}([0,1])$,

$$(X_1,X_2) = \sqrt{-2\ln U_1}\,(\cos(2\pi U_1),\sin(2\pi U_2))$$

1.5.2 Massaglia method

Let $(V_1,V_2)\sim\mathcal{U}(\mathcal{B}(0,1))$, where $\mathcal{B}(0,1)$ is the canonical euclidean unit ball in $\mathbb{R}^2$. Set $R^2 := V_1^2+V_2^2$ and,

$$X := \left(V_1\sqrt{\frac{-2\ln R^2}{R^2}}, V_2\sqrt{\frac{-2\ln R^2}{R^2}}\right).$$

We show that $R^2\sim\mathcal{U}([0,1])$, so $\sqrt{-2\ln R^2}\sim\mathcal{E}\left(\frac{1}{2}\right)$ and $\left(\frac{V_1}{R},\frac{V_2}{R}\right)\sim(\cos\Theta,\sin\Theta)$. We conclude with Box-Muller.

1.6 Simulation of Poisson distributions

The Poisson distribution with parameter $\lambda>0$, denoted $\mathcal{P}(\lambda)$, is an integral valued probability measure analytically defined by

$$\mathcal{P}(\lambda)(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!},$$

for all $k\in\mathbb{N}$.

To simulate this distribution in an exact way, one relies on its close connection with the Poisson counting process. The (normalized) Poisson counting process is the counting process induced by the Exponential random walk (with parameter 1). It is defined by

$$N_t = \sum_{n\geq 1}\mathbb{1}_{S_k\leq t},$$

for all $t\geq 0$, where $S_n = X_1+\cdots+X_n$, $(X_i)_i$ an iiid sequence of r.v. with distribution $\mathcal{E}(1)$.

2 Monte Carlo method and applications to option pricing

2.1 The Monte Carlo method

The Monte Carlo method is based the Strong Law of Large Numbers (SLLN) and its implementation on a computer. If $X_1,\ldots,X_M$ is a sequence of independent copies of an integrable r.v. X then,

$$\begin{aligned}\bar{X}_M(\omega) &:= \frac{X_1(\omega)+\cdots+X_M(\omega)}{M}\\ &\xrightarrow[M\to\infty]{} \mathbb{E}[X_1] =: m_X.\end{aligned}$$

Remark. To implement the SLLN on a computer one needs to be able to generate pseudo-random numbers, $U\sim\mathcal{U}([0,1])$; and then to be able to "represent" X.

The weak rate of convergence of the SLLN is ruled by the Central Limit Theorem (CLT) which says that if X is square integrable then,

$$\sqrt{M}(\bar{X}_M - m_X) \quad \xrightarrow[M\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,\sigma_X^2) \ ,$$

where $\sigma_X^2 = \mathrm{Var}[X]$.

Remark. It shows the main drawback of the Monte Carlo method: it is a slow method since dividing the error by 2 needs to increase the size of the simulation by 4.

We can then try to control the error with a confidence interval. Assume that $\sigma_X > 0$, the CLT also reads,

$$\sqrt{M}\frac{\bar{X}_M - m_X}{\sigma_X} \quad \xrightarrow[M\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1).$$

So we have with $a < b$, as the normal density has no atom,

$$\begin{aligned}\lim_{M\to\infty} \mathbb{P}\left\{\sqrt{M}\frac{\bar{X}_M - m_X}{\sigma_X} \in [a,b]\right\} &= \mathbb{P}\{\mathcal{N}(0,1) \in [a,b]\} \\ &= \Phi(b) - \Phi(a)/,\end{aligned}$$

where Φ denotes the distribution function of the standard normal distribution. So now we can design a probabilistic control of the error directly derived from statistical concepts: let $\alpha \in]0,1[$ denote a confidence level (close to 1) and let q_α be the two-sided α-quantile defined as the unique solution to the equation

$$\begin{aligned} & \mathbb{P}\{|\mathcal{N}(0,1)| \le q_\alpha\} = \alpha \\ \Leftrightarrow \quad & 2\Phi(q_\alpha) - 1 = \alpha.\end{aligned}$$

One defines the theoretical random confidence interval at level

$$J_M^\alpha \quad := \quad \left[\bar{X}_M - q_\alpha\frac{\sigma_X}{\sqrt{M}}, \bar{X}_M + q_\alpha\frac{\sigma_X}{\sqrt{M}}\right] \ ,$$

which satisfies,

$$\begin{aligned}\mathbb{P}\{m_X \in J_M^\alpha\} \quad &= \quad \mathbb{P}\left\{\left|\sqrt{M}\frac{\bar{X}_M - m_X}{\sigma_X}\right| \le q_\alpha\right\} \\ &\xrightarrow[M\to\infty]{} \quad \mathbb{P}\{|\mathcal{N}(0,1)| \le q_\alpha\} = \alpha.\end{aligned}$$

However, at this stage this procedure remains purely theoretical since the confidence interval J_M involves the standard deviation σ_M^2 of X which is usually unknown.

At this stage we will then evaluate this variance with the unbiased canonical estimator of the variance:

$$\begin{aligned}\bar{V}_M \quad &= \quad \frac{1}{M-1}\sum_{k=1}^{M}(X_k - \bar{X}_M)^2 \\ &= \quad \frac{1}{M-1}\sum_{k=1}^{M}X_k^2 \ - \ \frac{M}{M-1}\bar{X}_M^2.\end{aligned}$$

So $\mathbb{E}[\bar{V}_X] = \sigma_X^2$, and with the Slutsky theorem,

$$\begin{aligned}\sqrt{M}\frac{\bar{X}_M - m_X}{\sqrt{\bar{V}_X}} \quad &= \quad \sqrt{M}\frac{\bar{X}_M - m_X}{\sigma_X^2}\frac{\sigma_X^2}{\sqrt{\bar{V}_X}} \\ &\xrightarrow{\mathcal{L}} \quad \mathcal{N}(0,1).\end{aligned}$$

Remark. In fact $\bar{V}_X \sim \chi^2(M-1)$, and so $\sqrt{M}\frac{\bar{X}_M - m_X}{\sqrt{\bar{V}_X}} \sim \mathcal{T}(M-1)$; but with M big the Student law is like the standard Gaussian. And in numerical probability $M \approx 10^6$ so this approximation makes sense.

Finally, one defines the confidence interval at level α of the Monte Carlo simulation by

$$I_M^\alpha \quad := \quad \left[\bar{X}_M - q_\alpha\sqrt{\frac{\bar{V}_X}{M}}, \bar{X}_M + q_\alpha\sqrt{\frac{\bar{V}_X}{M}}\right].$$

2.2 Vanilla option pricing in a Black-Scholes model

Let us consider a 2-dimensional correlated Black & Scholes model under its risk neutral probability. The non risky asset is defined by

$$\mathrm{d}X_t^0 \quad = \quad rX_t^0 \ \mathrm{d}t \ , \quad X_0^0 = 1 \ ,$$

is the capitalisation of \$1 at the bank. Where r is the interest rate. And two risky assets,

$$\begin{aligned}\mathrm{d}X_t^i \quad &= \quad X_t^i(r \ \mathrm{d}t + \sigma_i \ \mathrm{d}W_t^i) \ , \quad i = 1,2 \\ X_0^i \quad &= \quad x_0^i.\end{aligned}$$

$W = (W^1, W^2)$ denotes a correlated brownian motion, $\mathrm{d}\langle W^1, W^2\rangle = \rho \ \mathrm{d}t$, with $\rho \in [-1,1]$. Then $W_t^2 = \rho W_t^1 + \sqrt{1-\rho^2}\tilde{W}_t^2$ defines the standard 2-dimensional brownian motion $(W_t^1, \tilde{W}_t^2)$. We also define the filtration $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$.

So for all $t \in [0,T]$,

$$\begin{aligned}X_t^0 \quad &= \quad e^{rt} \\ X_t^i \quad &= \quad x_0^i e^{\left(r - \frac{\sigma_i^2}{2}\right)t + \sigma_i W_t^i} \ , \quad i = 1,2\end{aligned}$$

A European vanilla option with maturity $T > 0$ is an option related to a European payoff $h_T := h(X_T)$ which only depends on X at time T. In such a complete market the option premium at time 0 is given by

$$V_0 \quad = \quad e^{-rT}\mathbb{E}[h(X_T)] \ ,$$

and more generally at any time $t \in [0,T]$,

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

The fact that W has independent stationary increments implies that X_1 and X_2 have independent stationary ratios *i.e.*

$$\frac{X_T^i}{X_t^i} \quad = \quad \frac{X_{T-t}^i}{x_0^i} \perp\!\!\!\perp \mathcal{F}_t.$$

As a consequence if $V_0 := v(x_0, T)$, then

$$\begin{aligned} V_t &= e^{-r(T-t)}\mathbb{E}[h(X_T)|\mathcal{F}_t] \\ &= e^{-r(T-t)}\mathbb{E}\left[h\left(\left(X_t^i\frac{X_T^i}{X_t^i}\right)_i\right)\middle|\mathcal{F}_t\right] \\ &= e^{-r(T-t)}\mathbb{E}\left[h\left(\left(x_t^i\frac{X_{T-t}^i}{x_0^i}\right)_i\right)\right] \\ &= v(x^i, T-t). \end{aligned}$$

One can then calculate V_0 and easily replace the parameters for V_t.

Example 2.1 (Best-of call)**.** Let us compute the Monte Carlo method for a best-of call with strike K, *i.e.* the payoff function

$$h_T = (\max(X_1, X_2) - K)^+ .$$

We need to write the payoff as a function of independent uniformly distributed random variables, or equivalently as a function of independent random variables that are simple functions of independent uniformly distributed random variables, namely a centered and normalized Gaussian. In our case, it amounts to writing

$$\begin{aligned} e^{-rT} &= \varphi(Z^1, Z^2) \\ &:= \Bigg(\max\Bigg(x_0^1 e^{-\frac{\sigma_1^2}{2}T+\sigma_1\sqrt{T}Z^1}, \\ &\qquad x_0^2 e^{-\frac{\sigma_2^2}{2}T+\sigma_1\sqrt{T}(\rho Z^1+\sqrt{1-\rho^2}Z^2)}\Bigg) - Ke^{-rT}\Bigg)^+ , \end{aligned}$$

where $Z = (Z^1, Z^2) \sim \mathcal{N}(0, \mathbf{I}_2)$. Then, simulating a M-sample $(Z_m)_{1\le m\le M}$ of the $\mathcal{N}(0, \mathbf{I}_2)$ distribution using *e.g.* the Box–Müller yields the estimate

$$\begin{aligned} \textit{Best-of Call}_0 &= \mathbb{E}[\varphi(Z^1, Z^2)] \\ &\approx \bar{\varphi}_M := \frac{1}{M}\sum_{m=1}^{M}\varphi(Z_m). \end{aligned}$$

One computes an estimate for the variance using the same sample

$$\bar{V}_m(\varphi) = \frac{1}{M-1}\sum_{m=1}^{M}\varphi^2(Z_m) - \frac{M}{M-1}\bar{\varphi}_M^2.$$

2.3 Greeks: a first approach

2.3.1 Background on differentiation of function defined by an integral

Theorem 2.1 (Interchanging differentiation and expectation)**.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let I be a non-trivial interval of $\mathbb{R}$. Let $\varphi : I \times \Omega \to \mathbb{R}$ be a $\mathcal{B}or(I) \otimes \mathcal{A}$-measurable function.*

(a) Local version. *Let $x_0 \in I$. If the function φ satisfies:*

(i) for every $x \in I$, the random variable $\varphi(x, \cdot) \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$;

(ii) $\mathbb{P}(d\omega)$-a.s. $\frac{\partial\varphi}{\partial x}(x_0, \omega)$ exists ;

(iii) There exists $Y \in L^1_{\mathbb{R}_+}(\mathbb{P})$ such that, for every $x \in I$,

$$P(d\omega)\text{-}a.s. \quad |\varphi(x, \omega) - \varphi(x_0, \omega)| \le Y(\omega)|x - x_0| ,$$

then the function $\Phi(x) := \mathbb{E}[\varphi(x, \cdot)]$ is defined at every $x \in I$, differentiable at x_0 with derivative

$$\Phi'(x_0) = \mathbb{E}\left[\frac{\partial\varphi}{\partial x}(x_0, \cdot)\right].$$

(b) Global version. *If φ satisfies (i) and*

(ii) $\mathbb{P}(d\omega)$-a.s. $\frac{\partial\varphi}{\partial x}(x, \omega)$ exists at every $x \in I$;

(iii) There exists $Y \in L^1_{\mathbb{R}_+}(\mathbb{P})$ such that, for every $x \in I$,

$$P(d\omega)\text{-}a.s. \quad \left|\frac{\partial\varphi(x, \omega)}{\partial x}\right| \le Y(\omega) ,$$

then the function $\Phi(x) := \mathbb{E}[\varphi(x, \cdot)]$ is defined and differentiable at every $x \in I$, with derivative

$$\Phi'(x) = \mathbb{E}\left[\frac{\partial\varphi}{\partial x}(x, \cdot)\right].$$

Exercise 2.2. Let $Z \sim \mathcal{N}(0, 1)$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\varphi(x, \omega) = (x - Z(\omega))^+$ and $\Phi(x) = \mathbb{E}[\varphi(x - Z)^+$, $x \in \mathbb{R}$.

1. Show that Φ is differentiable on the real line and compute its derivative.

2. Show that if I denotes a non-trivial interval of $\mathbb{R}$. Show that if $\omega \in \{Z \in I\}$ (*i.e.* $Z(\omega) \in I$), the function $x \mapsto (x - Z(\omega))^+$ is never differentiable on the whole interval I.

Proof. 1. We have the set $\{\omega : \frac{\partial\varphi}{\partial x}(x_0, \omega)$ exists $= \{\omega : X(\omega) \neq x_0\}$. And $\mathbb{P}\{X \neq x_0\} = 1 - \mathbb{P}\{X = x_0\} = 1$ as X has a density. Furthermore, for all $x \in I$ we have $|\varphi(x, \omega) - \varphi(x_,\omega)| \le |x - x_0|$, so $Y \equiv \mathbb{1} \in L^1$.

2. Let $I = [a, b]$, with $-\infty < a < b < \infty$, and we have $\frac{\partial\varphi}{\partial x}(x_0, \omega) = \mathbb{1}_{\{x_0 \ge X(\omega)\}}$ iff $X(\omega) \neq x_0$. So we have the set $\{\omega : \frac{\partial\varphi}{\partial x}(x_0, \omega)$ exists $\forall x_0 \in I\} = \{X(\omega) \notin I\}$, and $\mathbb{P}\{X \notin I\} = 1$ would imply $\mathbb{P}\{X \in I\} = 0$ which is absurd. So we don't have the global derivate.

□

2.3.2 Working on the scenarios space

Let us consider the BS model,

$$\mathrm{d}X_t^x = X_t^x(r\ \mathrm{d}t + \sigma\ \mathrm{d}W_t)\ , \qquad X_0^x = x > 0\ ,$$

so we have $X_t^x = xe^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma W_t}$. And let us define for every $x \in]0,\infty[$,

$$\Phi(x) = \mathbb{E}[\varphi(X_t^x)]\ ,$$

where $\varphi : \mathbb{R}_+^* \to \mathbb{R}$. We will first work on the scenarii space $(\Omega, \mathcal{A}, \mathbb{P})$, because this approach contains the "seed" of methods that can be developed in much more general settings in which the SDE has no explicit solution like it has in the Black-Scholes model. On the other hand, as soon as an explicit expression is available for the density $p_T(x,y)$ of X_T^x, it is more efficient to use the next section 2.3.3.

Proposition 2.2. *(a) If $\varphi :]0,\infty[\to \mathbb{R}$ is differentiable and φ' has polynomial growth (i.e. $|\varphi'| \leq C(1+|x|^p)$ with $C \geq 0$, $p > 0$), then the function Φ is differentiable and for all $x > 0$,*

$$\Phi'(x) = \mathbb{E}\left[\varphi'(X_T^x)\frac{X_T^x}{x}\right]. \tag{2.1}$$

(b) If φ is differentiable outside a countable set and is locally Lipschitz continuous with polynomial growth in the following sense

$$\exists m > 0\ , \forall u,v \in \mathbb{R}_+^*\ , \quad |\varphi(u) - \varphi(v)| \leq C|u-v|(1+|u|^m+|v|^m)\ ,$$

then Φ is differentiable everywhere on $\mathbb{R}_+^*$ and Φ' is given by (2.1). (c) If φ simply a Borel function with polynomial growth, then Φ is still differentiable and for all $x > 0$,

$$\Phi'(x) = \mathbb{E}\left[\varphi(X_T^x)\frac{W_T}{x\sigma T}\right]. \tag{2.2}$$

Proof. □

2.3.3 Direct differentiation on the state space

In fact, one can also achieve similar computations directly on the state space of a family of random variables (or vectors) X_T^x (indexed by its stating value x), provided this random variable (or vector) has an explicit probability density $p_T(x,y)$ with respect to a reference measure $\mu(\mathrm{d}y)$ on the real line (or $\mathbb{R}^d$). In general $\mu = \lambda_d$, Lebesgue measure.

One may imagine in full generality that X_T^x depends on a parameter θ a real parameter of interest : thus, $X_T^x = X_T^x(\theta)$ may be the solution at time T to a stochastic differential equation which coefficients depend on θ.

$$\begin{aligned}\Phi(\theta) &= \mathbb{E}[\varphi(X_T^x(\theta))] \\ &= \int_{\mathbb{R}} \varphi(y)p_T(\theta,x,y)\mu(\mathrm{d}y).\end{aligned}$$

So that,

$$\begin{aligned}\Phi'(\theta) &= \int_{\mathbb{R}} \varphi(y)\frac{\partial p_T}{\partial\theta}(\theta,x,y)\mu(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \varphi(y)\frac{\frac{\partial p_T}{\partial\theta}(\theta,x,y)}{p_T(\theta,x,y)}p_T(\theta,x,y)\mu(\mathrm{d}y) \\ &= \mathbb{E}\left[\varphi(y)\frac{\partial \ln p_T}{\partial\theta}(\theta,x,X_T^x)\right].\end{aligned}$$

Of course, the above computations need to be supported by appropriate assumptions (domination, *etc.*) to justify interchange of integration and differentiation.

3 Variance reduction

3.1 The Monte Carlo method revisited

Recall that for the MC method the confidence interval is

$$I_{\alpha,M} = \left[\bar{X}_M \pm q_\alpha\sqrt{\frac{\bar{V}_M}{M}}\right]\ ,$$

with the notations of 2.1. If we associate the precision ε to the confidence level α we can deduce the minimum number of computations

$$M(\varepsilon,\alpha) = \frac{q_\alpha^2\mathrm{Var}[X]}{\varepsilon^2}$$

As a first conclusion, this shows that, a confidence level being fixed, the size of a Monte Carlo simulation grows linearly with the variance of X for a given accuracy and quadratically as the inverse of the prescribed accuracy for a given variance.

Usually, the problem appears as follows: there exists a random variable $\xi \in L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$ such that

(i) $\mathbb{E}[\xi]$ can be computed at a very low cost by a deterministic method (closed form, numerical analysis method) ;

(ii) the random variable $X - \xi$ can be simulated with the same cost (complexity) than X ;

(iii) the variance $\mathrm{Var}[X-\xi] < \mathrm{Var}[X]$.

Then the r.v.

$$X' := X - \xi + \mathbb{E}[\xi]$$

can be simulated at the same cost as X.

Definition 3.1 (Control variate)**.** A r.v. ξ satisfying (i)-(ii)-(iii) is called a control variate for X.

3.1.1 Jensen's inequality and variance reduction

Proposition 3.1 (Jensen's Inequality)**.** *Let X be a random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a convex function. Suppose X and $g(X)$ are integrable. Then, for any subfield $\mathcal{B}$ of $\mathcal{A}$,*

$$g\left(\mathbb{E}[X|\mathcal{B}]\right) \;\leq\; \mathbb{E}\left[g(X)|\mathcal{B}\right] \;,\quad \mathbb{P}-a.s.$$

Proof. Immediate with the following characterization of a convex function:

$$g(x) \;=\; \sup_{\substack{a,b\in\mathbb{Q}\\ \varphi_{a,b}\leq g}} \phi_{a,b}(x) \;,$$

for all $x \in \mathbb{R}$, where $\varphi_{a,b}(x) = ax + b$. □

Example 3.1 (Basket or index option)**.** We consider a payoff on a basket of d (positive) risky assets (this basket can be an index). For the sake of simplicity we suppose it is a call with strike K *i.e.*

$$h_T \;=\; \left(\sum_{k=1}^{d} \alpha_i X_T^i \;-\; K\right)^+$$

where $(X^1, \ldots, X^d)$ models the price of d traded risky assets on a market and the α_k are some positive ($\alpha_i > 0$) weights satisfying $\sum_{i=1}^d \alpha_i = 1$. Then the convexity of the exponential implies that

$$0 \;\leq\; e^{\sum_{k=1}^d \alpha_i \ln X_T^i} \;\leq\; \sum_{k=1}^{d} \alpha_i X_T^i$$

so that

$$h_T \;\geq\; k_T := \left(e^{\sum_{k=1}^d \alpha_i \ln X_T^i}\right)^+ \;\geq 0$$

The correlated d-dimensional Black-Scholes model (under the risk-neutral probability measure with $r > 0$ denoting the interest rate) can be defined by the following system of SDE's which governs the price of d risky assets denoted $i \in [\![1, d]\!]$:

$$\mathrm{d}X_t^i \;=\; X_t^i\left(r\ \mathrm{d}t + \sum_{j=1}^{q} \sigma_{ij}\ \mathrm{d}W_t^j\right) \;,$$

where $W = (W^1, \ldots, W^q)$ is a standard q-dimensional Brownian motion and $\sigma = [\sigma_{ij}]_{\substack{1\leq i\leq d\\ 1\leq j\leq q}}$ is a is given by $d \times q$ matrix with real entries. Its solution

$$X_t^{i,x_i} \;=\; x_i \exp\left(\left(r - \frac{\sigma_{i\cdot}^2}{2}\right) t + \sum_{j=1}^{q} \sigma_{ij}\, \mathrm{d}W_t^j\right) \;,$$

where $\sigma_{i\cdot}^2 = \sum_{j=1}^q \sigma_{ij}^2$ for $i \in [\![1, d]\!]$.

We will now show that $\xi := e^{-rT} k_T$ is a pseudo-control variate.

Example 3.2 (Asian options and Kemna-Vorst control variate)**.** Let

$$h_T \;=\; \varphi\left(\frac{1}{T}\int_0^T X_T^x \,\mathrm{d}t\right)$$

be a generic Asian payoff where φ is non-negative, non-decreasing function defined on $\mathbb{R}_+$ and let X_t^x follow a regular Black-Scholes dynamic where we note $\mu := r - \frac{\sigma^2}{2}$. Then the Jensen inequality implies

$$\begin{aligned}\frac{1}{T}\int_0^T X_t^x \,\mathrm{d}t \;&\geq\; x \exp\left(\frac{1}{T}\int_0^T \mu t + \sigma W_t \,\mathrm{d}t\right)\\ &=\; x \exp\left(\mu\frac{T}{2} + \sigma\frac{1}{T}\int_0^T W_t \,\mathrm{d}t\right).\end{aligned}$$

We can show that

$$\frac{1}{T}\int_0^T W_t \,\mathrm{d}t \;\sim\; \mathcal{N}\left(0, \frac{T}{3}\right).$$

So we want to write the right hand side of the equality in a BS asset style *i.e.*

$$xe^{\alpha T} \exp\left(\left(r - \frac{\frac{\sigma^2}{3}}{2}\right) T + \sigma\frac{1}{T}\int_0^T W_t \,\mathrm{d}t\right)$$

where $\alpha = -\left(\frac{r}{2} + \frac{\sigma^2}{12}\right)$. This naturally leads to introduce the so-called Kemna-Vorst (pseudo-)control variate

$$k_T^{KV} \;:=\; \varphi\left(xe^{\alpha T} \exp\left(\left(r - \frac{\sigma^2}{3}\frac{1}{2}\right) T + \sigma\frac{1}{T}\int_0^T W_t \,\mathrm{d}t\right)\right)$$

3.1.2 Antithetic method

In this section we assume that X and X' have not only the same expectation m_X but also the same variance, *i.e.* $\mathrm{Var}[X] = \mathrm{Var}[X']$, and can be simulated with the same complexity $\kappa = \kappa_X = \kappa_{X'}$. In such a situation, choosing between X or X' may seem *a priori* a question of little interest but we can take advantage of this. Set

$$\chi \;:=\; \frac{X + X'}{2} \,,$$

it is reasonable to think that $\kappa_\chi \approx 2\kappa$. And we have

$$\mathrm{Var}[\chi] \;=\; \frac{\mathrm{Var}[X] + \mathrm{Cov}(X, X')}{2}.$$

So in term of complexity we will prefer χ to X iff

$$\begin{aligned}&\kappa_\chi \mathrm{Var}[\chi] \;\leq\; \kappa \mathrm{Var}[X]\\ \Leftrightarrow\quad & 2\kappa\frac{\mathrm{Var}[X] + \mathrm{Cov}(X, X')}{2} \;\leq\; \kappa \mathrm{Var}[X]\\ \Leftrightarrow\quad & \mathrm{Cov}(X, X') \;\leq\; 0.\end{aligned}$$

To use this remark in practice, one usually relies on the following result.

Proposition 3.2 (Co-monotony)**.** *Let $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and let $\varphi, \phi : \mathbb{R} \to \mathbb{R}$ be two monotone (hence Borel) functions with the same monotony. Assume that $\varphi(Z), \phi(Z) \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$\mathrm{Cov}(\varphi(Z), \psi(Z)) \quad \geq \quad 0.$$

If, mutatis mutandis, φ and ϕ have opposite monotony, then

$$\mathrm{Cov}(\varphi(Z), \psi(Z)) \quad \leq \quad 0.$$

3.2 Regression based control variate

3.2.1 Optimal mean square control variate

We come back to the original situation of two square integrable random variables X and X', having the same expectation $\mathbb{E}[X] = \mathbb{E}[X'] = m$, with nonzero variances and we assume that X and X' are not identical which is equivalent to $\mathrm{Var}[X - X'] > 0$.

This time we simply (and temporarily) set $\xi := X - X'$. The idea is simply to parametrize the impact of the control variate ξ by a factor λ *i.e.* we set for every $\lambda \in \mathbb{R}$,

$$X^\lambda \quad = \quad X - \lambda\xi.$$

Then the strictly convex parabolic function Φ defined by

$$\begin{aligned} \Phi(\lambda) \quad &:= \quad \mathrm{Var}[X^\lambda] \\ &= \quad \lambda^2 \mathrm{Var}[\xi] - 2\lambda \mathrm{Cov}(X, \xi) + \mathrm{Var}[X] \end{aligned}$$

reaches its minimum $\lambda_{\min}$ with

$$\lambda_{\min} \quad := \quad \frac{\mathrm{Cov}(X, \xi)}{\mathrm{Var}[\xi]}.$$

Consequently

$$\begin{aligned} \sigma^2_{\min} \quad &:= \quad \mathrm{Var}[X^{\lambda_{\min}}] \\ &= \quad \mathrm{Var}[X] - \frac{\mathrm{Cov}(X, \xi)^2}{\mathrm{Var}[\xi]} \\ &= \quad \mathrm{Var}[X'] - \frac{\mathrm{Cov}(X', \xi)^2}{\mathrm{Var}[\xi]}\,, \end{aligned}$$

so that $\sigma^2_{\min} \leq \min(\mathrm{Var}[X], \mathrm{Var}[X'])$ and $\sigma^2_{\min} = \mathrm{Var}[X]$ iff $\mathrm{Cov}(X, \xi) = 0$.

3.2.2 Implementation of the variance reduction

Let $(X_k, X'_k)_{k\geq 1}$ be an i.i.d. sequence of random vectors with the same distribution as (X, X') and let $\lambda \in \mathbb{R}$. Set for every $k \geq 1$

$$\begin{aligned} \xi_k \quad &:= \quad X_k - X'_k\,, \\ X^\lambda_k \quad &:= \quad X_k - \lambda\xi_k. \end{aligned}$$

Now we set for every size $M \geq 1$ of simulation

$$V_M \quad := \quad \frac{1}{M}\sum_{k=1}^M \xi_k^2\,, \tag{3.1}$$

$$C_M \quad := \quad \frac{1}{M}\sum_{k=1}^M X_k\xi_k^2\,, \tag{3.2}$$

$$\lambda_M \quad := \quad \frac{C_M}{V_M}. \tag{3.3}$$

The batch method The strong law of large numbers implies that both $V_M \xrightarrow[M\to\infty]{a.s.} \mathrm{Var}[\xi]$ and $C_M \xrightarrow[M\to\infty]{a.s.} \mathrm{Cov}[X, \xi]$ so that $\lambda_M \xrightarrow[M\to\infty]{a.s.} \lambda_{\min}$. This suggests to introduce the batch estimator of m, defined for every size $M \geq 1$ of the simulation by

$$\begin{aligned} \bar{X}^{\lambda_M}_M \quad &:= \quad \frac{1}{M}\sum_{k=1}^M X^{\lambda_M}_k \\ &= \quad \bar{X}_M - \lambda_M \xi_M. \end{aligned}$$

Proposition 3.3. *The batch estimator a.s. converges to m (consistency) i.e.*

$$\bar{X}^{\lambda_M}_M \quad \xrightarrow[M\to\infty]{a.s.} \mathbb{E}[X] = m\,,$$

and satisfies a CLT (asymptotic normality) with an optimal asymptotic variance $\sigma^2_{\min}$ i.e.

$$\sqrt{M}\left(\bar{X}^{\lambda_M}_M - m\right) \quad \xrightarrow[M\to\infty]{\mathcal{L}} \quad \mathcal{N}(0, \sigma^2_{\min}).$$

Remark. However, note that the batch estimator is a biased estimator of m since $\mathbb{E}[\lambda_M \bar{\xi}_M] \neq 0$.

The adaptive unbiased approach Another approach is to design an adaptive estimator of m by considering at each step k the (predictable) estimator λ_{k-1} of $\lambda_{\min}$.

Theorem 3.4. *Assume $X, X' \in L^{2+\delta}(\mathbb{P})$ for some $\delta > 0$. Let $(X_k, X'_k)_{k\geq 1}$ be an i.i.d. sequence with the same distribution as (X, X'). We set for every $k \geq 1$*

$$\begin{aligned} \tilde{X}_k \quad &= \quad X_k - \tilde{\lambda}_{k-1}\xi_k \\ &= \quad (1 - \tilde{\lambda}_{k-1})X_k + \tilde{\lambda}_{k-1}X'_k\,, \end{aligned}$$

where $\tilde{\lambda}_k = (-k) \vee (\lambda_k \wedge k)$. And λ_k is defined by (3.3). Then the adaptive estimator of m defined by

$$\bar{X}^{\tilde{\lambda}}_M \quad := \quad \frac{1}{M}\sum_{k=1}^M \tilde{X}_k$$

is unbiased, convergent and asymptotically normal with minimal variance.

3.3 Application to option pricing

The variance reduction by regression introduced in the former section still relies on the fact that $\kappa_X \approx \kappa_{X-\lambda\xi}$ or, equivalently that the additional complexity induced by the simulation of ξ given that of X is negligible. This condition may look demanding but we will see that in the framework of derivative pricing this requirement is always fulfilled as soon as the payoff of interest satisfies a so-called parity equation.

For a vanilla option the call-put parity is

$$C_0 - P_0 \;=\; S_0 - e^{-rT}K\,,$$

so that $C_0 = \mathbb{E}[X] = \mathbb{E}[X']$ with

$$\begin{aligned} X &:= e^{-rT}(S_T - K)^+\;; \\ X' &:= e^{-rT}(K - S_T)^+ + S_0 - e^{-rT}K. \end{aligned}$$

Note that the simulation of X involves that of S_T so that the additional cost of the simulation of ξ is definitely negligible.

3.4 Pre-conditioning

The principle of the pre-conditioning method – also known as the Blackwell-Rao method – is based on the very definition of conditional expectation.

For every subfield $\mathcal{B} \subset \mathcal{A}$

$$\mathbb{E}[X] \;=\; \mathbb{E}\left[\mathbb{E}[X|\mathcal{B}]\right]$$

and

$$\begin{aligned} \mathrm{Var}[\mathbb{E}[X|\mathcal{B}]] &= \mathbb{E}\left[\mathbb{E}[X|\mathcal{B}]^2\right] - \mathbb{E}[X]^2 \\ &\le \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \;=\; \mathrm{Var}[X]. \end{aligned}$$

The archetypal situation is the following: assume $X = g(Z_1, Z_2)$, where $Z_1 \perp\!\!\!\perp Z_2$. We have $\mathbb{E}[X] = \mathbb{E}[G(Z_2)]$ where

$$\begin{aligned} G(z_2) &= \mathbb{E}[g(Z_1, Z_2)|Z_2 = z_2] \\ &= \mathbb{E}[g(Z_1, z_2)]. \end{aligned}$$

At this stage, the pre-conditionning method can be implemented as soon as the following conditions are satisfied:

- a closed form is available for the function G and
- (the distribution of) Z_2 can be simulated with the same complexity as (the distribution of) X.

Example 3.3 (Exchange spread options)**.** Let $X_T^i = x_i \exp\left(\left(r - \frac{\sigma_i^2}{2}\right)T + \sigma_i W_T^i\right)$, x_i, $\sigma_i > 0$, $i = 1, 2$, be two "Black-Scholes" assets at time T related to two Brownian motions W_T^i , $i = 1, 2$, with correlation $\rho \in [-1, 1]$. One considers an exchange spread options with strike K *i.e.* related to the payoff

$$h_T \;=\; (X_T^1 - X_T^2 - K)^+.$$

Then on can write $(W_T^1, W_T^2) = \sqrt{T}(\sqrt{1-\rho^2}Z_1 + \rho Z_2, Z_2)$, where $(Z_1, Z_2) \sim \mathcal{N}(0, I_2)$. Then

$$\mathbb{E}[e^{-rT}h_T] \;=\; e^{-rT}\mathbb{E}[\mathbb{E}[h_T|Z_2]]$$

with a smaller variance than the original payoff.

3.5 Stratified sampling

Let us define the r.v. $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$.

Definition 3.2 (Strata)**.** We define a strata as $(A_i)_{i\in I}$ a finite partition of the space E, *i.e.* $A_i \cap A_j = \varnothing$ for all $i \neq j$, $\bigcup_{i\in I} A_i = \Omega$.

For this method we have to know:

(i) $p_i = \mathbb{P}\{X \in A_i\}$, $i \in I$.

(ii) how to simulate $X|X \in A_i$, *i.e.*, $X|X \in A_i \sim \varphi_i(U)$, $i \in I$, $\varphi_i = [0,1]^{r_i} \to E$, $U \sim \mathcal{U}([0,1]^{r_i})$, $r_i \in \mathbb{N} \cup \{\infty\}$.

Then with $F : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\begin{aligned} \mathbb{E}[F(X)] &= \sum_{i\in I} \mathbb{E}[F(X)\mathbb{1}_{\{X\in A_i\}}] \\ &= \sum_{i\in I} \mathbb{E}[F(X)|X \in A_i]\mathbb{P}\{X \in A_i\} \\ &= \sum_{i\in I} \mathbb{E}\left[F\left(\varphi_i(U)\right)\right] p_i. \end{aligned}$$

Then the estimator is

$$\begin{aligned} \hat{I}_M &= \sum_{i\in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F(\varphi_i(U_k^i)) \\ &= \frac{1}{M} \sum_{i\in I} \frac{p_i}{q_i} \sum_{k=1}^{M_i} F(\varphi_i(U_k^i))\,, \end{aligned}$$

with $q_i = \frac{M_i}{M}$, $M = \sum_i M_i$ the total budget. Then we have

$$\mathrm{Var}[\hat{I}_M] \;=\; \frac{1}{M} \sum_{i\in I} \frac{p_i^2}{q_i} \underbrace{\mathrm{Var}[F(\varphi_i(U))]}_{=:\ \sigma_{F,i}^2}.$$

We then have the optimizing problem

$$\min_{\substack{\sum q_i = 1 \\ q_i > 0}} \sum_i \frac{p_i^2}{q_i}\sigma_{F,i}^2.$$

3.5.1 Sub-optimal choice

We will take $q_i = p_i$,

$$\begin{aligned} \mathrm{Var}[\hat{I}_M] &= \frac{1}{M}\sum_i p_i \sigma_{F,i}^2 \\ &= \frac{1}{M}\sum_i \mathbb{P}\{X \in A_i\} \\ &\qquad \times \mathbb{E}[(F(X) - \mathbb{E}[F(X)|X \in A_i])^2 | X \in A_i] \\ &= \frac{1}{M}\mathbb{E}\left[F(X) - \sum_i \mathbb{E}[F(X)|X \in A_i]\mathbb{1}_{\{X\in A_i\}}\right]^2 \\ &= \frac{1}{M}\left\|X - \mathbb{E}[F(X)|\sigma(\{X \in A_i\})]\right\|_2^2 \\ &\le \frac{1}{M}\left\|X - \mathbb{E}[F(X)\right\|_2^2 \;=\; \frac{\mathrm{Var}[F(X)]}{M}. \end{aligned}$$

So this choice always reduces the variance of the estimator since we assumed that the stratification is not trivial. It corresponds in the opinion poll world to the so-called quota method.

3.5.2 Optimal choice

We can write

$$\begin{aligned}\sum_i \frac{p_i^2}{q_i}\sigma_{F,i}^2 &= \left(\sum_i\left(\frac{p_i}{\sqrt{q_i}}\sigma_{F,i}\right)^2\right)\left(\sum_i\sqrt{q_i}^2\right)\\ &\geq \left(\sum_i \frac{p_i}{\sqrt{q_i}}\sigma_{F,i}\sqrt{q_i}\right)^2,\end{aligned}$$

with the equality iff $\exists\lambda$ s.t. $\frac{p_i}{\sqrt{q_i}}\sigma_{F,i} = \lambda\sqrt{q_i}$, *i.e.*

$$\begin{aligned}q_i^* &= \frac{p_i\sigma_{F,i}}{\lambda}\\ &= \frac{p_i\sigma_{F,i}}{\sum_j p_j\sigma_{F,j}}.\end{aligned}$$

The problem is that the $\sigma_{F,i}$ are in the formula, hence q^* is hard to explicit, and the complexity is actually greater.

3.6 Importance sampling

We still have $X : (\Omega,\mathcal{A},\mathbb{P}) \to (E,\mathcal{E})$ and its density $\mathbb{P}_X = f\mu$ where μ is a reference measure. Let $h \in L^1(\mathbb{P}_X)$,

$$\mathbb{E}[h(X)] = \int_E h(x)f(x)\mu(\mathrm{d}x).$$

Now if we have Y, $\mathbb{P}_Y = g\mu$, $g > 0$, that can be simulated,

$$\begin{aligned}\mathbb{E}[h(X)] &= \int_E h(x)\frac{f(x)}{g(x)}\underbrace{g(x)\mu(\mathrm{d}x)}_{\mathbb{P}_Y(\mathrm{d}x)}\\ &= \mathbb{E}\left[h\frac{f}{g}(Y)\right].\end{aligned}$$

We will then take Y to calculate $\mathbb{E}[h(X)] = \mathbb{E}[h\frac{f}{g}(Y)]$ if

$$\begin{aligned}& \mathrm{Var}[h\frac{f}{g}(Y)] < \mathrm{Var}[h(X)]\\ \Leftrightarrow\quad & \mathbb{E}\left[\left(h\frac{f}{g}\right)^2(Y)\right] < \mathbb{E}[h^2(X)].\end{aligned}$$

But we have

$$\begin{aligned}\mathbb{E}\left[\left(h\frac{f}{g}\right)^2(Y)\right] &= \int h^2(y)\frac{f^2(y)}{g^2(y)}g(y)\mu(\mathrm{d}y)\\ &= \mathbb{E}\left[h^2(X)\frac{f(X)}{g(X)}\right]\end{aligned}$$

hence we will adopt Y iff $\mathbb{E}\left[h^2(X)\frac{f}{g}(X)\right] < \mathbb{E}[h^2(X)]$.

In practice we introduce $(Y_\theta)_{\theta\in\Theta}$ with $\mathbb{E}[h(X)] = \mathbb{E}\left[h(Y_\theta)\frac{f(Y_\theta)}{g(Y_\theta)}\right]$. The problem becomes a parametric optimization problem, typically solving the minimization problem

$$\min_{\theta\in\Theta}\left\{\mathbb{E}\left[\left(h\frac{f}{g_\theta}(Y_\theta)\right)^2\right] = \mathbb{E}\left[h^2(X)\frac{f}{g_\theta}(X)\right]\right\}$$

4 The Quasi-Monte Carlo method

4.1 Motivation and definitions

Definition 4.1 (Weak convergence)**.** Let $(\mu_n)_n$ be a sequence of probability measures and μ a probability measure. The sequence $(\mu_n)_n$ weakly converges to μ (denoted $\mu_n \Longrightarrow \mu$) if for every bounded function f,

$$\int f\,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int f\,\mathrm{d}\mu.$$

Theorem 4.1 (Glivenko–Cantelli)**.** *If $(U_n)_n$ is an i.i.d. sequence of uniformly distributed r.v. on $[0,1]^d$, then*

$$\frac{1}{n}\sum_{k=1}^n \delta_{U_k(\omega)} \Longrightarrow \lambda_{d|[0,1]^d} = \mathcal{U}([0,1]^d).$$

That is to say, $\forall f \in \mathcal{C}([0,1]^d,\mathbb{R})$,

$$\frac{1}{n}\sum_{k=1}^n f(U_k) \xrightarrow[n\to\infty]{} \int_{[0,1]^d} f(x)\lambda_d(\mathrm{d}x).$$

Definition 4.2. We note the box $[\![x,y]\!]$ defined for every $x = (x^1,\ldots,x^d)$, $y = (y^1,\ldots,y^d) \in [0,1]^d$, $x \leq y$ (2), by

$$[\![x,y]\!] := \{\xi \in [0,1]^d,\ x \leq \xi \leq y\}.$$

Theorem 4.2 (Portemanteau)**.** *Let $(\xi_n)_n$ be a $[0,1]^d$-valued sequence. The following assertions are equivalent.*

(i) $(\xi_n)_n$ is uniformly distributed on $[0,1]^d$.

(ii) For every $x \in [0,1]^d$,

$$\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{[\![0,x]\!]}(\xi_k) \xrightarrow[n\to\infty]{} \lambda_d([\![0,x]\!]) = \prod_{i=1}^d x^i.$$

(iii) ("Discrepancy at the origin")

$$D_n^*(\xi) := \sup_{x\in[0,1]^d}\left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{[\![0,x]\!]}(\xi_k) - \prod_{i=1}^d x^i\right| \xrightarrow[n\to\infty]{} 0.$$

(iv) ("Extreme discrepancy")

$$D_n^\infty(\xi) := \sup_{x,y\in[0,1]^d}\left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{[\![x,y]\!]}(\xi_k) - \prod_{i=1}^d (y^i - x^i)\right| \xrightarrow[n\to\infty]{} 0.$$

[2] This order relation if defined by: $x \leq y$ if $x^i \leq y^i$, $1 \leq i \leq d$.

(v) *(Weyl's criterion) For every integer $p \in \mathbb{N}^d \setminus \{0\}$*

$$\frac{1}{n}\sum_{k=1}^{n} e^{2i\pi(p|\xi_k)} \quad \xrightarrow[n\to\infty]{} \quad 0.$$

(vi) *(Bounded Riemann integrable function) For every bounded continuous function $f : [0,1]^d \to \mathbb{R}$,*

$$\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) \quad \xrightarrow[n\to\infty]{} \quad \int_{[0,1]^d} f(x)\lambda_d(\mathrm{d}x).$$

Proof. □

4.2 Application to numerical integration: functions with finite variations

Definition 4.3. A function $f : [0,1]^d \to \mathbb{R}$ has finite variation in the measure sense if there exists a signed measure ν s.t. $\nu(\{0\}) = 0$ and $\forall x \in [0,1]^d$,

$$f(x) \quad = \quad f(\mathbf{1}) \; + \; \nu([\![0, \mathbf{1} - x]\!]).$$

The variation $V(f)$ is defined by

$$V(f) \quad := \quad |\nu|([0,1]^d)\,,$$

where $|\nu|$ is the variation measure of ν.

Theorem 4.3 (Koksma–Hlawka)**.** *Let $\xi = (\xi_1, \dots, \xi_n)$ be a n-tuple of $[0,1]^d$-valued vectors and let $f : [0,1]^d \to \mathbb{R}$ be a function with finite variation in the measure sense. Then,*

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - \int_{[0,1]^d} f(x)\lambda(\mathrm{d}x)\right| \quad \leq \quad V(f)D_n^*(\xi).$$

Proof. □

Remark. It's actually very rare to find a funciton that has variation finite in the measure sense. For example $f(x^1, x^2) = (x^1 + x^2) \wedge 1$ has variation finite, but it's not the case for $f(x^1, x^2, x^3) = (x^1 + x^2 + x^3) \wedge 1$.

4.3 Sequences with low discrepancy

4.3.1 Back to MC on $[0,1]^d$

Let $(U_n)_n$ be an i.i.d. sequence of r.v. uniformly distributed on $[0,1]^d$. So it's natural to evaluate its discrepancy $D_n^*\left((U_k)_{k\geq 1}\right)$ and to wonder at which rate it goes to zero.

Theorem 4.4 (Chung)**.** *(i) CLT for the star discrepancy. We have*

$$\sqrt{n}D_n^*\left((U_k)_{k\geq 1}\right) \quad \xrightarrow{\mathcal{L}} \quad \sup_{x\in[0,1]^d} |Z_x^d|\,,$$

where $(Z_x^d)_{x\in[0,1]^d}$ denotes the centred Gaussian multi-index with covariance given by $x = (x^1, \dots, x^d)$, $y = (y^1, \dots, y^d) \in [0,1]^d$, $\mathrm{Cov}(Z_x^d, Z_y^d) = \prod_{i=1}^d x^i \wedge y^i - \left(\prod_{i=1}^d x^i\right)\left(\prod_{i=1}^d y^i\right)$. And

$$\mathbb{E}\left[\sqrt{n}D_n^*\left((U_k)_{k\geq 1}\right)\right] \quad \to \quad \mathbb{E}\left[\sup_{x\in[0,1]^d} |Z_x^d|\right].$$

(ii) LIL for the star discrepancy. We have

$$\limsup_n \sqrt{\frac{2n}{\ln\ln n}}D_n^*\left((U_k)_{k\geq 1}\right) \quad = \quad 1.$$

At this stage we have a first definition for a sequence with low discrepancy, ξ s.t.

$$D_n^*(\xi) \quad = \quad o\left(\sqrt{\frac{\ln\ln n}{n}}\right), \quad \text{when } n \to \infty.$$

Which means that its implementation with a function with finite variation will speed up the estimation with respect to the MC simulation.

4.3.2 Röth's lower bound

There exists a universal constant $c_d \in]0, \infty[$ s.t. for any $[0,1]^d$-valued n-tuple $(\xi_1, \dots, \xi_n)$,

$$D_n^*(\xi) \quad \geq \quad c_d \frac{\ln^{\frac{d-1}{2}} n}{n}.$$

Definition 4.4. A $[0,1]^d$-valued sequence $(\xi_n)_{n\geq 1}$ is a sequence with low discrepancy if

$$D_n^*(\xi) \quad = \quad O\left(\frac{\ln^d n}{n}\right), \qquad \text{as } n \to \infty.$$

4.3.3 Examples of sequences

Van des Corput Let $p_1, \dots, p_d$ be the first d prime numbers. The d-dimensional VdC sequence is defined, $\forall n \geq 1$, by:

$$\xi_n \quad = \quad \left(\Phi_{p_1}(n), \dots, \Phi_{p_d}(n)\right),$$

where the radical inverse functions Φ_p are defined by

$$\Phi_p(n) \quad = \quad \sum_{k=0}^{r} \frac{a_k}{p^{k+1}}\,,$$

with $n = a_0 + a_1 p + \dots + a_r p^r$, $a_i \in [\![0, p-1]\!]$, $a_r \neq 0$, denotes the p-adic expansion of n. Then for every $n \geq 1$,

$$\begin{aligned} D_n^*(\xi) \quad &\leq \quad \frac{1}{n}\left(1 + \prod_{i=1}^{d}\left((p_i - 1)\left\lfloor\frac{\ln(p_i n)}{\ln p_i}\right\rfloor\right)\right) \\ &= \quad O\left(\frac{\ln^d n}{n}\right), \qquad \text{as } n \to \infty. \end{aligned}$$

Kakutani

Faure Let p be the smallest prime integer s.t. $p \geq d$. The d-dimensional Faure sequence is defined for every $n \geq 1$, by

$$\xi_n = \big(\Phi_p(n-1), C_p(\Phi_p(n-1)), \ldots, C_p^{d-1}(\Phi_p(n-1))\big)$$

where Φ_p still denotes the radical inverse, and for every p-adic rational number u with (regular) p-adic expansion $u = \sum_{k\geq 0} u_k p^{-(k+1)} \in [0,1]$

$$C_p(u) = \sum_{k\geq 0} \left(\sum_{j \geq k} \binom{j}{k} u_j \mod p \right) p^{-(k+1)}.$$

These sequences' discrepancy at the origin satisfies

$$D_n^*(\xi) \leq \frac{1}{n} \left(\frac{1}{d!} \left(\frac{p-1}{2\ln p} \right)^d \ln^d p + O\left(\ln^{d-1} n\right) \right).$$

Sobol'

Niederraiter

5 Discretization scheme(s) of a Brownian diffusion

One considers a d-dimensional Brownian diffusion process $(X_t)_{t\in[0,T]}$ solution to the following SDE

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \tag{5.1}$$

where $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathcal{M}_{d,q}(\mathbb{R})$ are continuous functions, $(W_t)_{t\in[0,T]}$ denotes a q-dimensional standard Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $X_0 : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ is a random vector, independent of W.

We assume that b and σ are Lipschitz continuous continuous in x uniformly with respect to t, *i.e.* $\forall t \in [0,T]$, $\forall x, y \in \mathbb{R}^d$,

$$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \leq K|x-y|. \tag{5.2}$$

Theorem 5.1. *Under the above assumptions on b, σ, X_0 and W, the above SDE has a unique $(\mathcal{F}_t)$-adapted solution $X = (X_t)_{t\in[0,T]}$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, starting from X_0 at time 0, in the following sense: $\mathbb{P}$-a.s. $\forall t \in [0,T]$,*

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

This solution has $\mathbb{P}$-a.s. continuous paths.

5.1 Euler–Maruyama schemes

5.1.1 Discrete time

The discrete time Euler scheme is defined by

$$\begin{aligned} \bar{X}^n_{t^n_{k+1}} &= \bar{X}^n_{t^n_k} + \frac{T}{n} b(t^n_k, \bar{X}^n_{t^n_k}) \\ &\quad + \sigma(t^n_k, \bar{X}^n_{t^n_k}) \left(W_{t^n_{k+1}} - W_{t^n_k} \right) \qquad (5.3) \\ &= \bar{X}^n_{t^n_k} + \frac{T}{n} b(t^n_k, \bar{X}^n_{t^n_k}) + \sigma(t^n_k, \bar{X}^n_{t^n_k}) \sqrt{\frac{T}{n}} U^n_k, \end{aligned}$$

$\bar{X}_0 = X_0$, $k \in [\![0, n-1]\!]$. Where $t^n_k := \frac{kT}{n}$ and $(U_k)_k$ denotes a sequence of i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{I}_q)$-distributed random vectors.

5.1.2 Stepwise constant

For convenience, we denote from now on

$$\underline{t} := t^n_k \qquad \text{if } t \in [t^n_k, t^n_{k+1}[.$$

The stepwise constant Euler scheme, denoted $(\tilde{X}_t)_{t\in[0,T]}$ for convenience, is defined by

$$\tilde{X}^n_t = \bar{X}^n_{\underline{t}}, \qquad t \in [0,T].$$

5.1.3 Genuine (continuous)

At this stage it is natural to extend the definition (5.3) of the Euler scheme at every instant $t \in [0,T]$ by interpolating the drift with respect to time and the diffusion coefficient with respect to the Brownian motion, namely

$$\begin{aligned} \bar{X}_t &= \bar{X}^n_{\underline{t}} + (t - \underline{t}) b(\underline{t}, \bar{X}^n_{\underline{t}}) \\ &\quad + \sigma(\underline{t}, \bar{X}^n_{\underline{t}})(W_t - W_{\underline{t}}) \qquad (5.4) \\ &= \bar{X}^n_{\underline{t}} + \int_{\underline{t}}^t b(\underline{s}, \bar{X}^n_{\underline{s}})\,\mathrm{d}s + \int_{\underline{t}}^t \sigma(\underline{s}, \bar{X}^n_{\underline{s}})\,\mathrm{d}W_s. \end{aligned}$$

And then by concatenation,

$$\bar{X}_t = X_0 + \int_0^t b(\underline{s}, \bar{X}^n_{\underline{s}})\,\mathrm{d}s + \int_0^t \sigma(\underline{s}, \bar{X}^n_{\underline{s}})\,\mathrm{d}W_s.$$

5.2 Strong error rate and polynomial moments

5.2.1 Main results and comments

We consider the SDE and its Euler–Maruyama scheme(s) as defined by (5.1) and (5.3), (5.4).

Polynomial moment control

Proposition 5.2. *Assume that the coefficients b and σ of the SDE (5.1) are Borel functions that simply satisfy the following linear growth assumption: $\forall t \in [0,T]$, $\forall x \in \mathbb{R}^d$,*

$$|b(t,x)| + \|\sigma(t,x)\| \leq C(1 + |x|) \tag{5.5}$$

for some real constant $C>0$ and a "horizon" $T>0$. Then, for every $p\in[1,\infty[$, there exists a universal positive real constant κ_p such that every strong solution $(X_t)_{t\in[0,T]}$ (if any) satisfies

$$\left\|\sup_{t\in[0,T]}|X_t|\right\|_p \;\leq\; 2e^{\kappa_p CT}\left(1+\|X_0\|_p\right)$$

and, for every $n\geq 1$, the Euler scheme with step $\frac{T}{n}$ satisfies

$$\left\|\sup_{t\in[0,T]}|\bar{X}^n_t|\right\|_p \;\leq\; 2e^{\kappa_p CT}\left(1+\|X_0\|_p\right).$$

Uniform convergence rate in $L^p(\mathbb{P})$ First we introduce the following condition (H^β_T) which strengthens Assumption (5.2) by adding a time regularity assumption of the Hölder type:

$$(H^\beta_T) \;\equiv\; \begin{cases} \exists\beta\in[0,1],\ \exists C_{b,\sigma,T}>0 \text{ s.t.} \\ \forall s,t,\in[0,T],\ \forall x\in\mathbb{R}^d, \\ \\ |b(t,x)-b(s,x)|+\|\sigma(t,x)+\sigma(s,x)\| \\ \qquad \leq\; C_{b,\sigma,T}(1+|x|)|t-s|^\beta. \end{cases}$$

Theorem 5.3 (Strong Rate for the Euler scheme)**.** *(a) Continuous Euler scheme. Suppose the coefficients b and σ of the SDE (5.1) satisfy (5.2) the above regularity condition (H^β_T) for a real constant $C_{b,\sigma,T}>0$ and an exponent $\beta\in]0,1]$. Then the continuous Euler scheme $(\bar{X}^n_t)_{t\in[0,T]}$ converges toward $(X_t)_{t\in[0,T]}$ in every $L^p(\mathbb{P})$, $p>0$, such that $X_0\in L^p$, at a $O(n^{-(\frac{1}{2}\wedge\beta)})$-rate. To be precise, there exists a universal constant $\kappa_p>0$ only depending on p such that, for every $n\geq T$,*

$$\left\|\sup_{t\in[0,T]}|X_t-\bar{X}^n_t|\right\|_p \;\leq\; K(p,b,\sigma,T)(1+\|X_0\|_p)\left(\frac{T}{n}\right)^{\beta\wedge\frac{1}{2}}$$

(b) Stepwise constant Euler scheme. As soon as b and σ satisfy the linear growth assumption (5.5) with a real constant $L:=L_{b,\sigma,T}>0$, then, for every $p\in[1,\infty[$ and every $n\geq T$,

$$\left\|\sup_{t\in[0,T]}|\bar{X}^n_t-\bar{X}^n_{\underline{t}}|\right\|_p \leq \tilde{\kappa}_p e^{\tilde{\kappa}_p LT}(1+\|X_0\|_p)\sqrt{\frac{T(1+\ln n)}{n}}.$$

where $\tilde{\kappa}_p>0$ is a positive real constant only depending on p (and increasing in p).

5.3 Milstein scheme

5.3.1 The 1-dimensional setting

We are still in the setting

$$\mathrm{d}X_t \;=\; b(X_t)\,\mathrm{d}t \;+\; \sigma(X_t)\,\mathrm{d}W_t.$$

We have $d=1$, $q=1$, and we want to work on the approximation of $\int_{t_k}^{t_{k+1}}\sigma(X_s)\,\mathrm{d}W_s$. So Itô's formula on $\sigma\in\mathcal{C}^2$ gives us

$$\begin{aligned}\sigma(X_s) \;=\; & \underbrace{\sigma(X_{t_k})}_{=:\,(1)} + \underbrace{\int_{t_k}^s \sigma'(X_u)\sigma(X_u)\,\mathrm{d}W_s}_{=:\,(2)} \\ & + \underbrace{\int_{t_k}^s\left(b(X_u)\sigma'(X_u)+\frac{1}{2}\sigma''(X_u)\sigma^2(X_u)\right)\mathrm{d}u}_{=:\,(3)}\end{aligned}$$

We have for the first and third term,

$$\begin{aligned}\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}(1)\,\mathrm{d}W_s\right)^2\right] &= \underbrace{\mathbb{E}[\sigma^2(X_{t_k})]}_{<\infty}\underbrace{\mathbb{E}[(W_{t_{k+1}}-W_{t_k})^2]}_{=\frac{T}{n}} \\ \mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}(3)\,\mathrm{d}W_s\right)^2\right] &= \mathbb{E}\left[\int_{t_k}^{t_{k+1}}(3)^2\,\mathrm{d}s\right] \\ &= O\left(\left(\frac{T}{n}\right)^2\right).\end{aligned}$$

So we will see that both expressions are negligible compared to the second term:

$$\begin{aligned}(2) \;=\; & \int_{t_k}^s\left(\sigma\sigma'(W_u)-\sigma\sigma'(W_{t_k})\right)\mathrm{d}W_u \\ & + \sigma\sigma'(W_{t_k})(W_s-W_{t_k}),\end{aligned}$$

so that

$$\begin{aligned}\int_{t_k}^{t_{k+1}}(2)\,\mathrm{d}W_s \;=\; & \sigma\sigma'(X_{t_k})\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})\,\mathrm{d}W_s \\ & + \underbrace{\int_{t_k}^{t_{k+1}}\int_{t_k}^s(\underbrace{\sigma\sigma'(X_u)+\sigma\sigma'(X_{t_k})}_{=\,o(1)})\mathrm{d}W_u\,\mathrm{d}W_s}_{\text{negligible}} \\ \approx\; & \sigma\sigma'(X_{t_k})\frac{1}{2}\left((W_{t_{k+1}}-W_{t_k})^2-(t_{k+1}-t_k)\right).\end{aligned}$$

So we deduce the Milstein scheme:

$$\begin{aligned}\tilde{X}^{mil}_0 &= X_0\,; \\ \tilde{X}^{mil}_{t_{k+1}} &= \tilde{X}^{mil}_{t_k} + \left(b(\tilde{X}^{mil}_{t_k})-\frac{1}{2}\sigma\sigma'(\tilde{X}^{mil}_{t_k})\right)\frac{T}{n} \\ &\quad + \sigma(\tilde{X}^{mil}_{t_k})\sqrt{\frac{T}{n}}U_{k+1} + \frac{1}{2}\sigma\sigma'(\tilde{X}^{mil}_{t_k})\frac{T}{n}U^2_{k+1},\end{aligned}$$

where $U_k=\sqrt{\frac{n}{T}}(W_{t_n}-W_{t_{n-1}})\sim N(0,1)$. It is then easy to write the continuous scheme.

Theorem 5.4 (Strong rate for the Milstein scheme)**.** *(a) Assume that b and σ are $\mathcal{C}^1$ on $\mathbb{R}$ with bounded and Lipschitz derivatives. Then, for every $p\in[1,\infty[$, there exists a real constant $C_{b,\sigma,T,p}>0$ such that*

$$\left\|\max_{k\in[\![0,n]\!]}\left|X_{t_k}-\tilde{X}^{mil}_{t_k}\right|\right\|_p \;\leq\; C_{b,\sigma,T,p}\frac{T}{n}(1+\|X_0\|_p).$$

(b) As concerns the stepwise constant Milstein scheme, one has

$$\left\|\sup_{t\in[0,T]}\left|X_t - \tilde{X}^{mil}_{\underline{t}}\right|\right\|_p \le C_{b,\sigma,T,p}(1+\|X_0\|_p)\sqrt{\frac{T}{n}(1+\ln n)}.$$

5.3.2 Higher dimensional Milstein scheme

In higher dimension when the underlying diffusion process $(X_t)_{t\in[0,T]}$ is d-dimensional or the driving Brownian motion W is q-dimensional which means that the drift is a function $b:\mathbb{R}^d\to\mathbb{R}^d$ and the diffusion coefficient $\sigma=[\sigma_{ij}]:\mathbb{R}^d\to\mathcal{M}_{d,q}(\mathbb{R})$, the same reasoning as in the 1-dimensional setting leads to the following (discrete time) scheme.

$$\begin{aligned}
\tilde{X}^{mil}_0 &= X_0\ ;\\
\tilde{X}^{mil}_{t_{k+1}} &= \tilde{X}^{mil}_{t_k} + \frac{T}{n}b(\tilde{X}^{mil}_{t_k}) + \Delta W_{t_{k+1}}\sigma^*(\tilde{X}^{mil}_{t_k})\\
&\quad + \sum_{1\le i,j\le q}\partial\sigma_{\cdot i}\sigma_{\cdot j}(\tilde{X}^{mil}_{t_k})\int_{t_k}^{t_{k+1}}(W^i_s - W^i_{t_k})\,\mathrm{d}W^j_s
\end{aligned}$$

where $\Delta W_{t_{k+1}} := W_{t_{k+1}} - W_{t_k}$, $\sigma_{\cdot k}$ denotes the k-th column of σ, and $\forall x=(x_1,\dots,x_d)\in\mathbb{R}^d$,

$$\partial\sigma_{\cdot i}\sigma_{\cdot j}(x) := \sum_{l=1}^d \frac{\partial\sigma_{\cdot i}}{\partial x_l}(x)\sigma_{l,j}(x).$$

Here the hard part to simulate is the joint q^2-dimensional distribution

$$\left(W^1_t,\dots,W^d_t,\int_0^t W^i_s\,\mathrm{d}W^j_s,\ 1\le i,j\le q,\ i\ne j\right).$$

Proposition 5.5. *If the rectangular terms commute, i.e. if $\forall i\ne j$,*

$$\partial\sigma_{\cdot i}\sigma_{\cdot j} = \partial\sigma_{\cdot j}\sigma_{\cdot i}\,,$$

then the Milstein scheme reduces to

$$\begin{aligned}
\tilde{X}^{mil}_{t_{k+1}} &= \tilde{X}^{mil}_{t_k} + \frac{T}{n}\left(b(\tilde{X}^{mil}_{t_k}) - \frac{1}{2}\sum_{i=1}^q\partial\sigma_{\cdot i}\sigma_{\cdot i}(\tilde{X}^{mil}_{t_k})\right)\\
&\quad + \sigma(\tilde{X}^{mil}_{t_k})\Delta W_{t_{k+1}}\\
&\quad + \frac{1}{2}\sum_{1\le i,j\le q}\partial\sigma_{\cdot i}\sigma_{\cdot j}(\tilde{X}^{mil}_{t_k})\Delta W^i_{t_{k+1}}\Delta W^j_{t_{k+1}}.
\end{aligned}$$

5.4 Weak error for the Euler scheme

We recall that our problem is to calculate $\mathbb{E}[\varphi(X_T)]$, so actually we are looking at the weak error:

$$\mathbb{E}[\varphi(X_T)] - \mathbb{E}[\varphi(\bar{X}^n_T)]$$

where $\bar{X}^n_T$ is the final value of a discrete scheme.

We can show *e.g.* if φ is 1-Lipschitz,

$$\begin{aligned}
\left|\mathbb{E}[\varphi(X_T)] - \mathbb{E}[\varphi(\bar{X}^n_T)]\right| &\le \mathbb{E}\left[\left|\varphi(X_T)-\varphi(\bar{X}^n_T)\right|\right]\\
&\le \mathbb{E}\left[\left|X_T - \bar{X}^n_T\right|\right]
\end{aligned}$$

and $\mathbb{E}\left[\left|X_T-\bar{X}^n_T\right|\right]$ is in $O(\frac{1}{\sqrt{n}})$ if Euler and $O(\frac{1}{n})$ if Milstein. But this last majoration is too strong, we thinner majoration.

Theorem 5.6. *(a) Assume b and σ are 5 times continuously differentiable on $\mathbb{R}^d$ with bounded existing partial derivatives (this implies that b and σ are Lipschitz). Assume $f:\mathbb{R}^d\to\mathbb{R}$ is 5 times differentiable with polynomial growth as well as its existing partial derivatives. Then, for every $x\in\mathbb{R}^d$,*

$$\mathbb{E}\left[f(\bar{X}^n_T)\right] = \mathbb{E}\left[f(X_T)\right] + O\left(\frac{1}{n}\right) \qquad \text{as } n\to\infty.$$

(b) If $b,\sigma\in\mathcal{C}^{R+5}_b$ and $f\in\mathcal{C}^{R+5}_{pol}$, $R\ge 1$, then

$$\mathbb{E}\left[f(\bar{X}^n_T)\right] = \mathbb{E}\left[f(X_T)\right] + \sum_{r=1}^R c_r n^{-r} + O\left(n^{-(R+1)}\right)$$

as $n\to\infty$.

(c) If $b,\sigma\in\mathcal{C}^\infty_b$, f bounded and σ uniformly elliptic (i.e. $\sigma\sigma^*(x)\ge\xi\mathbf{I}_d$), then we have the result of (b) for any $R\ge 1$.

Remark. The part (b) has a great interest, indeed let

$$\begin{aligned}
\mathbb{E}\left[f(\bar{X}^n_T)\right] &= \mathbb{E}\left[f(X_T)\right] + \frac{c_1}{n} + \frac{c_2}{n^2} + O\left(\frac{1}{n^3}\right)\ ;\\
\mathbb{E}\left[f(\bar{X}^{2n}_T)\right] &= \mathbb{E}\left[f(X_T)\right] + \frac{c_1}{2n} + \frac{c_2}{4n^2} + O\left(\frac{1}{n^3}\right).
\end{aligned}$$

Hence

$$\begin{aligned}
2\mathbb{E}\left[f(\bar{X}^{2n}_T)\right] - \mathbb{E}\left[f(\bar{X}^n_T)\right] &= \mathbb{E}\left[f(X_T)\right] - \frac{c_2}{2n^2}\\
&\quad + O\left(\frac{1}{n^3}\right).
\end{aligned}$$

5.5 Richardson–Romberg extrapolation with consistent increments

To take advantage of the expansion, we will perform a Richardson-Romberg extrapolation. We introduce two Euler schemes, $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ with interval $\frac{T}{n}$ and $\frac{T}{2n}$, associated to $W^{(1)}$ and $W^{(2)}$. Hence

$$\begin{aligned}
\mathbb{E}\left[f(\bar{X}^{(1)}_T)\right] &= \mathbb{E}\left[f(X_T)\right] + \frac{c_1}{n} + \frac{c_2}{n^2} + O\left(\frac{1}{n^3}\right)\ ;\\
\mathbb{E}\left[f(\bar{X}^{(2)}_T)\right] &= \mathbb{E}\left[f(X_T)\right] + \frac{c_1}{2n} + \frac{c_2}{4n^2} + O\left(\frac{1}{(2n)^3}\right)
\end{aligned}$$

so by combining them,

$$\begin{aligned}
2\mathbb{E}\left[f(\bar{X}^{(2)}_T)\right] - \mathbb{E}\left[f(\bar{X}^{(1)}_T)\right] &= \mathbb{E}\left[f(X_T)\right] - \frac{c_1}{2n}\\
&\quad + O\left(\frac{1}{n^3}\right).
\end{aligned}$$

So the natural estimator is

$$\hat{I}_M := \frac{1}{M}\sum_{k=1}^M\left(2f\left((\bar{X}^{(2)}_T)_k\right) - f\left((\bar{X}^{(1)}_T)_k\right)\right),$$

and the quadratic error associated is

$$\begin{aligned}\left\|\mathbb{E}\left[f(X_T)\right]-\hat{I}_M\right\|_2^2 &\triangleq \mathrm{Var}\left[\hat{I}_M\right]+\mathbb{E}\left[\mathbb{E}\left[f(X_T)\right]-\hat{I}_M\right]^2 \\ &= \frac{\mathrm{Var}\left[2f(\bar{X}_T^{(2)})-f(\bar{X}_T^{(1)})\right]}{M} \\ &\quad + \left(\frac{c_2}{n^2}\right)^2 + O(n^{-5}).\end{aligned}$$

A lazy choice would lead in $W^{(1)} \perp\!\!\!\perp W^{(2)}$ and then

$$\mathrm{Var}\left[2f(\bar{X}_T^{(2)})-f(\bar{X}_T^{(1)})\right] = 5\mathrm{Var}\left[f(X_T)\right],$$

which is a bad idea because we have a much greater L^2 error than without the extrapolation.

The best choice is to take $W^{(1)} = W^{(2)}$, hence the previous variance is equal to $\mathrm{Var}\left[f(X_T)\right]$.

5.6 Link between PDE and simulation: Feynmann-Kac's formula

We have, with $d = q = 1$ the following SDE

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t.$$

Let $u \in \mathcal{C}^{1,2}([0,T],\mathbb{R})$, and with $\frac{\partial u}{\partial x}$ satisfies polynomial growth, solution of

$$\frac{\partial u}{\partial t} + Lu = 0, \qquad u(T,\cdot) = f \tag{5.6}$$

where the operator L is defined by

$$Lg(t,x) = b(t,x)\partial_x g(t,x) + \frac{1}{2}\sigma^2(t,x)\partial_x^2 g(t,x).$$

With Itô's Lemma we have

$$\begin{aligned}f(X_T^x) &= u(T, X_T^x) \\ &= u(0,x) + \int_0^T \underbrace{\left(\frac{\partial u}{\partial t}+Lu\right)}_{=\,0}(s,X_s^x)\,\mathrm{d}s \\ &\quad + \int_0^T \sigma(s,X_s^x)\frac{\partial u}{\partial x}(s,X_s^x)\,\mathrm{d}W_s.\end{aligned}$$

And as the third member is a martingale we have the direct link with the expectation:

$$\mathbb{E}[F(X_T^x)] = u(0,x).$$

6 Back to sensitivity computation

Let $Z : (\Omega, \mathcal{Z}, \mathbb{P}) \to (E, \mathcal{E})$ be a r.v., and with I a nonempty open space of $\mathbb{R}$, define $F : I \times E \to \mathbb{R}$. Then set

$$f(x) = \mathbb{E}\left[F(x,Z)\right].$$

Assume that the function f is regular, at least at some points. Our aim is to devise a method to compute by simulation $f'(x)$ at such points. If,

- the functional $F(x,z)$ is differentiable at x,
- if a domination or uniform integrability property holds,
- the partial derivative $\frac{\partial F}{\partial x}(x,z)$ can be computed at a reasonable cost,
- and Z is a simulatable random vector (still at a reasonable cost)

then it is natural to compute $f'(x)$ using a Monte Carlo simulation based on the representation formula

$$f'(x) = \mathbb{E}\left[\frac{\partial F}{\partial x}(x,Z)\right].$$

6.1 Finite difference method

6.1.1 The constant step approach

We will distinguish two cases: in the first one – called "regular setting" – the function $x \mapsto F(x, Z(\omega))$ is "not far" from being pathwise differentiable whereas in the second one – called "singular setting" – f remains smooth but F becomes "singular".

The regular setting

Proposition 6.1. *Let $x \in \mathbb{R}$. Assume that F satisfies the following local mean quadratic Lipschitz continuous assumption ("at x"), $\exists \varepsilon_0 > 0$, $\forall x' \in (x-\varepsilon_0, x+\varepsilon_0)$,*

$$|F(x,Z) - F(x',Z)|^2 \le C_{F,Z}|x-x'|.$$

Assume the function f is twice differentiable with a Lipschitz continuous second derivative on $]x-\varepsilon_0, x+\varepsilon_0[$. Let $(Z_k)_{k\ge 1}$ be a sequence of i.i.d. random vectors with the same distribution as Z, then for every $\varepsilon \in]0,\varepsilon_0[$,

$$\begin{aligned}&\left\|f'(x) - \frac{1}{M}\sum_{k=1}^M \frac{F(x+\varepsilon,Z_k)+F(x-\varepsilon,Z_k)}{2\varepsilon}\right\|_2 \\ &\le [f'']_{\mathrm{lip}}\frac{\varepsilon^2}{2} + \frac{C_{F,Z}}{\sqrt{M}}.\end{aligned}$$

Remark. In the above sum $[f'']_{\mathrm{lip}}\frac{\varepsilon^2}{2}$ represents the bias and $\frac{C_{F,Z}}{\sqrt{M}}$ is the statistical error.

From a practical point of view this means that, in order to reduce the error by a factor 2, we need to reduce ε and increase M as follows:

$$\begin{aligned}\varepsilon &\rightsquigarrow \frac{\varepsilon}{\sqrt{2}} \\ M &\rightsquigarrow 4M.\end{aligned}$$

The singular setting

Proposition 6.2. *Let $x \in \mathbb{R}$. Assume that F satisfies in a neighbourhood $]x\varepsilon_0, x+\varepsilon_0[$, $\varepsilon_0 > 0$, of x the following local mean quadratic θ-Hölder assumption ($\theta \in]0,1]$) assumption ("at x") i.e. there exists a positive real constant $C_{Hol,F,Z}$, $\forall x', x'' \in]x\varepsilon_0, x+\varepsilon_0[$,*

$$\|F(x',Z) - F(x'',Z)\| \leq C_{Hol,F,Z}|x'-x''|^{\theta}.$$

Assume the function f is twice differentiable with a Lipschitz continuous second derivative on $]x-\varepsilon_0, x+\varepsilon_0[$. Let $(Z_k)_{k\geq 1}$ be a sequence of i.i.d. random vectors with the same distribution as Z, then for every $\varepsilon \in]0,\varepsilon_0[$,

$$\left\| f'(x) - \frac{1}{M}\sum_{k=1}^{M} \frac{F(x+\varepsilon, Z_k) + F(x-\varepsilon, Z_k)}{2\varepsilon} \right\|_2 \leq [f'']_{\mathrm{lip}} \frac{\varepsilon^2}{2} + \frac{C_{Hol,F,Z}}{(2\varepsilon)^{1-\theta}\sqrt{M}}.$$

6.1.2 A recursive approach: finite difference with decreasing step

Let $(\varepsilon_k)_{k\geq 1}$ be a sequence of positive real numbers decreasing to 0. With the notations and the assumptions of the former section, consider the estimator

$$\widehat{f'(x)}_M := \frac{1}{M}\sum_{k=1}^{M} \frac{F(x+\varepsilon_k, Z_k) + F(x-\varepsilon_k, Z_k)}{2\varepsilon_k}$$

Remark. It can be computed recursively.

We can easily show that

$$\left\| f'(x) - \widehat{f'(x)}_M \right\|_2 \leq \frac{1}{\sqrt{M}} \sqrt{\frac{[f'']^2_{\mathrm{lip}}}{4M}\left(\sum_{k=1}^{M}\varepsilon_k^2\right)^2 + C^2_{F,Z}}.$$

In order to prove a $\frac{1}{\sqrt{M}}$ rate (like in a standard Monte Carlo M simulation) we need the sequence $(\varepsilon_m)_{m\geq 1}$ and the size M to satisfy

$$\left(\sum_{k=1}^{M}\varepsilon_k^2\right)^2 = O(M),$$

this leads to choose ε_k of the form $\varepsilon_k = O\left(k^{-\frac{1}{4}}\right)$ as $k \to \infty$.

6.2 Pathwise differentiation method

Theorem 6.3 (Kusuoka)**.**

Example 6.1. If $d = q = 1$, the above SDE reads

$$\begin{aligned} \mathrm{d}Y_t(x) &= Y_t(x)\,(b'_x(t,X_t)\,\mathrm{d}t + \sigma'_x(t,X^x_t)\,\mathrm{d}W_t) ; \\ Y_0(x) &= 1. \end{aligned}$$

and elementary computations show that

$$Y_t(x) = \exp\left(\int_0^t \left(b'_x(t,X^x_s) - \frac{1}{2}\sigma'_x(s,X^x_s)^2\right)\mathrm{d}s + \int_0^t \sigma'_x(s,X^x_s)\,\mathrm{d}W_s\right).$$

6.3 Sensitivity computation for non smooth payoffs