

Montecarlo methods

The goal of Montecarlo methods is to compute $\mathbb{E}(X)$, where X is a random variable. In dimension 1, deterministic methods are more efficient but in higher dimensions ($d \geq 4$), Montecarlo methods are more competitive.

1. | Foundations

As always, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $Y \in L^1$.

Principle

Definition 1. The main idea will be to approximate $\mathbb{E}(Y)$ by $\frac{1}{n} \sum_{i=1}^n Y_i := \bar{Y}_n$, where Y_i are i.i.d. random variables with same law as Y . The variable $\bar{Y}_n$ is called the *Montecarlo estimator* of $\mathbb{E}(Y)$.

Lemma 2. The Montecarlo estimator is consistent, i.e. $\bar{Y}_n \xrightarrow{\text{a.s.}} \mathbb{E}(Y)$, and unbiased, i.e. $\mathbb{E}(\bar{Y}_n) = \mathbb{E}(Y)$.

Proof. Use the ?? ?? $\square$

Lemma 3. Assume $Y \in L^2$ and let $\bar{Y}_n$ be the Montecarlo estimator of $\mathbb{E}(Y)$. Then:

$$\|\bar{Y}_n - \mathbb{E}(Y)\|_2 = \sqrt{\frac{\text{Var}(Y)}{n}}$$

Proof.

$$\begin{aligned} \|\bar{Y}_n - \mathbb{E}(Y)\|_2 &= \sqrt{\mathbb{E}((\bar{Y}_n - \mathbb{E}(Y))^2)} = \\ &= \sqrt{\text{Var}(\bar{Y}_n)} = \sqrt{\frac{\text{Var}(Y)}{n}} \end{aligned}$$

$\square$

Lemma 4. Let $Y \in L^2$ and $\bar{Y}_n$ be the Montecarlo estimator of $\mathbb{E}(Y)$. Then:

$$\sqrt{n}(\bar{Y}_n - \mathbb{E}(Y)) \xrightarrow{d} N(0, \text{Var}(Y))$$

Proof. Use ?? ?? $\square$

Remark. In practice, we do not know $\text{Var}(Y)$, so we use an estimator of it, such as $\bar{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$, which is a consistent unbiased estimator of $\text{Var}(Y)$. Thus:

$$\frac{\sqrt{n}}{\bar{\sigma}_n}(\bar{Y}_n - \mathbb{E}(Y)) \xrightarrow{d} N(0, 1)$$

by ?? ??.

Lemma 5. Let $Y \in L^2$ and $\bar{Y}_n$ be the Montecarlo estimator of $\mathbb{E}(Y)$. Then, a confidence interval for $\mathbb{E}(Y)$ of level $1 - \alpha$ is:

$$\text{CI}_\alpha := \left(\bar{Y}_n - z_{1-\alpha/2} \frac{\bar{\sigma}_n}{\sqrt{n}}, \bar{Y}_n + z_{1-\alpha/2} \frac{\bar{\sigma}_n}{\sqrt{n}} \right)$$

where $z_{\alpha/2}$ is the quantile of order $\alpha/2$ of the standard normal distribution.

Random number generator

In this chapter we will assume that we already now how to simulate sequences of i.i.d. random variables with uniform distribution on $[0, 1]$.

Remark. In summary, the computer generates a sequence $(x_i)_{0 \leq i \leq m}$, with m as large as possible, in the following way: $x_{i+1} = f(x_i)$ and then sets $u_i = \frac{x_i}{m}$. The value x_0 is called the *seed* of the sequence and f is chosen with periodicity as high as possible. In the early days of computers, $f(x) = ax + b \pmod{m}$, which had periodicity $m \sim 2^{31} - 1$. Nowadays, *Mersenne Twister algorithm* is used, which has periodicity $m \sim 2^{19937} - 1$.

Simulation of random variables

Lemma 6. Let $U, (U_i)_{0 \leq i \leq d} \sim U([0, 1])$. Then:

- If $a, b \in \mathbb{R}$ with $a < b$, then $a + (b - a)U \sim U([a, b])$.
- If $p \in (0, 1)$, then $\mathbf{1}_{U \leq p} \sim \text{Ber}(p)$.
- If $p \in (0, 1)$, then $\sum_{i=1}^d \mathbf{1}_{U_i \leq p} \sim \text{B}(d, p)$.
- If $(x_n), (p_n) \in \mathbb{R}$ be such that $\sum_{n \geq 0} p_n = 1$, then $\sum_{n \geq 0} x_n \mathbf{1}_{\sum_{k=0}^{n-1} p_k \leq U \leq \sum_{k=0}^n p_k} \sim U((x_n))$.
- If $\prod_{i=1}^d (a_i, b_i) \in \mathbb{R}^d$ with $a_i < b_i$, then $(a_i + (b_i - a_i)U_i)_{1 \leq i \leq d} \sim U\left(\prod_{i=1}^d (a_i, b_i)\right)$.

Proposition 7. Let X be a random variable with cdf F and $U \sim U([0, 1])$. Then,

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$$

satisfies $F^{-1}(U) \sim X$.

Proposition 8. Let $U \sim U([0, 1])$, X be a random variable with cdf F and $a, b \in \mathbb{R}$ with $a < b$ be such that $\mathbb{P}(a < X \leq b) > 0$. Then:

$$F^{-1}(F(a) + (F(b) - F(a))U) \sim \mathcal{L}(X \mid a < X \leq b)$$

Proposition 9 (Acceptance-rejection method). Let $(X_i)_{i \geq 1}$ be i.i.d. $\mathbb{R}^d$ -valued random variables, $D \in \mathcal{B}(\mathbb{R}^d)$ be such that $\mathbb{P}(X_1 \in D) > 0$ and set:

$$\nu := \inf\{i \geq 1 : X_i \in D\}$$

Then, $X_\nu \sim \mathcal{L}(X_1 \mid X_1 \in D)$.

Remark. The principle of the acceptance-rejection method is to simulate conditional distributions by rejecting samples that do not satisfy a prescribed condition.

Proposition 10. Let f be a pdf of some random variable, $(X_i)_{i \geq 1}$ be i.i.d. with pdf g and $(U_i)_{i \geq 1}$ be i.i.d. $U([0, 1])$ independent of $(X_i)_{i \geq 1}$. Assume that $\exists c \geq 1$ such that $f(x) \stackrel{\text{a.e.}}{\leq} cg(x)$ and set:

$$\nu := \inf\{i \geq 1 : cg(X_i)U_i \leq f(X_i)\}$$

Then, X_ν admits f as pdf.

Proposition 11. Let f be a pdf of some random variable and $a_1, a_2 \in \mathbb{R}$ with $a_2 > 0$ be such that

$$D := \{(u, v) \in \mathbb{R}_{>0} \times \mathbb{R} : 0 < u^2 < f\left(a_1 + a_2 \frac{v}{u}\right)\}$$

is bounded. If $(U, V) \sim U(D)$, then $a_1 + a_2 \frac{V}{U}$ admits f as pdf.

Gaussian distribution

Proposition 12 (Box-Muller method). Let U, V be i.i.d. $U([0, 1])$ and set:

$$X := \sqrt{-2\log(U)} \cos(2\pi V) \quad Y := \sqrt{-2\log(U)} \sin(2\pi V)$$

Then, X and Y are i.i.d. $N(0, 1)$.

Proof. Let $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be bounded and measurable. Then:

$$\begin{aligned} \mathbb{E}(\varphi(X, Y)) &= \\ &= \int_{(0,1)^2} \varphi\left(\sqrt{-2\log u} \cos(2\pi v), \sqrt{-2\log u} \sin(2\pi v)\right) du dv = \\ &= \int_{\mathbb{R}^2} \varphi(x, y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy \end{aligned}$$

by the change of variable formula. Thus, X and Y are i.i.d. $N(0, 1)$. $\square$

Proposition 13 (Polar method). Let U, V be i.i.d. $U(\mathbb{D})$, where $\mathbb{D} \subset \mathbb{R}^2$ is the open unit disk. Let $R^2 = U^2 + V^2$ and set:

$$X := U \sqrt{\frac{-2\log(R^2)}{R^2}} \quad Y := V \sqrt{\frac{-2\log(R^2)}{R^2}}$$

Then, X, Y are i.i.d. $N(0, 1)$.

Proposition 14. Let $\mathbf{X} \in N_d(0, \mathbf{I}_d)$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathcal{M}_d(\mathbb{R})$. Then, $\boldsymbol{\mu} + \mathbf{A}\mathbf{X} \sim N_d(\boldsymbol{\mu}, \mathbf{A}\mathbf{A}^T)$.

Remark. To simulate $\mathbf{Y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we proceed as follows:

1. Find $\mathbf{A} \in \mathcal{M}_d(\mathbb{R})$ such that $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ (e.g. by Cholesky decomposition).
2. Simulate $\mathbf{X} \sim N_d(0, \mathbf{I}_d)$.
3. Set $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{X}$.

2. | Variance reduction techniques

Antithetic control

Definition 15. Let $Y = g(X)$ be a random variable with $X \sim N(0, \sigma^2)$. The *antithetic method* consists in using the estimator:

$$\bar{Y}_n^A := \frac{1}{n} \sum_{i=1}^n \frac{g(X_i) + g(-X_i)}{2}$$

where $(X_i)_{1 \leq i \leq n}$ are i.i.d. $N(0, \sigma^2)$.

Lemma 16. The antithetic estimator $\bar{Y}_n^A$ is an unbiased and consistent estimator of $\mathbb{E}(Y)$. Furthermore:

$$\sqrt{n} \left(\bar{Y}_n^A - \mathbb{E}(Y) \right) \xrightarrow{d} N \left(0, \text{Var} \left(\frac{g(X) + g(-X)}{2} \right) \right)$$

Remark. In terms of computational cost, $\bar{Y}_n^A$ is more expensive than $\bar{Y}_n$ but cheaper than $\bar{Y}_{2n}$.

Proposition 17. We have that $\text{Var}(\bar{Y}_n^A) \leq \text{Var}(\bar{Y}_n)$. Moreover, if

$$\text{Cov}(g(X), g(-X)) \leq 0$$

then $\text{Var}(\bar{Y}_n^A) \leq \text{Var}(\bar{Y}_{2n})$.

Proof.

$$\begin{aligned} \text{Var}(\bar{Y}_n^A) &= \frac{1}{2n} (\text{Var}(g(X)) + \text{Cov}(g(X), g(-X))) \\ &\leq \text{Var}(\bar{Y}_n) \end{aligned}$$

And clearly if $\text{Cov}(g(X), g(-X)) \leq 0$, then $\text{Var}(\bar{Y}_n^A) \leq \text{Var}(\bar{Y}_{2n})$. $\square$

Proposition 18. If g is monotone, then:

$$\text{Cov}(g(X), g(-X)) \leq 0$$

Proof. Let X_1, X_2 be i.i.d. copies of X . Then, since g is monotone:

$$(g(X_1) - g(X_2))(g(-X_1) - g(-X_2)) \leq 0$$

Now taking expectations:

$$\begin{aligned} 0 &\geq \mathbb{E}((g(X_1) - g(X_2))(g(-X_1) - g(-X_2))) = \\ &= 2\text{Cov}(g(X), g(-X)) \end{aligned}$$

$\square$

Control variate

Definition 19. The principle of the *control variate* is to find a real-valued random variable X such that $\mathbb{E}(X)$ is known, and a constant $b \in \mathbb{R}$ such that:

$$\text{Var}(Y + b(X - \mathbb{E}(X))) \ll \text{Var}(Y)$$

We define $Y(b) := Y + b(X - \mathbb{E}(X))$. This suggests the following estimator:

$$\bar{Y}_n(b) := \frac{1}{n} \sum_{i=1}^n [Y_i + b(X_i - \mathbb{E}(X))]$$

where $(Y_i)_{1 \leq i \leq n}$ are i.i.d. copies of Y and $(X_i)_{1 \leq i \leq n}$ are i.i.d. copies of X .

Lemma 20. The control variate estimator $\bar{Y}_n(b)$ is unbiased and consistent. Furthermore:

$$\sqrt{n} (\bar{Y}_n(b) - \mathbb{E}(Y)) \xrightarrow{d} N(0, \text{Var}(Y(b)))$$

where $Y(b) := Y + b(X - \mathbb{E}(X))$.

Remark. If $b = 0$, the control variate estimator $\bar{Y}_n(b)$ coincides with the classical estimator $\bar{Y}_n$. Otherwise, the computational cost of $\bar{Y}_n(b)$ is higher than $\bar{Y}_n$, but it does not depend on the choice of $b \neq 0$.

Proposition 21. The minimum of $\text{Var}(Y(b))$ is attained for

$$\hat{b} = -\frac{\text{Cov}(Y, X)}{\text{Var}(X)}$$

and in that case, $\text{Var}(Y(\hat{b})) = \text{Var}(Y)(1 - \rho_{XY}^2)$, where ρ_{XY} is the correlation between X and Y .

Remark. Usually, $\hat{b}$ is unknown, but we can use an estimator of it, such as:

$$\hat{b}_n := -\frac{\sum_{i=1}^n (Y_i - \bar{Y}_n)(X_i - \bar{X}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

but if we know $\mathbb{E}(X)$ and $\text{Var}(X)$ explicitly, we can use them in the formula of $\hat{b}_n$.

Remark. The result of above tell us that we should pick X strongly correlated to Y but simple enough to know explicitly $\mathbb{E}(X)$.

Definition 22. Let $\mathbf{X}$ be a random vector such that $\mathbb{E}(\mathbf{X})$ is known, and $\mathbf{b} \in \mathbb{R}^d$. We define the *multiple control variate estimator* as:

$$\bar{Y}_n(\mathbf{b}) := \frac{1}{n} \sum_{i=1}^n [Y_i + \mathbf{b}^T(\mathbf{X}_i - \mathbb{E}(\mathbf{X}))]$$

where $(Y_i)_{1 \leq i \leq n}$ are i.i.d. copies of Y and $(\mathbf{X}_i)_{1 \leq i \leq n}$ are i.i.d. copies of $\mathbf{X}$. We also define $Y(\mathbf{b}) := Y + \mathbf{b}^T(\mathbf{X} - \mathbb{E}(\mathbf{X}))$.

Lemma 23. The multiple control variate estimator $\bar{Y}_n(\mathbf{b})$ is unbiased and consistent. Furthermore:

$$\sqrt{n}(\bar{Y}_n(\mathbf{b}) - \mathbb{E}(Y)) \xrightarrow{d} N(0, \text{Var}(Y(\mathbf{b})))$$

Proposition 24. The minimum of $\text{Var}(Y(\mathbf{b}))$ is attained for

$$\hat{\mathbf{b}} = -\text{Var}(\mathbf{X})^{-1} \text{Cov}(\mathbf{X}, Y)$$

and in that case, $\text{Var}(Y(\hat{\mathbf{b}})) = \text{Var}(Y)(1 - R^2)$, where:

$$R^2 := \frac{\text{Cov}(\mathbf{X}, Y)^T \text{Var}(\mathbf{X})^{-1} \text{Cov}(\mathbf{X}, Y)}{\text{Var}(Y)}$$

Importance sampling

Definition 25. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that a probability measure $\mathbb{Q}$ is *equivalent* to $\mathbb{P}$ if:

$$\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} = \{A \in \mathcal{F} : \mathbb{Q}(A) = 0\}$$

Lemma 26. Let $\mathbb{Q}$ be a probability measure equivalent to $\mathbb{P}$. Then, there exists a random variable $L > 0$ such that $\forall A \in \mathcal{F}$, we have $\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}(L\mathbf{1}_A)$. Furthermore, for all X bounded:

$$\mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{P}}(LX) \quad \mathbb{E}_{\mathbb{P}}(X) = \mathbb{E}_{\mathbb{Q}}\left(\frac{X}{L}\right)$$

Conversely, if we have a random variable $L > 0$ such that $\mathbb{E}_{\mathbb{P}}(L) = 1$, then

$$\begin{aligned} \mathbb{Q} : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{E}_{\mathbb{P}}(L\mathbf{1}_A) \end{aligned}$$

is a probability measure equivalent to $\mathbb{P}$.

Remark. The principle of the importance sampling method is to change the probability measure in order to give more weight to important outcomes.

Definition 27. Let $(\mathbb{Q}, L)$ be such that it defines a probability measure equivalent to $\mathbb{P}$. The *importance sampling estimator* is defined as:

$$\bar{Y}_n^{\mathbb{Q}} := \frac{1}{n} \sum_{i=1}^n L_i^{-1} Y_i$$

where (L_i, Y_i) are i.i.d. copies of (L, Y) (under $\mathbb{Q}$).

Lemma 28. The importance sampling estimator $\bar{Y}_n^{\mathbb{Q}}$ is unbiased and consistent. Furthermore:

$$\sqrt{n}(\bar{Y}_n^{\mathbb{Q}} - \mathbb{E}_{\mathbb{P}}(Y)) \xrightarrow{d} N(0, \text{Var}_{\mathbb{Q}}(L^{-1}Y))$$

Remark. In terms of computational cost, the importance sampling estimator $\bar{Y}_n^{\mathbb{Q}}$ is more expensive than the classical estimator $\bar{Y}_n$. In terms of precision, the best estimator is the one with the smallest variance so we have to compare $\text{Var}(Y)$ and $\text{Var}_{\mathbb{Q}}(L^{-1}Y)$. It holds:

$$\text{Var}_{\mathbb{Q}}(L^{-1}Y) = \mathbb{E}_{\mathbb{P}}(L^{-1}Y^2) - \mathbb{E}_{\mathbb{P}}(Y)^2$$

This quantity can be larger or smaller than $\text{Var}(Y)$ depending on the choice of L and thus the success of importance sampling relies on the choice of an effective change of probability measure.

3. | Simulation of diffusion processes

The aim of this section is to develop methods to simulate solutions to SDEs of the form:

$$\{d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t) dt + \boldsymbol{\sigma}(\mathbf{X}_t) d\mathbf{B}_t \mid \mathbf{X}_0 = \mathbf{x}_0\} \quad (1)$$

where $\mathbf{b} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\boldsymbol{\sigma} : \mathbb{R}^d \rightarrow \mathcal{M}_{d \times d}(\mathbb{R})$ are Lipschitz continuous.

Exact simulation

Proposition 29. To simulate a sample $(\mathbf{B}_{t_1}, \dots, \mathbf{B}_{t_m})$ of a d -dimensional Brownian motions, we can use the following algorithm:

1. Generate $(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$ i.i.d. $N_d(0, \mathbf{I}_d)$.
2. Set $\mathbf{B}_{t_0} := 0$ and for all $0 \leq i \leq m-1$, set:

$$\mathbf{B}_{t_{i+1}} = \mathbf{B}_{t_i} + \sqrt{t_{i+1} - t_i} \mathbf{Z}_{i+1}$$

Euler scheme

Definition 30. Consider the SDE of Eq. (1) and let h be a discretization step. The *Euler method* consists in:

$$\begin{aligned} \mathbf{X}_{t+h} &= \mathbf{X}_t + \int_t^{t+h} \mathbf{b}(\mathbf{X}_s) ds + \int_t^{t+h} \boldsymbol{\sigma}(\mathbf{X}_s) d\mathbf{B}_s \\ &\approx \mathbf{X}_t + h\mathbf{b}(\mathbf{X}_t) + \boldsymbol{\sigma}(\mathbf{X}_t)(\mathbf{B}_{t+h} - \mathbf{B}_t) \end{aligned}$$

More generally, if we want to obtain the solution at $(t_1, \dots, t_n)$, we can use the following algorithm. Set $\tilde{\mathbf{X}}_0^m := \mathbf{x}_0$ and for all $0 \leq i \leq m-1$, set:

$$\tilde{\mathbf{X}}_{t_{i+1}}^m = \tilde{\mathbf{X}}_{t_i}^m + (t_{i+1} - t_i)\mathbf{b}(\tilde{\mathbf{X}}_{t_i}^m) + \boldsymbol{\sigma}(\tilde{\mathbf{X}}_{t_i}^m)(\mathbf{B}_{t_{i+1}} - \mathbf{B}_{t_i})$$

Remark. Note that Euler scheme reduces to generating independent increments $\mathbf{B}_{t_{i+1}} - \mathbf{B}_{t_i} \sim \sqrt{t_{i+1} - t_i} N_d(0, \mathbf{I}_d)$.

Remark. Trying to build an implicit Euler scheme for SDEs is much more complicated than for ODEs, as we need to ensure that the process is still adapted.

Definition 31. Let $\mathbf{X}$ be the solution to Eq. (1). We define the *continuous Euler scheme* as:

$$\tilde{\mathbf{X}}_t := \mathbf{x}_0 + \int_0^t \mathbf{b}(\mathbf{X}_{\phi_s}) ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_{\phi_s}) d\mathbf{B}_s$$

where $\phi_s := \max\{t_i : t_i < s\}$.

Lemma 32. Let X be the solution to Eq. (1) with $d = 1$ and $\tilde{X}^m$ be the solution to the Euler scheme. Then, for $p \geq 1$:

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left(\sup_{0 \leq t \leq T} |\tilde{X}_t^m|^p \right) < \infty$$

Lemma 33. Let X be the solution to Eq. (1) with $d = 1$ and $\tilde{X}^m$ be the solution to the Euler scheme. Then, for $p \geq 1$:

$$\max_{i=0, \dots, m-1} \mathbb{E} \left(\sup_{t_i \leq t \leq t_{i+1}} |X_t - \tilde{X}_t^m|^p \right)^{\frac{1}{p}} \leq \frac{C}{\sqrt{m}}$$

Theorem 34 (Strong error of the Euler scheme). Let X be the solution to Eq. (1) with $d = 1$ and $\tilde{X}^m$ be the solution to the Euler scheme. Then, for $p \geq 1$:

$$\mathbb{E} \left(\sup_{0 \leq t \leq T} |X_t - \tilde{X}_t^m|^p \right)^{1/p} \leq \frac{C}{\sqrt{m}}$$

Theorem 35 (Weak error of the Euler scheme). If $\mathbf{b}, \boldsymbol{\sigma} \in \mathcal{C}^\infty(\mathbb{R}^d)$ with bounded derivatives and either one of the following conditions holds:

- $g \in \mathcal{C}_b^\infty(\mathbb{R}^d)$
- g is measurable with polynomial growth and $\boldsymbol{\sigma}$ is uniformly elliptic, i.e. $\exists \lambda > 0$ such that $\forall \mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^d$ we have:

$$\boldsymbol{\xi}^T (\boldsymbol{\sigma} \boldsymbol{\sigma}^T) (\mathbf{x}) \boldsymbol{\xi} \geq \lambda \|\boldsymbol{\xi}\|^2$$

Then:

$$|\mathbb{E}(g(\tilde{X}_T^m)) - \mathbb{E}(g(X_T))| \leq \frac{C}{m}$$

where T is the final time of the simulation.

Remark. We can write the error taking in the Montecarlo estimation using the Euler scheme as:

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n g(\tilde{X}_T^{m,(i)}) - \mathbb{E}(g(X_T)) = \\ &= \frac{1}{n} \sum_{i=1}^n g(\tilde{X}_T^{m,(i)}) - \mathbb{E}(g(\tilde{X}_T^m)) + \mathbb{E}(g(\tilde{X}_T^m)) - \mathbb{E}(g(X_T)) \end{aligned}$$

where $\tilde{X}_T^{m,(i)}$ are i.i.d. copies of $\tilde{X}_T^m$. This error consists in two terms:

- The statistical error:

$$\frac{1}{n} \sum_{i=1}^n g(\tilde{X}_T^{m,(i)}) - \mathbb{E}(g(\tilde{X}_T^m)) \sim \frac{1}{\sqrt{n}}$$

- The discretization error:

$$\mathbb{E}(g(\tilde{X}_T^m)) - \mathbb{E}(g(X_T)) \sim \frac{1}{m}$$

This suggests choosing $m \sim \sqrt{n}$. However, confidence intervals are of little interest in this setting as the amplitude of the bias is similar to the length of the confidence interval. Thus, we have no control on the error a posteriori. We conclude that it is best to be in between the setting $\sqrt{n} \ll m$ to have meaningful confidence intervals and $\sqrt{n} \sim m$ for efficiency, say $n^{\frac{1}{2}+\delta} \sim m$ for $\delta > 0$ small.

Corollary 36 (Romberg Extrapolation). The *Romberg Extrapolation* consists in the following result:

$$\mathbb{E}(2g(\tilde{X}_T^m) - g(\tilde{X}_T^m)) - \mathbb{E}(g(X_T)) = \frac{C}{m^2} + o\left(\frac{1}{m^2}\right)$$

This suggests using the estimator for $\mathbb{E}(g(X_T))$:

$$\frac{1}{n} \sum_{i=1}^n (2g(\tilde{X}_T^{m,(i)}) - g(\tilde{X}_T^{m,(i)}))$$

where $\tilde{X}_T^{m,(i)}$ are i.i.d. copies of $\tilde{X}_T^m$.

4. | Brownian bridge approach

Brownian bridge

Proposition 37. Let $X_t = x + \mu t + \sigma B_t$ with $\mu \in \mathbb{R}$, $\sigma \neq 0$ and $(B_t)_{t \geq 0}$ be a Brownian motion. Then, given $(X_u, X_v) = (y, z)$, the process $(X_t)_{u \leq t \leq v}$ is a continuous Gaussian process, independent of $(\tilde{X}_t)_{t \leq u}$ and $(X_t)_{t \geq v}$, such that for all $u \leq t \leq v$:

$$\mathbb{E}(X_t | X_u = y, X_v = z) = y + \frac{t-u}{v-u}(z-y)$$

$$\text{Cov}(X_t, X_s | X_u = y, X_v = z) = \frac{(t-u)(v-s)}{v-u} \sigma^2$$

Corollary 38. Let $X_t = x + \mu t + \sigma B_t$ with $\mu \in \mathbb{R}$, $\sigma \neq 0$ and $(B_t)_{t \geq 0}$ be a Brownian motion. Then, the distribution of the process $(X_t)_{u \leq t \leq v}$ given $(X_t)_{t \leq u}$ and $(X_t)_{t \geq v}$ is the same as the distribution of

$$(y + \sigma B_{t-u})_{u \leq t \leq v} \quad \text{given } B_{v-u} = \frac{z-y}{\sigma}$$

Exit times

In this section we are interested in approximate the quantity:

$$\mathbb{E}(\mathbf{1}_{\tau > T} g(X_T))$$

where $\tau := \inf\{t \geq 0 : X_t \notin D\}$ and $D \subset \mathbb{R}^d$ is an open connected set.

Remark. Due to ??, we have that $\mathbb{E}(\mathbf{1}_{\tau > T} g(X_T))$ is precisely the solution to the PDE:

$$\begin{cases} \partial_t u + \mathbf{b} \cdot \nabla u + \frac{1}{2} \text{tr}(\sigma \sigma^T \Delta u) = 0 & \text{in } [0, T) \times D \\ u(T, \cdot) = g & \text{in } x \in D \end{cases}$$

Proposition 39. Assume that:

- $\mathbf{b}, \sigma \in \mathcal{C}^3$ with bounded derivatives and σ uniformly elliptic.
- D is a half-space or ∂D is bounded of class $\mathcal{C}^3$.
- g is measurable with polynomial growth and vanishes at a neighborhood of ∂D .

Then:

$$\mathbb{E}(\mathbf{1}_{\tau > T} g(X_T)) - \mathbb{E}(\mathbf{1}_{\tilde{\tau}^m > T} g(\tilde{X}_T^m)) = O\left(\frac{1}{\sqrt{m}}\right)$$

where $\tilde{\tau}^m := \min\{t_i : \tilde{X}_{t_i}^m \notin D\}$ and $\tilde{X}^m$ is the Euler scheme.

Proposition 40. Assume that:

- $\mathbf{b}, \sigma \in \mathcal{C}^5$ with bounded derivatives and σ uniformly elliptic.
- D is a half-space or ∂D is bounded of class $\mathcal{C}^5$.
- g is measurable with polynomial growth and vanishes at a neighborhood of ∂D .

Then:

$$\mathbb{E}(\mathbf{1}_{\tau > T} g(X_T)) - \mathbb{E}(\mathbf{1}_{\tau^m > T} g(\tilde{X}_T^m)) = O\left(\frac{1}{m}\right)$$

Proposition 41. Assume that $\mathcal{D}$ is a half-space with hyperplane orthogonal to $\nu \in \mathbb{R}^d$ passing by $\mathbf{z} \in \mathbb{R}^d$, i.e.:

$$\mathcal{D} = \{\mathbf{y} \in \mathbb{R}^d : \nu^T(\mathbf{y} - \mathbf{z}) > 0\}$$

If $\mathbf{x}_i, \mathbf{x}_{i+1} \in \mathcal{D}$, then:

$$\begin{aligned} \mathbb{P}(\exists t \in [t_i, t_{i+1}], \tilde{X}_t^m \notin \mathcal{D} : \tilde{X}_{t_i}^m = \mathbf{x}_i, \tilde{X}_{t_{i+1}}^m = \mathbf{x}_{i+1}) = \\ = e^{-2 \frac{m}{T} \frac{\nu^T(\mathbf{x}_i - \mathbf{z}) \nu^T(\mathbf{x}_{i+1} - \mathbf{z})}{\|\sigma(\mathbf{x}_i)^T \nu\|^2}} \end{aligned}$$

Lemma 42. Let $(B_t)_{t \geq 0}$ be a Brownian motion. Then, $\forall a \leq 0$ and $b \geq a$ we have:

$$\mathbb{P}\left(\min_{t \in [0, h]} B_t \leq a \mid B_h = b\right) = e^{-\frac{2}{h} a(a-b)}$$

5. | Computation of sensitivities

In this chapter we aim to construct Montecarlo methods in order to compute sensitivities of the price of an option. From the PDE point of view, we aim to compute the derivatives of the solution $(\mathbf{X}_t^{\mathbf{x}})_{t \geq 0}$ of the SDE:

$$\begin{cases} d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t) dt + \sigma(\mathbf{X}_t) dB_t \\ \mathbf{X}_0 = \mathbf{x} \end{cases}$$

Finite difference method

Here we will focus on the case $d = 1$.

Definition 43. Let $u(0, x) := \mathbb{E}(g(X_T^x))$ and $\tilde{u}^n(0, x) := \frac{1}{n} \sum_{i=1}^n g(\tilde{X}_T^{x, (i)})$ where $\tilde{X}_T^{x, (i)}$ are i.i.d. copies of $\tilde{X}_T^x$. We define the *finite difference estimator* as:

$$\begin{aligned} \partial_x u(0, x) &\approx \frac{u(0, x + \varepsilon) - u(0, x - \varepsilon)}{2\varepsilon} \\ &\approx \frac{\tilde{u}^n(0, x + \varepsilon) - \tilde{u}^n(0, x - \varepsilon)}{2\varepsilon} \end{aligned}$$

Proposition 44. If g is smooth enough, then:

$$\text{Var}\left(\frac{\tilde{u}^n(0, x + \varepsilon) - \tilde{u}^n(0, x - \varepsilon)}{2\varepsilon}\right) \approx \frac{1}{n} \text{Var}(g'(X_T^x))$$

Remark. When g is irregular, the optimal choice of ε is not clear, as the bias increases with ε and the variance increases as ε decreases.

Black-Scholes model

Proposition 45. Recall the one-dimensional *Black-Scholes model*:

$$dX_t = rX_t dt + \sigma X_t dB_t$$

Then:

$$\partial_x \mathbb{E}(g(X_T^x)) = \frac{1}{\sigma x T} \mathbb{E}(g(X_T^x) B_T)$$

Pathwise differentiation

Theorem 46. If $\mathbf{b}, \sigma \in \mathcal{C}^2$ with bounded derivatives, then the flow $\mathbf{x} \mapsto \mathbf{X}_t^{\mathbf{x}}$ is $\mathcal{C}^1$ a.s. and the *tangent process* $\mathbf{D}\mathbf{X}^{\mathbf{x}}$ satisfies:

$$\mathbf{D}\mathbf{X}_t^{\mathbf{x}} = \mathbf{I}_d + \int_0^t \mathbf{D}\mathbf{b}(\mathbf{X}_s^{\mathbf{x}}) \mathbf{D}\mathbf{X}_s^{\mathbf{x}} ds + \sum_{j=1}^d \int_0^t \mathbf{D}\sigma_j(\mathbf{X}_s^{\mathbf{x}}) \mathbf{D}\mathbf{X}_s^{\mathbf{x}} dB_s^j$$

where $\mathbf{D}\sigma_j$ is the j -th column of $\mathbf{D}\sigma$.

Remark. In one dimension, we have:

$$\partial_x X_t^x = 1 + \int_0^t b'(X_s^x) \nabla X_s^x ds + \int_0^t \sigma'(X_s^x) \nabla X_s^x dB_s$$

which yields:

$$\partial_x X_t^x = \exp\left(\int_0^t \left[b'(X_s^x) - \frac{1}{2} \sigma'(X_s^x)^2\right] ds + \int_0^t \sigma'(X_s^x) dB_s\right)$$

Proposition 47. If $\mathbf{b}, \sigma \in \mathcal{C}^2$ and $g \in \mathcal{C}^1$ with bounded derivatives, then:

$$\nabla \mathbb{E}(g(\mathbf{X}_T^x)) = \mathbb{E}(\nabla g(\mathbf{X}_T^x) \nabla \mathbf{X}_T^x)$$

Remark. In practice, to find the derivative of $\mathbf{X}_T^x$ we proceed as in the deterministic case, i.e. solving the coupled variational equations.

Malliavin differentiation

Proposition 48. If $\mathbf{b}, \sigma \in \mathcal{C}^2$ with bounded derivatives, σ is uniformly elliptic and g is measurable with polynomial growth, then:

$$\nabla \mathbb{E}(g(\mathbf{X}_T^x)) = \mathbb{E} \left[g(\mathbf{X}_T^x) \frac{1}{T} \left(\int_0^T (\sigma^{-1}(X_t^x) \mathbf{D} \mathbf{X}_t^x)^T \mathbf{d} \mathbf{B}_t \right)^T \right]$$

6. | American options

Definition 49. In a frictionless market, the price of an American option is given by:

$$v(0, x) = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}(e^{-r\tau} g(X_\tau))$$

where r is the risk-free interest rate, $\mathcal{T}_{0,T}$ is the set of stopping times with values in $[0, T]$ and:

$$dX_t = rX_t dt + \sigma(X_t) dB_t \quad X_0 = x$$

In this section we will introduce efficient algorithms to approximate the price of an American option.

Discretization

Definition 50. Fix a time grid $(t_i)_{0 \leq i \leq m}$ with $t_0 = 0$ and $t_m = T$. The discretization method consists in replacing:

1. $\mathcal{T}_{0,T}$ by $\tilde{\mathcal{T}}_{0,T}^m$, the set of stopping times with values in $(t_i)_{0 \leq i \leq m}$.
2. X by $\tilde{X}^m$, the Euler scheme.

Proposition 51. We can compute the price of an American option using the discretization method by the following recursive formula:

$$\begin{cases} \tilde{v}^m(t_m, \tilde{X}_{t_m}^m) = g(\tilde{X}_{t_m}^m) \\ \tilde{v}^m(t_i, \tilde{X}_{t_i}^m) = \\ \quad = \max \left\{ g(\tilde{X}_{t_i}^m), e^{-\frac{rT}{m}} \mathbb{E} \left(\tilde{v}^m(t_{i+1}, \tilde{X}_{t_{i+1}}^m) \mid \tilde{X}_{t_i}^m \right) \right\} \end{cases}$$

Proposition 52. If g is Lipschitz continuous, then:

$$|v(0, x) - \tilde{v}^m(0, x)| \leq \frac{C}{\sqrt{m}}$$

Remark. In the sequel, we assume that $r = 0$ and we write X instead of $\tilde{X}^m$ for the sake of simplicity.

Naive approach

Definition 53. The naive approach consists in proceeding as follows:

1. Generate $(X_{t_1}^j)_{1 \leq j \leq n}$ i.i.d. copies of X_{t_1} given $X_0 = x$ and approximate:

$$\tilde{v}^m(0, x) \approx \max \left\{ g(x), \frac{1}{n} \sum_{j=1}^n \tilde{v}^m(t_1, X_{t_1}^j) \right\}$$

2. For each $1 \leq j \leq n$, generate $(X_{t_2}^{j,k})_{1 \leq k \leq n}$ i.i.d. copies of X_{t_2} given $X_{t_1} = X_{t_1}^j$ and approximate:

$$\tilde{v}^m(t_1, X_{t_1}^j) \approx \max \left\{ g(X_{t_1}^j), \frac{1}{n} \sum_{k=1}^n \tilde{v}^m(t_2, X_{t_2}^{j,k}) \right\}$$

3. For each $(j_1, \dots, j_{m-1}) \in \{1, \dots, n\}^{m-1}$, generate $(X_{t_m}^{j_1, \dots, j_{m-1}, k})_{1 \leq k \leq n}$ i.i.d. copies of X_{t_m} given $X_{t_{m-1}} = X_{t_{m-1}}^{j_1, \dots, j_{m-1}}$ and approximate:

$$\begin{aligned} \tilde{v}^m(t_{m-1}, X_{t_{m-1}}^{j_1, \dots, j_{m-1}}) &\approx \\ &\approx \max \left\{ g(X_{t_{m-1}}^{j_1, \dots, j_{m-1}}), \frac{1}{n} \sum_{k=1}^n \tilde{v}^m(t_m, X_{t_m}^{j_1, \dots, j_{m-1}, k}) \right\} \end{aligned}$$

Remark. This method provides a consistent estimator. However, it requires to generate $\sum_{i=1}^m n^i \sim n^m$ random variables. So the computational cost of the method increases exponentially with the number of exercise dates and becomes prohibitive for applications to the pricing of American options.

Regression methods

Definition 54 (Tsitsiklis-Van Roy method). The Tsitsiklis-Van Roy method consists in approximate the conditional expectation by a projection on a finite dimensional subspace of L^2 . Namely, it holds:

$$\begin{aligned} \mathbb{E}(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) \mid X_{t_i}) &= \\ &= \arg \min_{Y \in L^2(X_{t_i})} \mathbb{E}((\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) - Y)^2) \end{aligned}$$

Here $L^2(X_{t_i})$ is the collection of square-integrable $\sigma(X_{t_i})$ -measurable random variables. Then, we choose a family of basis functions $\varphi = (\varphi_1, \dots, \varphi_\ell)$ and approximate:

$$\mathbb{E}(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) \mid X_{t_i}) \approx \sum_{j=1}^{\ell} \alpha_j^i \varphi_j(X_{t_i})$$

where:

$$\alpha^i = \arg \min_{\alpha \in \mathbb{R}^\ell} \mathbb{E} \left[\left(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) - \sum_{j=1}^{\ell} \alpha_j \varphi_j(X_{t_i}) \right)^2 \right]$$

One can check that

$$\alpha^i = \mathbb{E}(\varphi(X_{t_i}) \varphi(X_{t_i})^T)^{-1} \mathbb{E}(\varphi(X_{t_i}) \tilde{v}^m(t_{i+1}, X_{t_{i+1}}))$$

provided that $\mathbb{E}(\varphi(X_{t_i}) \varphi(X_{t_i})^T)$ is non-degenerate.

Proposition 55. An implementation of the Tsitsiklis-Van Roy method is as follows:

1. Generate $(X_{t_1}^j, \dots, X_{t_m}^j)_{1 \leq j \leq n}$ i.i.d. copies of $(X_{t_1}, \dots, X_{t_m})$.
2. Set $V_m^j = g(X_{t_m}^j)$ for all $1 \leq j \leq n$.
3. Recursively for $i = m-1, \dots, 1$, compute:

$$\tilde{\alpha}^i = \arg \min_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_{j=1}^n \left(V_{i+1}^j - \sum_{k=1}^{\ell} \alpha_k \varphi_k(X_{t_i}^j) \right)^2$$

and set:

$$V_i^j = \max \left\{ g(X_{t_i}^j), \sum_{k=1}^{\ell} \tilde{\alpha}_k^i \varphi_k(X_{t_i}^j) \right\}$$

4. Set:

$$V_0 = \max \left\{ g(x), \frac{1}{n} \sum_{j=1}^n V_1^j \right\}$$

Theorem 56. If

$$\mathbb{E}(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) \mid X_{t_i}) = \sum_{j=1}^{\ell} \alpha_j^i \varphi_j(X_{t_i})$$

then the Tsitsiklis-Van Roy estimator V_0 is consistent, i.e. $V_0 \xrightarrow{\mathbb{P}} v(0, x)$ as $n \rightarrow \infty$.

Definition 57 (Longstaff-Schwartz method). The *Longstaff-Schwartz method* consists in approximate the optimal stopping time instead of the value function itself. Recall that:

$$\tilde{v}^m(0, x) = \sup_{\tau \in \tilde{\mathcal{T}}_{0,T}^m} \mathbb{E}(g(X_{\tau})) = \mathbb{E}(g(X_{\tau^*}))$$

where

$$\tau^* = \inf\{t_i : g(X_{t_i}) \geq \mathbb{E}(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) \mid X_{t_i})\}$$

Proposition 58. The implementation of the Longstaff-Schwartz method is as follows:

1. Generate $(X_{t_1}^j, \dots, X_{t_m}^j)_{1 \leq j \leq n}$ i.i.d. copies of $(X_{t_1}, \dots, X_{t_m})$.
2. Define the stopping rule $\tilde{\tau}_m = t_m$ and apply it to the trajectories simulated just before, i.e. set $V_m^j = g(X_{\tilde{\tau}_m}^j) = g(X_{t_m}^j)$ for all $1 \leq j \leq n$.
3. Recursively for $i = m-1, \dots, 1$, compute:

$$\tilde{\alpha}^i = \arg \min_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_{j=1}^n \left(V_{i+1}^j - \sum_{k=1}^{\ell} \alpha_k \varphi_k(X_{t_i}^j) \right)^2$$

Then, define for any sample path $(X_{t_1}, \dots, X_{t_m})$, the stopping rule:

$$\tilde{\tau}_i = \begin{cases} t_i & \text{if } g(X_{t_i}) \geq \sum_{k=1}^{\ell} \tilde{\alpha}_k^i \varphi_k(X_{t_i}) \\ \tilde{\tau}_{i+1} & \text{otherwise} \end{cases}$$

and apply it to the trajectories simulated just before, i.e. set for $1 \leq j \leq n$:

$$V_i^j = \begin{cases} g(X_{t_i}^j) & \text{if } g(X_{t_i}^j) \geq \sum_{k=1}^{\ell} \tilde{\alpha}_k^i \varphi_k(X_{t_i}^j) \\ V_{i+1}^j & \text{otherwise} \end{cases}$$

4. Define the stopping rule

$$\tilde{\tau}_0 = \begin{cases} 0 & \text{if } g(x) \geq \frac{1}{n} \sum_{j=1}^n V_1^j \\ \tilde{\tau}_1 & \text{otherwise} \end{cases}$$

and apply it to the trajectories simulated just before, i.e. set:

$$V_0 = \frac{1}{n} \sum_{j=1}^n g(X_{\tilde{\tau}_0}^j) = \max \left\{ g(x), \frac{1}{n} \sum_{j=1}^n V_1^j \right\}$$

Theorem 59. If

$$\mathbb{E}(\tilde{v}^m(t_{i+1}, X_{t_{i+1}}) \mid X_{t_i}) = \sum_{j=1}^{\ell} \alpha_j^i \varphi_j(X_{t_i})$$

then the Longstaff-Schwartz estimator V_0 is consistent, i.e. $V_0 \xrightarrow{\mathbb{P}} v(0, x)$ as $n \rightarrow \infty$. Otherwise, the limit corresponds to the value of the option under a sub-optimal stopping rule and so it underestimates the true price.

Remark. However, when n is finite, $\tilde{\tau}_0$ is not a stopping time since it uses information about the future. Thus, we should add a fifth step to the algorithm:

5. Generate $(X_{t_1}^{n+j}, \dots, X_{t_m}^{n+j})_{1 \leq j \leq \tilde{n}}$ i.i.d. copies of $(X_{t_1}, \dots, X_{t_m})$ and apply the stopping rule $\tilde{\tau}_0$ to these new trajectories, i.e. set:

$$\bar{V}_0 = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} g(X_{\tilde{\tau}_0}^{n+j})$$

Lemma 60 (Rogers's lemma). We have:

$$v(0, x) = \inf_{M \in \mathcal{M}_{0,T}} \mathbb{E} \left(\sup_{t \in [0,T]} g(X_t) - M_t \right)$$

where $\mathcal{M}_{0,T}$ is the set of continuous martingales on $[0, T]$.

Remark. Roughly speaking, we can construct a nearly optimal martingale $\tilde{M}$ of the problem above and simulate i.i.d. copies of $(X, \tilde{M})$ to compute the Monte Carlo estimator:

$$\bar{V}_0 = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \sup_{t \in [0,T]} g(X_t^j) - \tilde{M}_t^j$$

This provides a confidence interval for the true price given by:

$$\left[\bar{V}_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\text{Var}(g(X_{\tilde{\tau}_0}))}{\tilde{n}}}, \right.$$

$$\left. \bar{V}_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\text{Var} \left(\sup_{t \in [0,T]} \{g(X_t) - \tilde{M}_t\} \right)}{\tilde{n}}} \right]$$