Introduction to Numerical Methods in Probability for Finance

Gilles Pagès

Introduction to Numerical Methods in Probability for Finance

—

School Mathematical Finance - Hanoi 2007

—

CIMPA

—

Gilles Pagès

LPMA-Université Paris 6

—

E-mail: gpa@ccr.jussieu.fr

—

http: //www.proba.jussieu.fr/pageperso/pages

25.04.2007

Useful links

▷ The website of the Master 2 Probabilité & Finance (Univ. Pierre & Marie Curie, Paris, France):

`www.masterfinance.proba.jussieu.fr`

▷ A website devoted to Quantitative Finance:

`www.maths-fi.com/uk_default.asp`

`Job & Education in Quantitative/IT Finance and Maths`

1 Simulation of random variables

1.1 Pseudo-random numbers

From a mathematical point of view, the definition of a sequence of (uniformly distributed) random numbers (over the unit interval $[0,1]$) should be :

"**Definition.**" *A sequence $x_n, n \geq 1$, of $[0,1]$-valued real numbers is a sequence of random numbers if there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a sequence $U_n, n \geq 1$, of i.i.d. random variables with uniform distribution $\mathcal{U}([0,1])$ and $\omega \in \Omega$ such that $x_n = U_n(\omega)$ fo every $n \geq 1$.*

But this naive and abstract definition is not satisfactory because the "scenario" $\omega \in \Omega$ is may be not a "good" one *i.e.* not a "generic" ... ? Many probabilistic properties (like the law of large numbers to quote the most basic one) are only satisfied $\mathbb{P}$-*a.s.*. Thus, if ω precisely lies in the negligible set that does not satisfy one of them.

Whatever, one usually cannot have access to an i.i.d. sequence of random variables (U_n) with distribution $\mathcal{U}([0,1])$! Any physical device would too slow and not reliable. Some works by logicians like Martin-Löf lead to consider that a sequence (x_n) that can be generated by an algorithm cannot be considered as "random" one. Thus the digits of π are not random in that sense. This is quite embarrassing since an essential requested feature for such sequences is to be generated almost instantly on a computer!

The approach coming for computer and algorithmic sciences is not really more tractable since their definition of a sequence of random numbers is that the complexity of the algorithm to generate the first n terms behaves like $O(n)$. The rapidly growing need of good (pseudo-)random sequences with the explosion of Monte Carlo simulation in many fields of Science and Technology (I mean not only neutronics) after World War II lead to adopt a more pragmatic approach – say – heuristic – based on a statistical tests. The idea is to submit some sequences to statistical tests (uniform distribution, block non correlation, rank tests, etc)

For practical implementation, such sequences are finite , as is the accuracy of computers. One considers some sequences (x_n) of so-called *pseudo-random* numbers displaying as

$$x_n = \frac{y_n}{N}, \quad y_n \in \{0, \ldots, N-1\}.$$

One classical process is to generate the y_n by a congruential induction :

$$y_{n+1} \equiv a y_n + b \mod. N$$

where $\gcd(a, N) = 1$, so that $\overline{a}$ is invertible in the multiplicative group $((\mathbb{Z}/N\mathbb{Z})^*, \times)$ (invertible elements of $\mathbb{Z}/N\mathbb{Z}$ for the product).

If $b = 0$ (the most common case), one speaks of *homogenous generator*.

One will choose N as large as possible given the computation capacity of the computers (with integers) ($N = 2^{31} - 1$ for a 32 *bits* architecture, etc).

Still if $b = 0$, the length of the sequence will be settled by the period $\tau := \min\{t \ / \ a^t \equiv 1 \mod. N\}$ of a in $((\mathbb{Z}/N\mathbb{Z})^*, +)$.

Since $\mathrm{card}(\mathbb{Z}/N\mathbb{Z})^* = \varphi(N)$ where $\varphi(N) := \mathrm{card}\{1 \leq k \leq N-1, \text{ tq } \mathrm{pgcd}(k, N) = 1\}$ is the Euler function, it follows from Lagrange theorem that:

$$\tau = \mathrm{card}(<\overline{a}>) \mid \varphi(N)$$

(where $<\overline{a}>$:= multiplicative sub-group of $(\mathbb{Z}/N\mathbb{Z})^*$ generated by $\overline{a}$). Let us recall

$$\varphi(N) = N \prod_{p|N,\ p\ \mathrm{prime}} \left(1 - \frac{1}{p}\right).$$

The (difficult) study of $((\mathbb{Z}/N\mathbb{Z})^*, \times)$ when N is a primary integer leads to the following theorem:

Theorem 1 *Let $N = p^\alpha$, p prime, $\alpha \in \mathbb{N}^*$.*

(a) *If $\alpha = 1$ (i.e. $N = p$ prime), then $((\mathbb{Z}/N\mathbb{Z})^*, \times)$ (whose cardinality is $p-1$) is a cyclic group. This means that there exists $\overline{a} \in \{1, \ldots, p-1\}$ s.t. $(\mathbb{Z}/p\mathbb{Z})^* = <\overline{a}>$. Hence the maximal period is $\tau = p - 1$.*

(b) *If $p = 2$, $\alpha \geq 3$, $(\mathbb{Z}/N\mathbb{Z})^*$, (whose cardinality is $2^{\alpha-1} = \dfrac{N}{2}$), is not cyclic. The maximal period is then $\tau = 2^{\alpha-2}$ with $a \equiv \pm 3$ mod. 8.*

(c) *If $p \neq 2$, then $(\mathbb{Z}/N\mathbb{Z})^*$, (whose cardinality is $p^{\alpha-1}(p-1)$), is cyclic, hence $\tau = p^{\alpha-1}(p-1)$. It is generated by any element a whose class $\tilde{a}$ in $(\mathbb{Z}/p\mathbb{Z})$ spans the cyclic group $((\mathbb{Z}/p\mathbb{Z})^*, \times)$.*

However, one should be aware that the length of a sequence, if it is a necessary asset of a sequence, provides no guarantee or even clue that a sequence is a good as a sequence of pseudo-random numbers! Thus, the generator of the FORTRAN IMSL library does not fit in the formerly described setting: one sets $N := 2^{31} - 1$ (which is a prime number), $a := 7^5, b := 0$ ($a \not\equiv 0$ mod.8).

Another approach to random number generation is based on shift register and relies upon the theory of finite fields.

1.2 The fundamental principle of simulation

Theorem 2 *Let (E, d) be a Polish space (complete and separable) and $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E,)$ be a random variable with distribution $\mathbb{P}_X$. Then there exists a Borel function $\varphi : ([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]}) \to (E, \mathcal{B}or(E), P_X)$ such that*

$$\mathbb{P}_X = \lambda \circ \varphi^{-1}$$

where $\lambda \circ \varphi^{-1}$denotes the image of the Lebesgue measure $\lambda_{[0,1]}$ by φ.

As a consequence this means that, if U denotes a uniformly distributed random variable on a probability space, then

$$X \stackrel{d}{=} \varphi(U).$$

The interpretation is that any E-valued random variable can be simulated from a uniform distribution. In practice this turns out to be a purely theoretical result which is of no help for practical simulation.

1.3 The distribution function method

Let μ be a probability distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ having a continuous increasing distribution function F. Then F has an inverse function F^{-1} defined $(0,1)$.

Proposition 1 *If $\mathcal{L}(U) = U((0,1))$, then $X := F^{-1}(U) \stackrel{d}{=} \mu$.*

Proof. *: Let $x \in \mathbb{R}$, $\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x)$. Now F^{-1} is increasing, hence $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$. Hence $\mathbb{P}(X \leq x) = F(x)$.* $\diamond$

Remark. • If μ has a probability density f satisfying $\{f = 0\}$ has an empty interior, then $F(x) = \int_{-\infty}^{x} f(u)du$ is continuous, increasing.

• One can replace $\mathbb{R}$ by any interval $[a,b] \subset \overline{\mathbb{R}}$.

• When F is simply non-decreasing (or discontinuous at some points), one defines the canonical right continuous inverse F_r^{-1} by:

$$\forall u \in (0,1), \quad F_r^{-1}(u) = \inf\{s / F(s) \geq u\}.$$

One shows that F_r^{-1} is non-decreasing, right continuous and

$$F_r^{-1}(u) \leq x \Rightarrow F(x) \geq u \text{ et } F(x) > u \Rightarrow F_r^{-1}(u) \leq x.$$

Hence $X \stackrel{d}{=} F_r^{-1}(U)$ since

$$\mathbb{P}(X \leq x) = \mathbb{P}(F_r^{-1}(U) \leq x) \left\{ \begin{array}{l} \leq \mathbb{P}(F(x) \geq U) = F(x) \\ \geq \mathbb{P}(F(x) > U) = F(x) \end{array} \right\} = F(x).$$

If X takes finitely many values in a $\mathbb{R}$, on retrieves the standard simulation method.

• One could have considered the left continuous inverse defined by

$$\forall u \in (0,1), \quad F_l^{-1}(u) = \inf\{s / F(s) > u\}.$$

with the same result.

Examples : • Simulation of an exponential distribution.

Let $X \stackrel{d}{=} \mathcal{E}(\lambda)$, $\lambda > 0$. Then

$$\forall\, x \in (0,\infty), \qquad F_X(x) = \lambda \int_0^x e^{-\lambda\xi} d\xi = 1 - e^{-\lambda x}.$$

Consequently, for every $y \in (0,1)$, $F^{-1}(u) = -\log(1-u)/\lambda$. Now, using that $U \stackrel{d}{=} 1-U$ if $U \stackrel{d}{=} U((0,1))$ yields

$$X = -\log(U)/\lambda \stackrel{d}{=} \mathcal{E}(\lambda).$$

• Simulation of a Cauchy(c), $c > 0$, distribution.

We know that $\mathbb{P}_X(dx) = \dfrac{c}{\pi(x^2+c^2)} dx$.

$$\forall x \in \mathbb{R}, \quad F_X(x) = \int_{-\infty}^{x} \frac{c}{u^2+c^2} \frac{du}{\pi} = \frac{1}{\pi}\left(\text{Arctan}(\frac{x}{c}) + \frac{\pi}{2}\right),$$

hence $F_X^{-1}(x) = c\tan(\pi(u - 1/2))$. It follows that

$$X = c\tan(\pi(U - 1/2)) \stackrel{d}{=} \text{Cauchy}(c).$$

• Simulation of a Pareto(θ), $\theta > 0$, distribution.

We know that $\mathbb{P}_X(dx) = \dfrac{\theta}{x^{1+\theta}} \mathbf{1}_{\{x \geq 1\}} dx$. $F_X(x) = 1 - x^{-\theta}$ so that

$$X = U^{-\frac{1}{\theta}} \stackrel{d}{=} \text{Pareto}(\theta).$$

• Simulation of a purely discrete distribution supported by $E \subset \mathbb{R}$.

Let $E := \{x_1, \ldots, x_N\}$ and $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ an E-valued r.v. with distribution $\mathbb{P}(X = x_k) = p_k$, $1 \le k \le N$. Then, one checks that

$$\forall u \in (0,1), \qquad F_X^{-1}(u) = \sum_{k=1}^{N} x_k \mathbf{1}_{\{p_1+\cdots+p_{k-1} < u \le p_1+\cdots+p_k\}}$$

so that

$$X \stackrel{d}{=} \sum_{k=1}^{N} x_k \mathbf{1}_{\{p_1+\cdots+p_{k-1} < U \le p_1+\cdots+p_k\}}.$$

The yield of the procedure is $\bar{r} = 1$ but when implemented naively its complexity – which corresponds to (at most) N comparisons for every simulation – may be quite high. See [17] for some considerations (in the spirit of quick sort algorithms) which lead to a $O(\log N)$ complexity. Furthermore, this procedure underlines that the p_k are known (as real numbers) which is not always the case even in *a priori simple* situations

• Simulation of a Bernoulli random variable $B(p)$, $p \in (0,1)$.

This is the simplest application of the previous method since

$$X = \mathbf{1}_{\{U \le p\}} \stackrel{d}{=} B(p).$$

The yield of the method is 1.

• Simulation of a Binomial random variable $B(n,p)$, $p \in (0,1)$, $^{\geq}1$.

One relies on the very definition of the binomial distribution as the law of the sum of n independent $B(p)$-distributed random variables *i.e.*

$$X = \sum_{k=1}^{n} \mathbf{1}_{\{U_k \le p\}} \stackrel{d}{=} B(n,p).$$

where $(U_1, \ldots, U_n)$ are i.i.d. $B(p)$-distributed r.v. Note that this procedure has a very bad yield, namely $\frac{1}{n}$ and needs n comparisons like the standard method (without any shortcut). Its asset is that it does not require the computation of the probabilities p_k's.

1.4 The rejection method (Von Neumann)

Let $f, g : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \longrightarrow \mathbb{R}_+$ be two probability densities with respect to a nonnegative measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Assume $g > 0$ μ-*a.s.* and

$$\forall x \in \mathbb{R}^d, \qquad f(x) \le c g(x).$$

Proposition 2 *Let $(U_n, Y_n)_{n\ge1}$ be a sequence of i.i.d. r.v. with distribution $U([0,1]) \otimes \mathbb{P}_Y$ (independent marginals) defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Assume $\mathbb{P}_Y(dy) = g(y)\mu(dy)$*

Let

$$\tau := \min\{k \ge 1 \,|\, c\, U_k g(Y_k) < f(Y_k)\}.$$

Then, τ has a geometric distribution $G^*(p)$ with parameter $p := \mathbb{P}(c\, U_1 g(Y_1) < f(Y_1)))$ and

$$X := Y_\tau \stackrel{d}{=} f(x)\mu(dx).$$

In practice, one needs

– to simulate the distribution of Y,

– to compute the functions f and g,
ms on a computer at a reasonable cost.

The *yield* of the method is obviously $\frac{1}{\tau}$ and its mean yield

$$\mathbb{E}\,\frac{1}{\tau} = -\frac{p}{1-p} logp.$$

Proof. Step 1: Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a bounded Borel test function. By Fubini's Theorem,

$$\begin{aligned}
\mathbb{E}\left(\varphi(Y)\mathbf{1}_{\{cUg(Y)\le f(Y)\}}\right) &= \int_{\mathbb{R}^d}\varphi(y)\int_0^1 \mathbf{1}_{\{cug(y)\le f(y)\}}g(y)\mu(dy)\\
&= \int_{\mathbb{R}^d}\varphi(y)\int_0^1 \mathbf{1}_{\{cug(y)\le f(y)\}\cap\{g(y)>0\}}g(y)\mu(dy)\\
&= \int_{\mathbb{R}^d}\varphi(y)\int_0^1 \mathbf{1}_{\{u\le \frac{f(y)}{cg(y)}\}\cap\{g(y)>0\}}g(y)\mu(dy)\\
&= \int_{\{g(y)>0\}}\varphi(y)\frac{f(y)}{cg(y)}g(y)\mu(dy)\\
&= c\int_{\mathbb{R}^d}\varphi(y)f(y)\mu(dy).
\end{aligned}$$

If $\varphi \equiv 1$, then $c = \mathbb{P}(Uf(Y) \le cg(Y))$, hence, elementary conditioning yields

$$\mathbb{E}\left(\varphi(Y)|\{cUg(Y)\le f(Y)\}\right) = \int_{\mathbb{R}^d}\varphi(y)f(y)\mu(dy)$$

i.e.

$$\mathcal{L}\left(Y|\{cUg(Y)\le f(Y)\}\right) = f(y)\mu(dy).$$

Step 2: Let $B \in \mathcal{B}(\mathbb{R}^d)$. Then

$$\begin{aligned}
\mathbb{P}(X\in B) &= \sum_{n\ge1}\mathbb{E}\left(\mathbf{1}_{\{\tau=n\}}\mathbf{1}_{\{Y_n\in B\}}\right)\\
&= \sum_{n\ge1}\mathbb{P}(\{cU_1g(Y_1)\ge f(Y_1)\})^{n-1}\mathbb{P}(\{cU_1g(Y_1)< f(Y), Y_1\in B\})
\end{aligned}$$

where we used that the sequence $(U_n, Y_n)_{n\ge1}$ is i.i.d.. Hence

$$\begin{aligned}
\mathbb{P}(X\in B) &= \sum_{n\ge1}\mathbb{E}\left(\mathbf{1}_{\{\tau=n\}}\mathbf{1}_{\{Y_n\in B\}}\right)\\
&= \sum_{n\ge1}\mathbb{P}(\{cU_1g(Y_1)\ge f(Y_1)\})^{n-1}\mathbb{P}(\{cUg(Y)< f(Y_1), Y_1\in B\})\\
&= \mathbb{P}(Y\in B\,|\,cU_1g(Y_1)\le f(Y_1)).
\end{aligned}$$

Combining this with Step 1 yields $\mathbb{P}_{X} = f.\mu$. $\diamond$

Corollary 1 *Set by induction for every* $n \geq 1$

$$\tau_1 := \min\{k \geq 1 \,|\, c\, U_k g(Y_k) < f(Y_k)\} \quad \textit{and} \quad \tau_{n+1} := \min\{k \geq \tau_n + 1 \,|\, c\, U_k g(Y_k) < f(Y_k)\}.$$

then the sequence

$$X_n := Y_{\tau_n}$$

is an i.i.d. $\mathbb{P}_X$*-distributed sequence of r.v.*

The easy proof is left to the reader

Remark. The average yield of the rejection method is defined by $\bar{r} := \mathbb{P}(c\,U_1 g(Y_1) \leq f(Y_1))$: this means that in average one needs to simulate $\frac{1}{\bar{r}}$ to obtain one $\mathbb{P}_X$-distributed number.

Applications. $\triangleright$ *Uniform distributions on bounded domains* D. Let $D \subset [-M, M]^d$, $\lambda_d(D) > 0$ and let $Y \stackrel{d}{=} U([-M, M]^d)$ and $\tau := \min\{n \,|\, Y_n \in D\}$. Then,

$$X_\tau \stackrel{d}{=} U(D).$$

This follows (exercise) from the above proposition with

$$g(u) := (2M)^{-d} \underbrace{\mathbf{1}_{[-M,M]^d}(y).\lambda_d(dy)}_{\mu(dy)}$$

and

$$f(x) = \frac{1}{\lambda_d(D)} \mathbf{1}_D(x) \lambda_d(dx) \leq \frac{(2M)^d}{\lambda_d(D)} g(x).$$

A standard application is to consider the unit ball of $\mathbb{R}^d$, $D := B_d(0; 1)$. When $d = 2$, this is involved in the so-called *polar method*, see below, for the simulation of $\mathcal{N}(0; I_2)$ random vectors.

$\triangleright$ *The* $\gamma(\alpha)$*-distribution* Let $\alpha > 0$ and $\mathbb{P}_X(dx) = f_\alpha(x)\lambda(dx)$ where

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \mathbf{1}_{(0,+\infty)}(x).$$

(Keep in mind $\Gamma(a) = \int_0^{+\infty} u^{a-1} e^{-u} du$). Note that when $\alpha = 1$ the gamma distribution is but the exponential distribution.

– If $0 < \alpha < 1$, one uses the rejection method, based on the probability density

$$g_\alpha(x) = \frac{\alpha e}{\alpha + e} \left(x^{\alpha-1} \mathbf{1}_{\{0<x<1\}} + e^{-x} \mathbf{1}_{\{x \geq 1\}} \right).$$

First, one checks that $f_\alpha(x) \leq c_\alpha g_\alpha(x)$ where

$$c_\alpha = \frac{\alpha + e}{\alpha e \Gamma(\alpha)}.$$

Then, one uses the inverse distribution function to simulate the random variable with distribution $\mathbb{P}_Y(dy) = g_\alpha(y)\lambda(dy)$. Namely, if G_α denotes the distribution function of Y, one checks that, for every $u \in (0, 1)$,

$$G_\alpha^{-1}(u) = \left(\frac{\alpha + e}{e} u \right)^{\frac{1}{a}} \mathbf{1}_{\{u < \frac{e}{\alpha+e}\}} - \log \left((1 - u) \frac{\alpha + e}{\alpha e} \right) \mathbf{1}_{\{u \geq \frac{e}{\alpha+e}\}}$$

– If $\alpha \geq 1$, Then $X = X' + X''$, with X' and X'' are independent and X' has a gamma distribution with parameter $[\alpha]$ and X" has a gamma distribution with parameter $\{\alpha\} = \alpha - [\alpha]$. Consequently one may assume that $\alpha = n \in \mathbb{N}$. Then,

$$X = \xi_1 + \cdots + \xi_n$$

where ξ_k are i.i.d. with exponential distribution. Consequently, if $U_1, \ldots, U_n$ are i.i.d. uniformly distributed random variables

$$X \stackrel{d}{=} \log\left(\prod_{k=1}^{n} U_k\right).$$

1.5 The Box-Müller method for normal vectors

1.5.1 d-dimensional Normal vectors

One relies on the Box-Müller method, which is probably the most efficient method to simulate *couples* of bi-variate normal distributions.

Proposition 3 *Let R^2 et $\Theta : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be two independent r.v. with distributions $\mathcal{L}(R^2) = \mathcal{E}(\frac{1}{2})$ and $\mathcal{L}(\Theta) = U([0, 2\pi])$ respectively. Then*

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

where $R := \sqrt{R^2}$.

Proof. : Let f be a bounded Borel function.

$$\iint_{\mathbb{R}^2} f(x_1, x_2)\exp(-\frac{x_1^2 + x_2^2}{2})\frac{dx_1dx_2}{2\pi} = \iint f(\rho\cos\theta, \rho\sin\theta)e^{-\frac{\rho^2}{2}}\mathbf{1}_{\mathbb{R}_+^*}(\rho)\mathbf{1}_{]0,2\pi[}(\theta)\rho\frac{d\rho d\theta}{2\pi}$$

using the standard change of variable: $x_1 = \rho\cos\theta, x_2 = \rho\sin\theta$. Setting now $\rho = \sqrt{r}$, one has:

$$\iint_{\mathbb{R}^2} f(x_1, x_2)\exp(-\frac{x_1^2 + x_2^2}{2})\frac{dx_1dx_2}{2\pi} = \iint f(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta)\frac{e^{-\frac{r}{2}}}{2}\mathbf{1}_{\mathbb{R}_*^+}(\rho)\mathbf{1}_{]0,2\pi[}(\theta)\frac{drd\theta}{2\pi}$$

$$= \mathbb{E}\left(f(\sqrt{R^2}\cos\Theta, \sqrt{R^2}\sin\Theta)\right) = \mathbb{E}(f(X)). \qquad \diamond$$

Corollary 2 *One can simulate a distribution $\mathcal{N}(0; I_2)$ from a couple (U_1, U_2) of independent r.v. with distribution $U([0, 1])$ by setting*

$$X := \left(\sqrt{-2\log(U_1)}\cos(2\pi U_2), \sqrt{-2\log(U_1)}\sin(2\pi U_2)\right).$$

The yield of the simulation is $\rho = 1$.

Proof. Simulate the exponential distribution using the inverse distribution function and note that if $U \sim U([0, 1])$, then $\mathcal{L}(2\pi U) = U([0, 2\pi])$. $\diamond$

To simulate a d-dimensional vector $\mathcal{N}(0; I_d)$, one may assume that d is even and "concatenate" the above process using a d-tuple $(U_1, \ldots, U_d)$ of i.i.d. $U([0, 1])$ r.v..

Exercise (Polar method) Let $(U_1, U_2) \stackrel{d}{=} U(B(0; 1))$. Set

$$X := (U_1\sqrt{-2\log(R^2)/R^2}, U_2\sqrt{-2\log(R^2)/R^2}).$$

Show that $X \stackrel{d}{=} \mathcal{N}(0; I_2)$. Derive a simulation method for $\mathcal{N}(0; I_2)$ combining the above identity and some rejection algorithm.

1.5.2 d-dimensional Gaussian vectors (with general covariance matrix)

Let Σ be a covariance matrix and $X \stackrel{d}{=} \mathcal{N}(0;\Sigma)$. Σ is a symmetric nonnegative so there exists a unique symmetric nonnegative matrix commuting with Σ, denoted $\sqrt{\Sigma}$ such that $\sqrt{\Sigma}^2 = \Sigma$. A straightforward computation shows that if

$$Z \stackrel{d}{=} \mathcal{N}(0;I_d) \qquad \text{then} \qquad \sqrt{\Sigma}Z \stackrel{d}{=} \mathcal{N}(0;\Sigma)$$

One can compute $\sqrt{\Sigma}$ by diagonalizing Σ in the orthogonal group: since if $\Sigma = P^* Diag(\lambda_1,\ldots,\lambda_d)P$ then $\sqrt{\Sigma} = P^* Diag(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_d})P$ where P stands for the transpose of the orthogonal matrix P.

However, one will prefer usually rely on the Cholevsky method (see *e.g.* Numerical Recipes [37] by decomposing

$$\Sigma = TT^*$$

where T is an lower triangular matrix. Then

$$TZ \stackrel{d}{=} \mathcal{N}(0;\Sigma).$$

1.6 Vanilla options pricing in a Black-Scholes model by Monte Carlo

1.7 Premium computation

For the sake of simplicity, one considers a 2-dimensional correlated Black-Scholes model (under its unique risk neutral probability)but a general d-dimensional can be defined likewise.

$$\begin{array}{rcl} dX^0_t &=& rX^0_t dt, \quad X^0_0 = 1,\\ dX^1_t &=& X^1_t(rdt + \sigma_1 W^1_t), \quad X^1_0 = x^1_0,\\ dX^2_t &=& X^2_t(rdt + \sigma_1 W^2_t), \quad X^2_0 = x^2_0, \end{array}$$

with the usual notations (r interest rate, σ_1, σ_2 volatility. In particular, $W = (W^1, W^2)$ denotes a correlated bi-dimensional Brownian motion such that $<W^1, W^2>_t = \rho\, dt$. The filtration $\mathcal{F}$ is the augmented filtration of W.

Then, for every $t \in [0,T]$

$$\begin{aligned} X^0_t &= e^{rt}\\ X^1_t &= x^1_0 e^{(r-\frac{\sigma_1^2}{2})t+\sigma_1 W^1_t},\\ X^2_t &= x^2_0 e^{(r-\frac{\sigma_2^2}{2})t+\sigma_2 W^2_t}. \end{aligned}$$

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

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

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

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

The fact that W has independent stationary increments implies that X^1 and X^2 have independent stationary ratios so that if

$$V_0 := v(x_0, T)$$

then

$$\begin{aligned}
V_t &= e^{-r(T-t)}\mathbb{E}(h(X_T)\,|X_t) \\
&= e^{-r(T-t)}\mathbb{E}\left(h\left(\left(\frac{X^i_T}{X^i_t}\times X^i_t\right)_{i=1,2}\right)|S_t\right) \\
&= e^{-r(T-t)}\mathbb{E}(h((x^i\frac{X^i_{T-t}}{X^i_0})_{i=1,2})_{x^i=X^i_t} \\
&= v(X_t, T-t).
\end{aligned}$$

Examples. • Vanilla call: $h(x^1,x^2) = (x^1-K)_+$. There is a closed form for this option which is but the celebrated Black-Scholes formula

$$\mathrm{Call}_0^{BS} = C(x_0,K,T,r,\sigma) = s_0\Phi_0(d_1) - e^{-rT}K\Phi_0(d_2)$$

with

$$d_1 = \frac{\log(x_0^1/K) + (r+\frac{\sigma_1^2}{2})T}{\sigma_1\sqrt{T}}, \qquad d_2 = d_1 - \sigma_1\sqrt{T}.$$

(Φ_0 denotes the distribution function of the $\mathcal{N}(0;1)$-distribution.

• Best of call with strike price K:

$$h_T = (\max(X^1_T, X^2_T) - K)_+.$$

A quasi-closed form is available involving the distribution function of the bi-variate (correlated) normal distribution. It may be interesting to price it by MC (although PDE is also quite appropriate).

• Exchange Call Spread:

$$h_T = ((X^1_T - X^2_T) - K)_+.$$

For this payoff no closed form is available. One has the choice between a PDE approach (quite appropriate in this 2-dimensional setting) and a Monte Carlo simulation.

We will illustrate on this last example the regular Monte Carlo procedure.

Pricing by Monte Carlo : the regular procedure A crude Monte Carlo amounts to writing

$$e^{-rT}h_T \stackrel{d}{=} \varphi(Z^1,Z^2) := \left(s_0^1\exp\left(\frac{\sigma_1^2}{2}T + \sigma_1\sqrt{T}Z^1\right) - s_0^2\exp -\frac{\sigma_2^2}{2}T + \sigma_2\sqrt{T}Z^2) - Ke^{-rT}\right)_+$$

where $Z = (Z^1,Z^2) \stackrel{d}{=} \mathcal{N}(0;I_2)$ (the dependence of φ in x_0^i, etc is dropped). Then, simulating a M-sample $(Z_m)_{1\le m\le M}$ of the $\mathcal{N}(0;I_2)$ distribution using *e.g.* the Box-Müller yields the estimate

$$ExchSpread_0 = e^{-rT}\mathbb{E}((X^1_T - X^2_T) - K)_+) = \mathbb{E}(\varphi(Z^1,Z^2)) \approx \overline{\varphi}_M := \frac{1}{M}\sum_{m=1}^{M}\varphi(Z_m).$$

One computes an estimate for the variance using the same sample

$$\overline{V}_M(\varphi) = \frac{1}{M-1}\sum_{k=1}^{M}\varphi(Z_m)^2 - \frac{M}{M-1}(\overline{\varphi}_M)^2 \approx \mathrm{Var}(\varphi(Z))$$

since M is large enough. Then one designs a confidence interval for $\mathbb{E}\,\varphi(Z)$ at level $\alpha\in(0,1)$ by setting

$$I_M=\left[\overline{\varphi}_M-a_\alpha\sqrt{\frac{\overline{V}_M(\varphi)}{M}},\overline{\varphi}_M+a_\alpha\sqrt{\frac{\overline{V}_M(\varphi)}{M}}\right]$$

where a_α is defined by $\mathbb{P}(|\mathcal{N}(0;1)|\le a_\alpha)=\alpha$ (or equivalently $2\Phi_0(a_\alpha)-1=\alpha$. Some approximations are hidden in what precedes from a statistical viewpoint, in particular the fact that $\frac{\overline{\varphi}_M-\mathbb{E}\,\varphi(Z)}{\sqrt{\overline{V}_M(\varphi)}}$ does have a normal distribution, which is in fact only true asymptotically. However, within the usual range of simulation implemented for numerical purpose, this approximation is quite satisfactory, in fact more satisfactory than in many statistical applications where it is usually made.

1.8 Greeks (sensitivity to the option parameters: the elementary approach)

The *greeks* or sensitivities denote the set of parameters obtained as derivatives of the premium of an option with respect to some of its parameters: the starting value, the volatility, etc. In many reasonably elementary situations, one simply needs to apply some more or less standard theorem like

Theorem 3 *(Inverting differentiation and expectation) (a) Let $\Psi : I\times(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}$ where I denotes a nonempty interval of $\mathbb{R}$. Let $x_\infty\in I$. If the function Ψ satisfies*

(i) for every $x\in I$, the r.v. $\Psi(\xi,.)\in L^1_{\mathbb{R}}(\mathbb{P})$,

(ii) $\mathbb{P}(d\omega)$-a.s., $\dfrac{\partial\Psi}{\partial x}(x_\infty,\omega)$ exists,

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

$$\mathbb{P}(d\omega)\text{-a.s. }\;|\Psi(x,\omega)-\Psi(u_\infty,\omega)|\le Y(\omega)|x-x_\infty|,$$

then, the function $\psi(x):=\mathbb{E}(\Psi(x,.))$ is defined at every $\xi\in I$, differentiable at x_∞ with derivative

$$\psi'(x_\infty)=\mathbb{E}\left(\frac{\partial\Psi}{\partial x}(x_\infty,.)\right).$$

(b) If Ψ satisfies (i) and

$(ii)_{glob}$ $\mathbb{P}(d\omega)$-a.s., $\dfrac{\partial\Psi}{\partial x}(x,\omega)$ exists at every $x\in I$,

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

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

then, the function $\psi(x):=\mathbb{E}(\Psi(x,.))$ is defined and differentiable at every $\xi\in I$, with derivative

$$\psi'(x)=\mathbb{E}\left(\frac{\partial\Psi}{\partial x}(x,.)\right).$$

Remarks. • This is a special case of a general result of Integration theory and one can replace the probability space $(\Omega,\mathcal{A},\mathbb{P})$ by any measured space $(E,\mathcal{E},\mu)$.

• Some variants of the result can be established to get a theorem for differentiability of functions defined on $\mathbb{R}^d$ or for holomorphic functions, etc.

1.8.1 Working with the random process

To illustrate the different methods to compute the sensitivity, we will consider the 1-dimensional case of the Black-Scholes and temporarily change our notation by setting

$$dX_t^x = X_t^x(rdt + \sigma dW_t), \quad X_0^x = x > 0$$

so that $X_t^x = x \exp\left((r - \frac{\sigma^2}{2})t + \sigma W_t\right)$. Then we consider, for every $x \in (0, \infty)$,

$$f(x) = \mathbb{E}(\varphi(X_T^x)),$$

where $\varphi : (0, +\infty) \to \mathbb{R}$ is in $L^1(\mathbb{P}_{X_T^x})$. 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 in the Black-Scholes model. However, as soon as a closed form is available for the density of X_T^x, it is more efficient to use the next paragraph.

Proposition 4 *(a) If φ is differentiable with polynomial growth, then f is differentiable and*

$$f'(x) = \mathbb{E}\left(\varphi'(X_T^x)\frac{X_T^x}{x}\right).$$

(b) If φ is simply Borel function with polynomial growth, then f is still differentiable and

$$f'(x) = \mathbb{E}\left(\varphi(X_T^x)\frac{W_T}{x\sigma T}\right).$$

Proof. (a) This straightforwardly follows from the explicit expression for X_T^x and the above differentiation Theorem 3.

(b) Now, still under the assumption (a) (with $\mu := r - \frac{\sigma^2}{2}$),

$$\begin{aligned}
f'(x) &= \int_{\mathbb{R}} \varphi'(x \exp(\mu T + \sigma\sqrt{T}u)) \exp\left(\mu T + \sigma\sqrt{T}u\right) - \frac{u^2}{2}\frac{du}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} \frac{\partial \varphi(x \exp(\mu T + \sigma\sqrt{T}u))}{\partial u} \frac{\exp(-u^2/2)}{x\sigma\sqrt{T}}\frac{du}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} \varphi(x \exp(\mu T + \sigma\sqrt{T}u))u\frac{\exp(-u^2/2)}{x\sigma\sqrt{T}}\frac{du}{\sqrt{2\pi}}
\end{aligned}$$

where we used an integration by part in the last line. Finally, coming back to Ω,

$$f'(x) = \frac{1}{x\sigma T}\mathbb{E}\left(\varphi(X_T^x)W_T\right). \tag{1.1}$$

When φ is not differentiable, let us sketch the extension by density when φ has compact support. Then φ can be approximated in every $L^p(\mathbb{P})$ by differentiable functions φ_n with compact support (use a *mollifier* and a convolution approximation). Then, with obvious notations, $f_n'(x)$ converges uniformly on compact sets of $(0, \infty)$ to $f'(x)$ defined by (1.1). Furthermore $f_n(x)$ converges toward $f(x)$. Consequently f is differentiable with derivative f'.

Remark. Using item (a) of Theorem 3 and the fact that $\mathbb{P}(X^x = y) = 0$ for every $y > 0$, one derives that claim (a) in the proposition may remain true if φ is not differentiable at countably many points. This extends *e.g.* the first formula to the case of functions $\varphi(x) = (x - K)+$ or $(K - x)_+$.

▷ **Exercise:** *Application to the computation of the* γ*.* As a result

$$f"(x) := \frac{1}{x^2\sigma T}\mathbb{E}\left(\varphi'(X^x_{_T})W_{_T}X^x_{_T} - \varphi(X^x_{_T})W_{_T}\right)$$

if φ is differentiable with a derivative having linear growth, and

$$f"(x) := \frac{1}{x^2\sigma T}\mathbb{E}\left(\varphi(X^x_{_T})\left(\frac{W^2_{_T}}{\sigma T} - W_{_T} - \frac{1}{\sigma}\right)\right).$$

if φ is simply Borel with linear growth.

▷ *Computation of the vega.* One shows likewise the expression for the *vega i.e.* the derivative of the premium with respect to the volatility parameter σ under the same assumptions on φ, namely

$$\frac{\partial}{\partial\sigma}\mathbb{E}(\varphi(X^x_{_T})) = \mathbb{E}\left(\varphi'(X^x_{_T})(W_{_T} - \sigma T)X^x_{_T}\right) = \frac{1}{\sigma T}\mathbb{E}\left(\varphi(X^x_{_T})W_{_T}\right)$$

if φ is differentiable with a derivative with polynomial growth. An integration by part then shows that

$$\frac{\partial}{\partial\sigma}\mathbb{E}(\varphi(X^x_{_T}) = \mathbb{E}\left(\varphi(X^x_{_T})\left(\frac{W^2_{_T}}{\sigma T} - W_{_T} - \frac{1}{\sigma}\right)\right).$$

1.8.2 A direct approach based on differentiation on the state space

In fact, one can also carry on the computations directly on the state space of the process without specifying the Black-Scholes model provided the solution at time t, X^x_t, has an explicit probability density $p_t(x,y)\mu(dy)$ with respect to a reference measure μ on the real line. If θ denotes a real parameter of interest of the structure equation that defines X^x (which may turn to be x itself), one has at least formally

$$f(\theta) = \mathbb{E}(\varphi(X^x_{_T})) = \int_{\mathbb{R}}\varphi(y)p_{_T}(\theta,x,y)\mu(dy)$$

so that

$$\begin{aligned} f'(\theta) &= \int_{\mathbb{R}}\varphi(y)\frac{\partial p_{_T}}{\partial\theta}(\theta,x,y)\mu(dy)\\ &= \int_{\mathbb{R}}\varphi(y)\frac{\frac{\partial p_{_T}}{\partial\theta}}{p_t(\theta,x,y)}(\theta,x,y)p_{_T}(\theta,x,y)\mu(dy)\\ &= \mathbb{E}\left(\varphi(X^x_{_T})\frac{\partial\log(p_{_T})}{\partial\theta}(\theta,x,X^x_{_T})\right) \end{aligned} \tag{1.2}$$

Of course, the above computations need to be supported by appropriate assumptions (domination, etc).

Exercise (a) Compute the probability density $p_T(x,y)$ of $X^x_{_T}$.

(b) Re-establish all the above formulae using this approach.

A multi-dimensional version of this result can be established the same way round. However, this straightforward and simple approach to "greek" computation remains marginal outside the Black-Scholes world since it needs to have access to an explicit form for the probability density of the asset at time T. When dealing with path-dependent options (or even worse American options), this approach already fails even in a Black-Scholes model.

This is why we are lead to “go back” on the “scenarii” space Ω. Then, some extensions of the first approach are possible: if the function and the diffusion coefficients (when the risky asset prices follow a Brownian diffusion) are smooth enough, one usually relies on the so-called tangent process. A more sophisticate method is to introduce some Malliavin calculus methods which correspond to a differentiation theory with respect to the generic Brownian paths. This second topic is beyond the scope of the present course.

1.8.3 The tangent process method

In fact if the coefficients of the SDE are regular enough, one can differentiate directly the processes with respect to a given parameter. We refer to section 4.

2 Variance reduction

2.1 Static control variate

Let $X,\ X' \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ satisfying

$$\mathbb{E}\,X = \mathbb{E}\,X' = m \in \mathbb{R}, \qquad \mathrm{Var}(X),\ \mathrm{Var}(X'),\ \mathrm{Var}(X - X') > 0$$

$\triangleright$ The parameter m is to be computed by a Monte Carlo simulation. Let X_k, $k \geq 1$ be a sequence of i.i.d. copies of X. Then $(SLLN)$

$$m = \lim_{M\to\infty} \overline{X}_M \quad \mathbb{P}\text{-}a.s., \qquad \text{with} \quad \overline{X}_M := \frac{1}{M}\sum_{k=1}^{M} X_k$$

with a convergence ruled by the Central Limit Theorem (CLT)

$$\sqrt{M}\left(\overline{X}_M - m\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0; \mathrm{Var}(X)) \quad \text{as} \quad M \to \infty.$$

so that

$$\mathbb{P}\left(m \in \left[\overline{X}_M - a\frac{\sigma(X)}{\sqrt{M}}, \overline{X}_M + a\frac{\sigma(X)}{\sqrt{M}}\right]\right) \;\approx\; 2\Phi_0(a) - 1.$$

where $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ and $\Phi_0(x) = \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} \frac{d\xi}{\sqrt{2\pi}}$.

$\triangleright$ Question: Which random vector (distribution...) is more appropriate?

A natural answer is : if both X and X' can be simulated with an equivalent cost (complexity), then the one with the lowest variance is the best choice *i.e.*

$$X \quad \text{if} \quad \mathrm{Var}(X) < \mathrm{Var}(X'), \qquad X' \ \text{otherwise.}$$

provided this is known *a priori*.

$\triangleright$ Practical implementation. 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 r.v. $X - \Xi$ can be simulated with the same cost (complexity) than X.

(iii) the variance $\mathrm{Var}(X - \Xi) < \mathrm{Var}(X)$.

Then, the random variable

$$X' = X - \Xi + \mathbb{E}\,\Xi$$

can be simulated at the same cost as X, $\mathbb{E}X' = \mathbb{E}\,X$ and $\mathrm{Var}(X') < \mathrm{Var}(X)$.

$\triangleright$ In option pricing the payoffs are usually nonnegative. In that case, any r.v. Ξ satisfying (i)-(ii) and

$$0 \leq \Xi \leq X$$

can be considered as a good candidate to reduce the variance, especially if Ξ is not too far from X in some way. However, note that it does not imply (iii) (if $X \equiv 1$, then $\mathrm{Var}(X) = 0$ whereas a uniformly distributed random variable Ξ on $[1-\eta, 1]$ has clearly a nonzero variance...).

2.1.1 Jensen inequality and variance reduction

Jensen inequality is an efficient tool to design control variate when dealing with path-dependent exotic option pricing as illustrated by the following examples:

Examples: • *Asian options and Kemna-Vorst control variate in a Black-Scholes dynamics (see [29])*
Let $h_T = \varphi\left(\frac{1}{T}\int_0^T X_t^x dt\right)$ where φ is nonnegative and non-decreasing and

$$X_t^x = x \exp\left((r - \frac{\sigma^2}{2})t + \sigma W_t\right), \quad x > 0$$

is a Black-Scholes dynamics with volatility σ and interest rate r. Then, Jensen inequality applied to the probability measure $\frac{1}{T}\mathbf{1}_{[0,T]}(t)dt$ implies

$$\begin{aligned}\frac{1}{T}\int_0^T X_t^x\,dt &\geq x\exp\left(\frac{1}{T}\int_0^T (r - \sigma^2/2)t + \sigma W_t)dt\right)\\ &= x\exp\left(\frac{1}{2}(r-\sigma^2/2)T + \frac{\sigma}{T}\int_0^T W_t dt)\right).\end{aligned}$$

Now

$$\int_0^T W_t dt = T\,W_T - \int_0^T s dW_s = \int_0^T (T-s)dW_s$$

so that

$$\frac{1}{T}\int_0^T W_t dt \stackrel{d}{=} \mathcal{N}\left(0; \frac{1}{T^2}\int_0^T s^2 ds\right) = \mathcal{N}\left(0; \frac{T}{3}\right).$$

Hence

$$\frac{1}{T}\int_0^T X_t^x\,dt \geq e^{-(r/2+\sigma^2/12)T}\,x\exp\left((r - (\sigma^2/3)/2)T + (\sigma/\sqrt{3})\sqrt{T}Z\right)$$

for some normally distributed r.v. Z, so that

$$h_T \geq h_T^{KV} \stackrel{d}{=} \varphi\left(xe^{-(r/2+\sigma^2/12)T}\exp\left((r-(\sigma^2/3)/2)T + (\sigma/\sqrt{3})W_T)\right)\right)$$

Rule: If the vanilla option with payoff $\varphi(X_T^x)$ has a closed form, so is the case for the Kemna-Vorst payoff h_T^{KV}.

• *Basket option.* One considers a payoff on a basket of options (*e.g.* an index), say a call

$$h_T = \left(\sum_{k=1}^{N} \alpha_k X_T^{k,x_k} - K\right)_+$$

where $(X^1, \dots, X^d)$ is d-dimensional basket of risky assets and all the weights $\alpha_k > 0$, $\sum_{1\le k\le d} \alpha_k = 1$. Then the convexity of the exponential implies that

$$e^{\sum_{1\le k\le d} \alpha_k \log(X_T^{k,x_k})} \le \sum_{k=1}^{d} \alpha_k X^{k,x_k}$$

In a d-dimensional Black-Scholes model, $\sum_{1\le k\le d} \alpha_k \log(X_T^{k,x_k})$ still has a normal distribution so that the payoff $k_T := (e^{\sum_{1\le k\le d} \alpha_k \log(X_T^{k,x_k})} - K)_+$ gives raise to a closed form. The extension to more general payoffs of the basket is straightforward provided a closed form is available when the basket is replaced by a single *B-S* risky asset.

• *Best of call option.* The payoff is given by

$$h_T = (\max(X_T^1, X_T^2) - K)_+$$

Using that $\sqrt{ab} \le \max(a, b)$, a, $b > 0$, its is clear that

$$h_T \ge k_T = \left(\sqrt{X_T^1 X_T^2} - K\right)_+$$

and that, still in a 2-dimensional Black-Scholes model, the option with payoff k_T has a closed form. One may improve the procedure by noting that more generally $a^\theta b^{1-\theta} \le \max(a, b)$ when $\theta \in (0, 1)$.

2.2 Negatively correlated variables with the same expectation and variance

When $\mathrm{Var}(X) = \mathrm{Var}(X')$, choosing X or X' seems of little interest. however, it may be possible to take advantage of this situation to induce a variance reduction.

Assume $\mathrm{Var}(X) = \mathrm{Var}(X')$, X and X' can be simulated with the same complexity κ. Then set

$$\Xi = \frac{X + X'}{2}$$

It is reasonable (when no further information on (X, X') is available) to assume that the simulation complexity of Ξ is twice that of X and X', *i.e.* 2κ. On the other hand

$$\mathrm{Var}\,(\Xi) = \frac{\mathrm{Var}(X) + \mathrm{Cov}(X, X')}{2}$$

The size of the simulation using X (or X') and Ξ respectively to enter a given interval $[m - \varepsilon, m + \varepsilon]$ with the same confidence level $2\Phi_0(a) - 1 > 0$ $(a > 0)$ is

$$M^X = \frac{a^2 \mathrm{Var}(X)}{\varepsilon^2} \quad \text{with } X \text{ and} \quad M^\Xi = \frac{a^2 \mathrm{Var}(\Xi)}{\varepsilon^2} \quad \text{with } \Xi.$$

Taking into account the complexity, that means essentially the computation *CPU* time, one should better use Ξ if and only iff $2\kappa\, M^\Xi < \kappa\, M^X$ *i.e.*

$$2\mathrm{Var}(\Xi) < \mathrm{Var}(X)$$

which amounts finally to

$$\mathrm{Cov}(X, X') < 0.$$

To use this remark in practice, one may rely on the following simple result.

Proposition 5 *Let* $X = \varphi(Z) \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$*, where* $\varphi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is a monotone function. Assume that there exists a non-increasing transform* $T : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that* $Z \stackrel{d}{=} T(Z)$*. Then*

$$\mathrm{Cov}(f(Z), f(T(Z))) \le 0.$$

Proof. Without loss of generality one may assume f non-decreasing. Let Z, Z' two independent random variables defined on the same probability space with distribution $\mathbb{P}_Z$. Then T being non-increasing

$$(f(Z) - f(Z')(f(T(Z)) - f(T(Z')) \le 0$$

hence its expectation. Consequently

$$\mathbb{E}(f(Z)f(T(Z))) + \mathbb{E}(f(Z')f(T(Z'))) - \mathbb{E}(f(Z)f(T(Z')) - \mathbb{E}(f(Z')f(T(Z))) \quad \le \quad 0$$

so that using that $Z' \stackrel{d}{=} Z$ and $Z\ Z'$ are independent

$$2\mathbb{E}(f(Z)f(T(Z))) \le \mathbb{E}(f(Z))\mathbb{E}(f(T(Z')) + \mathbb{E}(f(Z'))\mathbb{E}(f(T(Z))) = 2\mathbb{E}(f(Z))\mathbb{E}(f(T(Z))$$

that is

$$\mathrm{Cov}(f(Z), f(T(Z))) = \mathbb{E}(f(Z)f(T(Z))) - \mathbb{E}(f(Z))\mathbb{E}(f(T(Z)) \le 0 \qquad \diamond$$

The classical situation in which such an approach successfully applies is when $T(z) = -z$.

Example. European option Pricing in BS model. Let $h_T = h(X^x_T)$ with h monotone (like for Calls, Puts, spreads, etc). Then $h_T = h(x \exp{(r - \frac{\sigma^2}{2}T + \sqrt{T}Z)})$, $Z \stackrel{d}{=} \mathcal{N}(0;1)$. The function $z \mapsto h(x \exp{(r - \frac{\sigma^2}{2}T + \sqrt{T}z)})$ is monotone as the composition of two monotone functions and $W_T \stackrel{d}{=} -W_T$.

2.3 Adaptive control variate

The situation of two square integrable r.v. X and X', $X \not\equiv X'$ having the same expectation

$$\mathbb{E}\,X = \mathbb{E}\,X' = m$$

(and nonzero variances $\mathrm{Var}(X)$ and $\mathrm{Var}(X')$) can be reformulated by setting

$$Y := X - X' \qquad \text{with} \quad \mathbb{E}\,Y = 0 \qquad \text{and } \mathrm{Var}(Y) > 0$$

We saw that if $\mathrm{Var}(X - Y) \ll \mathrm{Var}(X)$, one will choose $X - Y$ to implement the Monte Carlo simulation and we provided several classical examples in that direction.

However there is a way to optimally use this idea which is to parametrize the problem as follows. For convenience, set

$$X^\lambda = X - \lambda\, Y.$$

Then

$$\Phi(\lambda) = \lambda^2 \mathrm{Var}(Y) - 2\,\lambda\,\mathrm{Cov}(X, Y) + \mathrm{Var}(X)$$

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

$$\boxed{\lambda_{\min} := \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} = 1 - \frac{\mathrm{Cov}(X',Y)}{\mathrm{Var}(Y)}}$$

and (if $\rho_{X,Y}$ denotes the correlation coefficient of X and Y)

$$\boxed{\sigma^2_{\min} := \mathrm{Var}(X^{\lambda_{\min}}) = \mathrm{Var}(X) - \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(Y)} = \mathrm{Var}(X') - \frac{(\mathrm{Cov}(X',Y))^2}{\mathrm{Var}(Y)}.}$$

Hence

$$\boxed{\sigma^2_{\min} \le \min\left(\mathrm{Var}(X), \mathrm{Var}(X')\right)}$$

and

$$\boxed{\sigma^2_{\min} = \mathrm{Var}(X)(1-\rho^2_{X,Y}) = \mathrm{Var}(X')(1-\rho^2_{X',Y}).}$$

A more symmetric expression for $\mathrm{Var}(X^{\lambda_{\min}})$ is

$$\begin{aligned}
\sigma^2_{\min} &= \frac{\mathrm{Var}(X)\mathrm{Var}(X')(1-\rho^2_{X,X'})}{(\mathrm{Var}(X)-\mathrm{Var}(X'))^2 + 2\sqrt{\mathrm{Var}(X)\mathrm{Var}(X')}(1-\rho_{X,X'})} \\
&\le \sqrt{\mathrm{Var}(X)\mathrm{Var}(X')}\,\frac{1+\rho_{X,X'}}{2}.
\end{aligned}$$

2.3.1 Implementation of the adaptive Variance reduction

Let X_k, X'_k, $k \ge 1$, be (simulated...) independent copies of (X, X').
Set, for (a large enough *fixed*) integer $M \ge 1$:

$$V_M := \frac{1}{M}\sum_{k=1}^{M}(X_k - X'_k)^2 \qquad C_M := \frac{1}{M}\sum_{k=1}^{M}(X_k - \overline{X}_M)(X_k - X'_k)$$

$$\lambda_M := \frac{C_M}{V_M} \qquad \text{[can be updated recursively]}$$

▷ The "batch" approach (Glasserman, [20])
Then $\lambda_M \to \lambda_{\min}$ $\mathbb{P}$-*a.s.* so that one derives by elementary arguments that

$$\frac{1}{M}\sum_{k=1}^{M} X_k^{\lambda_M} \xrightarrow{a.s.} \mathbb{E}X = m$$

Not recursive at all... And what about rates ?...

▷ Recursive approach/implementation

Theorem 4 *If $X \in L^p(\mathbb{P})$ for any $p \ge 1$ and $\frac{1}{Y} = \frac{1}{X-X'} \in L^{1+\eta}(\mathbb{P})$ for some $\eta > 0$ then the expected variance reduction does occur in the following sense:*

$$\sqrt{M}\left(\frac{1}{M}\sum_{k=1}^{M} X''_k - m\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0;\sigma^2_{\min})$$

where $X''_k = X_k - \lambda_{k-1}Y_k = (1-\lambda_{k-1})X_k + \lambda_{k-1}X'_k$ and λ_k is defined as above.

The rest of this paragraph can be omitted at the occasion of a first reading.

Proof. • Assume (temporarily) that

$$\lambda_M = \frac{C_M}{V_M} \stackrel{L^2(\mathbb{P})}{\longrightarrow} \lambda_{\min} \qquad \text{as } M \to \infty.$$

Note that (C_M, V_M) can be recursively computed from (C_{M-1}, V_{M-1}) and (X_M, X'_M).

Let $\mathcal{F}_k := \sigma(X_1, X'_1, \dots, X_M, X'_k)$, $k \geq 1$, denote the filtration of the simulation and set for every $k \geq 1$,

$$\boxed{X''_k = X_k - \lambda_{k-1} Y_k = (1-\lambda_{k-1})X_k + \lambda_{k-1} X'_k.}$$

$$\begin{aligned}
\forall\, k \geq 1,\ \mathbb{E}(X''_k \,|\, \mathcal{F}_{k-1}) &= \mathbb{E}(X_k \,|\, \mathcal{F}_{k-1}) - \lambda_{k-1}\mathbb{E}(Y_k \,|\, \mathcal{F}_{k-1}) = m \\
\mathrm{Var}(X''_k) &= \mathbb{E}\left(\mathbb{E}((X''_k - m)^2 \,|\, \mathcal{F}_{k-1})\right) = \mathbb{E}\left(\mathrm{Var}(X^\lambda_k)_{|\lambda=\lambda_{k-1}}\right) \\
&= \mathbb{E}\left(\Phi(\lambda_{k-1})\right) \stackrel{k\to\infty}{\longrightarrow} \Phi(\lambda_{\min}) = \min_\lambda \Phi(\lambda).
\end{aligned}$$

• Now, set for every $M \geq 1$,

$$N_M := \sum_{k=1}^{M} \frac{X''_k - m}{k}.$$

$(N_M)_{M\geq 1}$ is an $L^2((\mathcal{F}_k)_k, \mathbb{P})$-martingale since $X''_k, k \geq 1$ is a sequence of $\mathcal{F}_k$-martingale increments and

$$\mathbb{E}(N^2_M) = \sum_{k=1}^{M} \frac{\mathbb{E}((X''_k - m)^2)}{k^2} = \sum_{k=1}^{M} \frac{\mathrm{Var}(X''_k)}{k^2} \leq C \sum_{k\geq 1} \frac{1}{k^2} < +\infty.$$

Hence $N_M \to N_\infty \in L^2(\mathbb{P})$ ($\mathbb{P}$-*a.s.* and in $L^2(\mathbb{P})$) as $M \to \infty$. Consequently, by the *Kronecker Lemma* (see below),

$$\frac{1}{M}\sum_{k=1}^{M} X''_k - m \stackrel{a.s.}{\longrightarrow} 0$$

$$\Longrightarrow \qquad \boxed{\overline{X}''_M := \frac{1}{M}\sum_{k=1}^{M} X''_k \xrightarrow{a.s.\&L^2(\mathbb{P})} m \qquad \text{as} \qquad M \to \infty.}$$

Lemma 1 Kronecker Lemma *Let $(a_n)_{n\geq 1}$ be a sequence of real numbers and let $(b_n)_{n\geq 1}$ be a non decreasing sequence of positive real numbers with* $\lim_n b_n = +\infty$. *Then*

$$\left(\sum_{n\geq 1} \frac{a_n}{b_n} \textit{ converges in } \mathbb{R} \textit{ as a series}\right) \Longrightarrow \left(\frac{1}{b_n}\sum_{k=1}^{n} a_k \longrightarrow 0 \ \textit{ as } \ n \to \infty\right).$$

• (Weak) Rate of convergence: One applies the Lindeberg *CLT* (see Hall & Heyde) to the *array of martingale increments* defined by

$$X_{M,k} := \frac{X''_k - m}{\sqrt{M}},\ 1 \leq k \leq M.$$

One checks that

$$\begin{aligned}\sum_{k=1}^{M}\mathbb{E}(X_{M,k}^2\,|\,\mathcal{F}_{k-1}) &= \frac{1}{M}\sum_{k=1}^{M}\mathbb{E}((X_k''-m)^2\,|\,\mathcal{F}_{k-1})\\ &= \frac{1}{M}\sum_{k=1}^{M}\Phi(\lambda_{k-1})\\ &\longrightarrow \sigma^2_{\min}:=\min_{\lambda}\Phi(\lambda)\end{aligned}$$

+ Lindeberg condition... which needs $\sup_M \mathbb{E}(\lambda_{_M}^{2+\eta})<+\infty\ldots$

$$\Longrightarrow \qquad \boxed{\sqrt{M}\left(\frac{1}{M}\sum_{k=1}^{M}X_k''-m\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0;\sigma^2_{\min})}$$

• Remaining task : a criterion for the assumption $\sup_M \mathbb{E}(\lambda_{_M}^{2+\eta})<+\infty$.

$$\begin{aligned}\lambda_{_M} &= \frac{\sum_{k=1}^{M}(X_k-\overline{X}_{_M})(X_k-X_k')}{\sum_{k=1}^{M}(X_k-X_k')^2}\\ \text{so that}\quad \lambda^2_{_M} &\le \frac{\sum_{k=1}^{M}(X_k-\overline{X}_{_M})^2}{\sum_{k=1}^{M}(X_k-X_k')^2}\qquad\text{(Schwarz Inequality)}\\ &= \frac{1}{M}\sum_{k=1}^{M}(X_k-\overline{X}_{_M})^2\times\frac{M}{\sum_{k=1}^{M}(X_k-X_k')^2}\\ &\le \frac{1}{M}\sum_{k=1}^{M}(X_k-\overline{X}_{_M})^2\times\frac{1}{M}\sum_{k=1}^{M}\frac{1}{(X_k-X_k')^2}\end{aligned}$$

where we used the convexity inequality

$$\frac{M}{\sum_{1\le k\le M}a_k}\le\frac{1}{M}\sum_{1\le k\le M}\frac{1}{a_k}.$$

By Hölder Inequality with conjugate exponents p, $q\in(1,+\infty)$

$$\begin{aligned}\forall\,\varepsilon>0,\qquad \left\|\lambda^2_{_M}\right\|_{1+\varepsilon} &\le \left\|\frac{1}{M}\sum_{k=1}^{M}(X_k-\overline{X}_{_M})^2\right\|_{p(1+\varepsilon)}\left\|\frac{1}{M}\sum_{k=1}^{M}\frac{1}{(X_k-X_k')^2}\right\|_{q(1+\varepsilon)}\\ &\le \left\|X_1-\overline{X}_{_M}\right\|^2_{2p(1+\varepsilon)}\left\|\frac{1}{|X-X'|}\right\|^2_{q(1+\varepsilon)}\\ &\le 4\,\|X\|^2_{2p(1+\varepsilon)}\left\|\frac{1}{|X-X'|}\right\|^2_{q(1+\varepsilon)}.\end{aligned}$$

2.3.2 Application to option pricing: using parity equations

One often starts form the (intuitive and well-known) Parity equations. Furthermore these relations are *model free* so they can be applied for various dynamics for the underlying asset.

We denote in this example by X_t the risky asset (with $X_0=x_0$) and set $X^0_r=e^{rt}$.

▷ Vanilla Call-Put parity ($d=1$):

$$\mathrm{Call}_0=e^{-rT}\mathbb{E}\left((X_{_T}-K)_+\right)\qquad\text{and}\qquad \mathrm{Put}_0=e^{-rT}\mathbb{E}\left((K-X_{_T})_+\right)$$

$$\boxed{\mathrm{Call}_0 - \mathrm{Put}_0 = x_0 - e^{-rT}K}$$

so that $\mathrm{Call}_0 = \mathbb{E}(X) = \mathbb{E}(X')$ with $X := e^{-rT}(X_T - K)_+$ and $X' := e^{-rT}(K - X_T)_+ + x_0 - e^{-rT}K$. As a result on set

$$Y = e^{-rT}X_T - x_0$$

which turns out te be the terminal value of a martingale

$\triangleright$ Asian Call-Put parity:

At $T_0 \in [0, T)$ starts the *averaging* interval $[T_0, T]$.
$\mathrm{Call}_0 = e^{-rT}\mathbb{E}\left(\left(\frac{1}{T-T_0}\int_{T_0}^{T} X_t dt - K\right)_+\right)$
$\mathrm{Put}_0 = e^{-rT}\mathbb{E}\left(\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} X_t dt\right)_+\right)$.

Using that $\widetilde{X}_t = e^{-rt}X_t$ is a $\mathbb{P}$-martingale and Fubini's theorem yield

$$\boxed{\mathrm{Call}_0^{As} - \mathrm{Put}_0^{As} = x_0\frac{1 - e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K}$$

so that

$$\mathrm{Call}_0^{As} = \mathbb{E}(X) = \mathbb{E}(X')$$

with

$$\begin{aligned} X &:= e^{-rT}\left(\frac{1}{T-T_0}\int_{T_0}^{T} X_t dt - K\right)_+ \\ X' &:= s_0\frac{1 - e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K + e^{-rT}\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} X_t dt\right)_+. \end{aligned}$$

which leads to

$$Y = e^{-rT}\frac{1}{T-T_0}\int_{T_0}^{T} X_t dt - s_0\frac{1 - e^{-r(T-T_0)}}{r(T-T_0)}.$$

Remarks. • In both cases, this relies on the $\mathbb{P}$-*martingale property* of $\widetilde{S}_t$.

• *Vanilla Call & Put* (model-free): The assumptions of the theorem are satisfied as soon as

$$X_T \in L^p(\mathbb{P}),\ p \in [1 + \infty) \qquad \text{and} \qquad \forall\, x \geq 0,\ \frac{1}{X_T - x} \in L^{1+\eta}(\mathbb{P}) \ \text{ for some } \eta > 0$$

This is satisfied by the *B-S* model.

• *Asian Call & Put* (in any model-free): Same condition involving

$$\frac{1}{T-T_0}\int_{T_0}^{T} X_t\, dt.$$

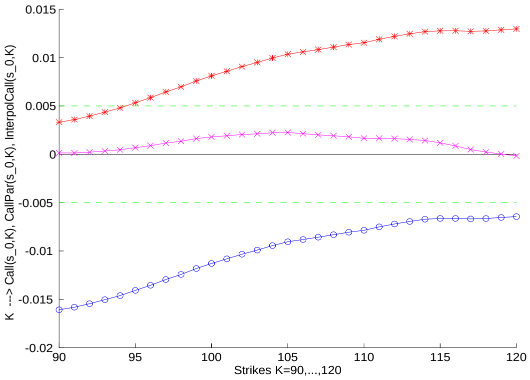


Figure 1: BLACK-SCHOLES CALLS: *Error= Reference BS−(MC Premium).* $K = 90, \ldots, 120$. *–o–o–o– Crude Call. –∗–∗–∗– Synthetic Parity Call. –×–×–×– Interpolated synthetic Call.*

2.3.3 Complexity:

In practical implementation, one often neglects the cost of the computation of $\lambda_{\min}$ since one only needs a rough estimate or it: this leads to stop its computation after the first 10% or 20% of the simulation.

– In the examples derived form "Parity equations" developed in the above subsection, the r.v. Y is involved in the simulation of X, so the complexity of the simulation process *is not* increased: updating λ_M and (the empirical mean) $\overline{X}''_M$ is (almost) costless. Subsequently, in that setting, the complexity remains the same!

– WARNING ! This no longer true in general . . . When the complexity is doubled, the method is efficient iff

$$\sigma^2_{\min} < \frac{1}{2}\min(\mathrm{Var}(X), \mathrm{Var}(X')).$$

if one neglects the cost of the estimation of the coefficient $\lambda_{\min}$.

2.3.4 Some Simulations

▷ Vanilla B-S Calls

Model parameters:

$$\boxed{T = 1,\ x_0 = 100,\ r = 5\,\%,\ \sigma = 20\,\%,\ K = 90, \ldots, 120.}$$

MC parameter: $M = 10^6$.

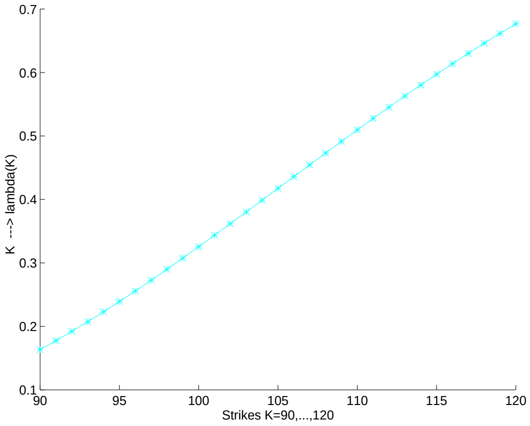


Figure 2: BLACK-SCHOLES CALLS: $K \mapsto 1 - \lambda_{\min}(K)$, $K = 90, \ldots, 120$, *for the Interpolated synthetic Call.*

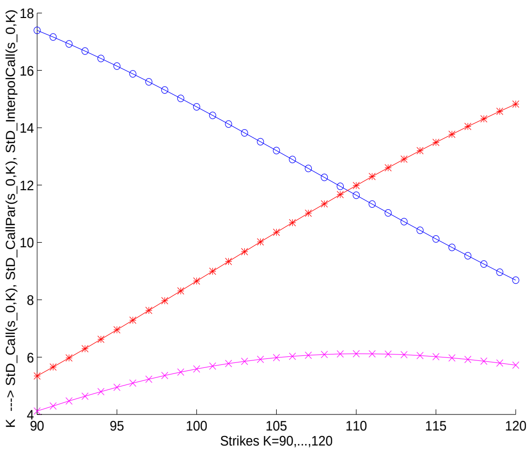


Figure 3: BLACK-SCHOLES CALLS. *Standard Deviation(MC Premium).* $K = 90, \ldots, 120$. –o–o–o– *Crude Call.* –∗–∗–∗– *Parity Synthetic Call.* –×××– *Interpolated Synthetic Call.*

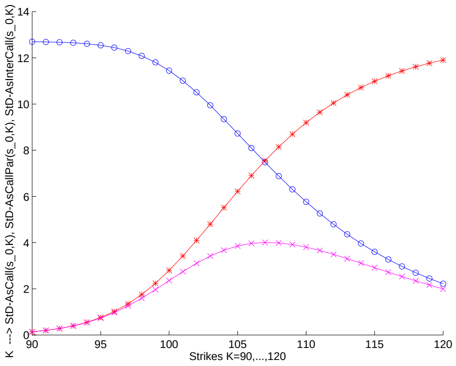


Figure 4: HESTON ASIAN CALLS. *Standard Deviation (MC Premium).* $K = 90, \ldots, 120$. –o–o–o– *Crude Call.* –∗–∗–∗– *Synthetic Parity Call.* –×–×–×– *Interpolated synthetic Call.*

▷ ASIAN HESTON CALLS

– The dynamics: Let ϑ, k, a s.t. $\vartheta^2/(2ak) < 1$.

$$dX_t = X_t(r\,dt + \sqrt{v_t})dW_t^1, \qquad X_0 = x_0 > 0, \quad \text{(risky asset)}$$

$$dv_t = k(a - v_t)dt + \vartheta\sqrt{v_t}dW_t^2,\ v_0 > 0 \quad \text{with } <W^1, W^2>_t = \rho\,t,\ \rho \in [-1, 1].$$

– The payoff and the premium: no closed form available for Asian payoffs!

$$\text{AsCall}^{Hest} = e^{-rT}\mathbb{E}\left(\left(\frac{1}{T}\int_0^T X_s ds - K\right)_+\right).$$

Note however that (quasi-)closed forms do exist for vanilla European options in this model (see [27]) which is the origin of its success.

– Parameters of the model:

$$\boxed{x_0 = 100,\ k = 2,\ a = 0.01,\ \rho = 0.5,\ v_0 = 10\%,\ \vartheta = 20\%.}$$

– Parameters of the option portfolio:

$$\boxed{T = 1,\ \ K = 90, \cdots, 120 \qquad \text{(31 strikes)}.}$$

Exercise. One considers a 1-dimensional Black-Scholes model with market parameters

$$r = 0, \quad \sigma = 0.3, \quad x_0 = 100, \quad T = 1$$

1. One considers a vanilla Call with strike $K = 80$. The r.v. Y is defined as above. Estimate the $\lambda_{\min}$ (one should be not far from 0.825). Then compute a confidence interval for the Monte Carlo pricing of the Call with and without the linear variance reduction for the following sizes of the simulation: $M = 5\,000,\ 10\,000, 100\,000,\ 500\,000$.

2. Proceed as above but with $K = 150$ (true price 1.49). What do you observe ? Provide an interpretation.

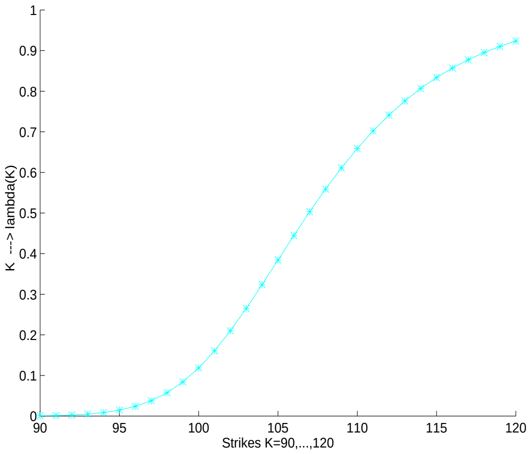


Figure 5: HESTON ASIAN CALLS. *$K \mapsto 1 - \lambda_{\min}(K)$, $K = 90, \ldots, 120$, for the Interpolated Synthetic Asian Call.*

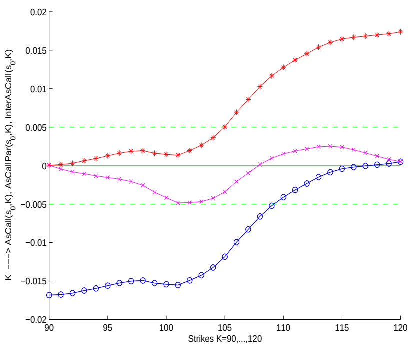


Figure 6: HESTON ASIAN CALLS. *$M = 10^6$ (Reference: MC with $M = 10^8$). $K = 90, \ldots, 120$. -o-o-o- Crude Call. -∗-∗-∗- Parity Synthetic Call. -×-×-×- Interpolated Synthetic Call.*

2.4 Multidimensional case

$\triangleright$ Let $X := (X^1, \ldots, X^d),\ Y := (Y^1, \ldots, Y^q) : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^q$, *square integrable* random vectors.

$$\mathbb{E}\, X = m \in \mathbb{R}^d, \qquad \mathbb{E}(Y) = 0 \in \mathbb{R}^q$$

Let $D(X) := [\mathrm{Cov}(X^i, X^j)]_{1 \le i,j \le d}$ and $D(Y)$ denote the covariance (dispersion) matrices of X and Y respectively.

$$D(X) \ \text{ and } \ D(Y) > 0$$

i.e. positive definite symmetric.

$\triangleright$ Problem Find a matrix $\Lambda \in \mathcal{M}(d, q)$ solution of the optimization problem

$$\mathrm{Var}(X - \Lambda\, Y) = \min \left\{ \mathrm{Var}(X - L\, Y),\ L \in \mathcal{M}(d, q) \right\}.$$

$\triangleright$ Solution

$$\Lambda = (D(Y))^{-1}(C(X, Y))$$

where

$$C(X, Y) = [\mathrm{Cov}(X^i, Y^j)]_{1 \le i \le d, 1 \le j \le q}.$$

$\triangleright$ Examples: Traded assets $X_t = (X_t^1, \ldots, X_t^d)$, $t \in [0, T]$.

– Options on various baskets

$$X^i = \left(\sum_{j=1}^{d} \theta_j^i X_T^j - K \right)_+ , \ i = 1, \ldots, d$$

Remark. Also produces an optimal *asset selection* (PCA) which helps for hedging.

– Portfolio of *forward start options*

$$X^{i,j} = \left(X_{T_{i+1}}^j - X_{T_i}^j \right)_+, \ i = 1, \ldots, d$$

2.5 Importance sampling (introduction to)

The basic principle of importance sampling is the following: let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ be an E-valued r.v. . Let μ be a σ-finite measure on $(E, \mathcal{E})$ satisfying $\mathbb{P}_X \ll \mu$ *i.e.* there exists a probability density $f : (E, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that

$$\mathbb{P}_X = f.\mu$$

In practice this means that one has to simulate several r.v. , whose distributions are all absolutely continuous with respect to this *reference* measure μ. For a first reading one may assume that $E = \mathbb{R}$ and μ is the Lebesgue measure but what follows can also be applied on more general measured spaces like the Wiener space, etc. Then

$$\mathbb{E} h(X) = \int_E h(x) \mathbb{P}_X(dx) = \int_E h(x) f(x) \mu(dx).$$

Now for any probability distribution g defined on $(E, \mathcal{E})$ (with respect o μ), one has

$$\mathbb{E} h(X) = \int_E h(x) f(x) \mu(dx) = \int_E \frac{h(x) f(x)}{g(x)} g(x) \mu(dx).$$

One can always enlarge (if necessary) the original probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to design a random variable $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ having g as a probability density with respect to μ. Then, going back on the probability space yields

$$\mathbb{E}\, h(X) = \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right). \tag{2.3}$$

So, in order to compute $\mathbb{E}\, h(X)$ on can also implement a Monte Carlo simulation based on the simulation of independent copies of the r.v. Y *i.e.*

$$\mathbb{E}h(X) = \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right) = a.s. \lim_{M\to\infty} \frac{1}{M}\sum_{\ell=1}^{M} \frac{h(Y_\ell)f(Y_\ell)}{g(Y_\ell)}.$$

$\triangleright$ *Necessary conditions ...* to undertake the simulation. To proceed it is necessary to simulate Y and to compute the ratio of density functions f/g at a reasonable cost (note that only the ratio is needed which can make useless the computation of some "structural" constants).

$\triangleright$ *Sufficient conditions ...* to undertake the simulation. Once the above conditions are fulfilled, the question is: is it profitable to proceed like that? So is the case if the complexity of the simulation for a given accuracy (in terms of confidence interval) is lower with the second method. If one assume for simplicity that simulating X and Y on the one hand and computing $h(x)$ and $(hf/g)(x)$ on the other hand is comparable *the question amounts to comparing the variances.*

Now

$$\begin{aligned}
\mathrm{Var}\left(\frac{h(Y)f(Y)}{g(Y)}\right) &= \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2 - (\mathbb{E}\, h(X))^2\\
&= \int_E \left(\frac{h(x)f(x)}{g(x)}\right)^2 g(x)\mu(dx) - (\mathbb{E}\, h(X))^2\\
&= \int_E \frac{(h(x)f(x))^2}{g(x)}\mu(dx) - (\int_E h(x)\mu(dx))^2
\end{aligned}$$

As a consequence simulating Ywill reduce the variance iff

$$\int_E \frac{(h(x)f(x))^2}{g(x)}\mu(dx) < \int_E h^2(x)f(x)\mu(dx).$$

Remarks. • In fact, theoretically, one may reduce the variance of the new simulation to $\ldots 0$. Assume $\mathbb{E}\, h(X) \neq 0$ and $g(x) > 0$ $\mu(dx)$-*a.s.*. As a matter of fact, using Schwarz Inequality one gets,

$$\begin{aligned}
(\int_E h(x)\mu(dx))^2 &= \int_E \frac{h(x)f(x)}{\sqrt{g(x)}}\sqrt{g(x)}\mu(dx)\\
&\leq \int_E \frac{(h(x)f(x))^2}{g(x)}\sqrt{g(x)}\mu(dx) \times \int_E g\, d\mu\\
&= \int_E \frac{(h(x)f(x))^2}{g(x)}\mu(dx)
\end{aligned}$$

since g is a probability density. Now the equality case in Schwarz inequality says that the variance is 0 iff $\sqrt{g(x)}$ and $\frac{h(x)f(x)}{\sqrt{g(x)}}$ are $\mu(dx)$-*a.s.* proportional *i.e.* $h(x)f(x) = cg(x)\ \mu(dx)$-*a.s.* for some (deterministic) nonnegative real constant c. Finally this leads to

$$g(x) = f(x)\frac{h(x)}{\mathbb{E}h(X)}\ \mu(dx) \text{ a.s.}$$

This condition is clearly impossible to reach, the simplest argument being that if it were, this would mean that $\mathbb{E}h(X)$ is known since it is involved in the formula... and would then be of no use. *A contrario* this may suggest a direction to design the (distribution) of Y.

• The intuition that must guide the user when calling upon some importance sampling method is to replace a r.v. X by another r.v. which is closer to their common mean. When pricing derivatives, the r.v. is a payoff like $(X_T - K)_+$ which is very often equal to 0 as soon as $x_0 \ll K$ *i.e.* the option is deep-out-of-the money at the origin of time. So the idea is to change the dynamics of the risky asset so that, endowed with its new distribution, it turns to be more often larger than K.

As concerns vanilla option in simple models, one usually work on the state space $E = \mathbb{R}_+$ and importance sampling amounts to a change of variable in integrals. In a more general framework, one works on the scenarii space *i.e.* one sets (in some way) $E = \Omega$ and use Girsanov theorem.

Example (option pricing). (a) In a 1-dimensional Black-Scholes model

$$X^x_T = x\exp(\mu t + \sigma W_T) = x\exp(\mu T + \sigma\sqrt{T}Z), \qquad Z \stackrel{d}{=} \mathcal{N}(0;1).$$

Then, a standard change of variable, shows

$$\begin{aligned}
\mathbb{E}\,\varphi(X^x_T) &= \mathbb{E}\,h(Z)\\
&= \int_{\mathbb{R}} h(x\exp(\mu T + \sigma\sqrt{T}z))\exp(-z^2/2)\frac{dz}{\sqrt{2\pi}}\\
&= \int_{\mathbb{R}} h(x\exp(\mu T + \sigma\sqrt{T}(u+\theta))\exp(-\theta^2/2 - \theta u - u^2/2)\frac{du}{\sqrt{2\pi}}\\
&= \exp(-\theta^2/2)\,\mathbb{E}\left(\exp(-\theta Z)h(Z+\theta)\right)\\
&= \exp(\theta^2/2)\,\mathbb{E}\left(\exp(-\theta(Z+\theta))h(Z+\theta)\right).
\end{aligned}$$

This identity is sometimes known as the Cameron-Martin formula. Viewed through the above notations related to "abstract" importance sampling, it corresponds to switch from X to Y with

$$X \longleftarrow Z, \qquad Y \longleftarrow Z + \theta$$

in (2.3) (with the related probability densities). It is to be noticed that there is need to know a numerical value for π. We leave the computations an exercise.

At this stage the underlying idea is to choose a "good" θ. This choice highly depends on the function h as emphasized above.

• If $\varphi(x) = (x-K)_+$ *i.e.* $h(z) = (x\exp(\mu T + \sigma\sqrt{T}z) - K)_+$, with $x \ll K$ (deep-out-of-the-money), most simulations of $h(Z)$ will produce 0 as a result. So, one idea can be to re-center the simulation of X^x_T around K *i.e.* choose θ satisfying

$$\mathbb{E}\left(x\exp(\mu T + \sigma\sqrt{T}(Z+\theta)\right) = K$$

which yields $\theta := \dfrac{\log(K/x)}{\sigma\sqrt{T}} - \mu T$. This choice is not optimal but reasonable.

Exercise. Set $r = 0$, $\sigma = 0.2$, $X_0 = x = 70$, $T = 1$.One wishes to price a Call with strike price $K = 100$ (*i.e.* deep-out-of-the-money). The true Black-Scholes price is 0.248.

– Make a "crude" Monte Carlo simulation and then

– apply the above importance sampling

Several methods have been developed to approximate the optimal θ *i.e.* solution to the mimimization problem

$$\min \theta \mathbb{E} \left(\exp\left(\theta^2/2\right) \exp\left(-\theta(Z+\theta)\right) h(Z+\theta) \right)^2.$$

One is based on large deviations techniques (see [20]) seems to be strongly dependent on the regularity of the function h (or say the payoff). Another approach based on stochastic approximation techniques has been recently introduced by Arouna in [1]. Although quite promising, one faces a problem: the resulting procedure does not fit (at all) in the regular theory of recursive stochastic algorithms. Practical implementation then often lead to exploding behaviours. To overcome this problem, some further developments are needed, like those developed in [1] which are essentially based on the so-called projection "à la Chen".

(b) When dealing with path-dependent options, one usually relies on the Girsanov theorem to modify in an appropriate way the drift of the risky asset dynamics. Of course all this can be implemented for multi-dimensional models...

3 Euler scheme(s) of a Brownian diffusion

One considers a d-dimensional Brownian diffusion process $(X_t)_{t\in[0,T]}$ solution of the following S.D.E.

$$(SDE) \;\equiv\; dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{3.4}$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to \mathcal{M}(d \times q)$ are continuous functions and $(W_t)_{t\in[0,T]}$ denotes a q-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ (the filtration satisfying the usual conditions). We assume that b and σ are Lipschitz continuous in x uniformly with respect to t *i.e.*, if $|\,.\,|$ denotes any norm on $\mathbb{R}^d$ and $\|\,.\,\|$ any norm on the matrix space,

$$\forall\, t \in [0, T],\ \forall\, x,\, y \in \mathbb{R}^d, \quad |b(t, x) - b(t, y)| + \|\sigma(t, x) - \sigma(t, y)\| \le K|x - y|.$$

The starting random variable X_0 is defined on $(\Omega, \mathcal{A}, \mathbb{P})$, square integrable and independent of W. Let $\mathcal{F} := (\mathcal{F}_t)_{t\in[0,T]}$ the (augmented) filtration generated by X_0 and $\sigma(W_s,\, 0 \le s \le t)$.

Then, one shows that the above SDE has a unique strong solution X with initial value X_0 (at time 0). When $X_0 = x \in \mathbb{R}^d$ one denotes the solution of (SDE) by X^x.

Remark. By adding the component t to X *i.e.* be setting $Y_t := (t, X_t)$ one may always assume that the (SDE) is homogenous *i.e.* that the coefficients b and σ only depend on the space variable. This is often enough for applications although it induces some uselessly stringent assumption on the time variable in many theoretical results. Furthermore, when some *ellipticity* assumptions are required, this way of considering the equation no longer works since the equation $dt = 1dt + 0dW_t$ is completely degenerate.

3.1 Euler-Maruyama schemes: stepwise constant and continuous schemes.

Except for some very specific equations, it is impossible to process an exact simulation of the process X even at a fixed time T (by exact simulation, we mean writing $X_T = \chi(U)$, $U \sim U([0,1])$) (nevertheless, when $d = 1$ and $\sigma \equiv 1$, see [13]). Consequently, to approximate $\mathbb{E}(f(X_T))$ by a Monte Carlo method, one needs to approximate X by a process that can be simulated (at least at a fixed number of instants). To this end one first introduces the stepwise constant Brownian Euler scheme $\bar{X} = (\bar{X}_{\frac{kT}{n}})_{0\le k\le n}$ with step $\frac{T}{n}$ associated to the SDE.

$\triangleright$ *Stepwise constant Euler scheme.* It is defined by

$$\bar{X}_{t^n_{k+1}} = \bar{X}_{t^n_k} + b(t^n_k, \bar{X}_{t^n_k})\frac{T}{n} + \sigma(t^n_k, \bar{X}_{t^n_k})\sqrt{\frac{T}{n}}\,U_{k+1}, \qquad \bar{X}_0 = X_0, \quad k = 0, \ldots, n-1, \tag{3.5}$$

where $t^n_k = \frac{kT}{n}$, $k = 0, \ldots, n-1$ and $(U_k)_{1\le k\le n}$ denotes a sequence of i.i.d. $\mathcal{N}(0;1)$-distributed random vectors given by

$$U_k := \sqrt{\frac{n}{T}}\,(W_{t^n_k} - W_{t^n_{k-1}}), \quad k = 1, \ldots, n.$$

Moreover, set for convenience

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

The stepwise constant (sometimes called "discrete") Euler scheme is defined by

$$\widetilde{X}_t = \bar{X}_{\underline{t}}, \qquad t \in [0,T].$$

$\triangleright$ *Continuous Euler scheme.* At this stage it is natural to extend the definition of the Euler scheme at every real instant $t \in [0,T]$ by setting

$$\bar{X}_t = \bar{X}_{\underline{t}} + b(\underline{t}, \bar{X}_{\underline{t}})(t - \underline{t}) + \sigma(\underline{t}, X_{\underline{t}})(W_t - W_{\underline{t}}).$$

This continuous Euler scheme satisfies (SDE) with *frozen coefficients*, namely

$$d\bar{X}_t = b(\underline{t}, X_{\underline{t}})dt + \sigma(\underline{t}, X_{\underline{t}})dW_t, \quad \bar{X}_0 = X_0$$

i.e.

$$\bar{X}_t = X_0 + \int_0^t b(\underline{s}, X_{\underline{s}})ds + \int_0^t \sigma(\underline{s}, X_{\underline{s}})dW_s.$$

Then, it is classical background that under the above assumptions on the coefficients b and σ mentioned above, $\sup_{t\in[0,T]} |X_t - \bar{X}_t|$ goes to zero in every $L^p(\mathbb{P})$, $0 < p < \infty$. Let us be more specific on that topic by providing error rates under slightly more stringent assumptions.

How to use this continuous for practical simulation seems not obvious, at least not as obvious as the stepwise constant Euler scheme. However this turns out to be an important method to improve the rate of convergence of MC simulations *e.g.* for option pricing. Using this scheme in simulation relies on the so-called diffusion bridge method and will be detailed further on.

3.2 Strong error rate

Theorem 5 *Assume b and σ satisfies for some index $\alpha \in (0,1)$,*

$$\forall\, t \in [0,T],\ \forall\, x, y \in \mathbb{R}^d, \quad |b(s,x) - b(t,y)| + \|\sigma(s,x) - \sigma(t,y)\| \le C(|t-s|^\alpha + |x-y|) \tag{3.6}$$

and if $X_0 \in L^p$ $(p \ge 2)$.

(*a*) *Then, for every $n \geq 1$,*

$$\| \sup_{t\in[0,T]} |X_t - \bar{X}_t| \|_p \leq C_{b,\sigma,p} e^{C_{b,\sigma,p}T}(1 + \|X_0\|_p) \left(\frac{T}{n}\right)^{\frac{1}{2}\wedge\alpha}.$$

In particular if b and σ are Lipschitz in (t,x) the L^p-rate is $O(n^{-\frac{1}{2}})$.

(*b*) *Then, for every $n \geq 1$,*

$$\| \sup_{t\in[0,T]} |X_t - \widetilde{X}_t| \|_p \leq C_{b,\sigma,p} e^{C_{b,\sigma,p}T}(1 + \|X_0\|_p) \sqrt{\frac{\log n}{n}}.$$

Remarks. • Note that the second rate is universal since it holds as a sharp rate for the Brownian motion itself

$$\begin{aligned}
\| \sup_{t\in[0,T]} |W_t - W_{\underline{t}}| \|_p &= \| \max_{k=0,\ldots,n-1} \sup_{t\in[t_k^n, t_{k+1}^n)} |W_t - W_{t_k^n}| \|_p \\
&= \sqrt{\frac{T}{n}} \| \max_{k=0,\ldots,n-1} \sup_{t\in[k,k+1)} |W_t - W_k| \|_p \qquad \text{by scaling} \\
&\approx \sqrt{\frac{T}{n}} C_p \sqrt{\log n}.
\end{aligned}$$

• It follows from the above theorem that the continuous Euler scheme (and the stepwise constant one as well) converge $\mathbb{P}$-*a.s.* to X. This straightforwardly follows from the Borel-Cantelli Lemma since for large enough p

$$\sum_{n\geq 1} \mathbb{E} \sup_{t\in[0,T]} |X_t - \bar{X}_t|^p) \leq c \sum_{n\geq 1} n^{-p(\frac{1}{2}\wedge\alpha)} < +\infty$$

which implies that

$$\sum_{n\geq 1} \mathbb{E} \sup_{t\in[0,T]} |X_t - \bar{X}_t|^p) < +\infty \quad \mathbb{P}\text{-}a.s.$$

One derives likewise some *a.s.* convergence $n^{-(\frac{1}{2}\wedge\alpha-\eta)}$-rate or any η small enough.

Beyond these rates, it is often useful to have at hand the following bounds for solutions of (SDE) and its Euler schemes.

Proposition 6 *If*

$$\forall\, t\in[0,T],\ \forall\, x\in\mathbb{R}^d, \quad |b(t,x)| + \|\sigma(t,x)\| \leq C(1+|x|)$$

then, for every $p\in(0,+\infty)$, there exists a real constant $C_{p,b,\sigma}\in(0,\infty)$ such that, for every $n\geq 1$,

$$\mathbb{E} \sup_{t\in[0,T]} |X_t|^p + \mathbb{E} \sup_{t\in[0,T]} |\bar{X}_t|^p \leq C_{p,b,\sigma}(1 + \mathbb{E}|X_0|^p) e^{C_{p,b,\sigma}T}.$$

We provide below a partial proof of these results, in the 1-dimensional homogenous case, for the continuous Euler scheme with $p = 2$ without optimization of the behaviour of the constants (a complete proof can be found *e.g.* in [14]).

Lemma 2 *(Gronwall Lemma) Let $f : \mathbb{R}_+ \to \mathbb{R}_+$, a Borel non-negative locally bounded function and $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ a non-decreasing function satisfying*

$$(G) \equiv f(t) \le \alpha \int_0^t f(s)\, ds + \psi(t)$$

for some $\alpha > 0$. Then

$$\forall\, t \ge 0, \qquad \sup_{0\le s\le t} f(s) \le e^{\alpha t}\psi(t).$$

Proof. It is clear that the non-decreasing (finite) function $\varphi(t) := \sup_{0\le s\le t} f(s)$ satisfies (G) instead of f. Now the function $e^{-\alpha t}\int_0^t \varphi(s)\, ds$ has right derivative at every $t \ge 0$ and that

$$\begin{aligned}\left(e^{-\alpha t}\int_0^t \varphi(s)ds\right)'_r &= e^{-\alpha t}(\varphi(t+) - \alpha\int_0^t \varphi(s)\, ds)\\ &\le e^{-\alpha t}\psi(t+)\end{aligned}$$

Then, it follows from the fundamental theorem of calculus that

$$e^{-\alpha t}\int_0^t \varphi(s)ds - \int_0^t e^{-\alpha s}\psi(s+) \quad \text{is non-increasing}$$

so that, applying that between 0 and t yields

$$\int_0^t \varphi(s) \le e^{\alpha t}\int_0^t e^{-\alpha s}\psi(s+)ds$$

Plugging this in the above inequality implies

$$\begin{aligned}\varphi(t) &\le \alpha e^{\alpha t}\int_0^t e^{-\alpha s}\psi(s+)ds + \psi(t)\\ &= \alpha e^{\alpha t}\int_0^t e^{-\alpha s}\psi(s)ds + \psi(t)\\ &\le e^{\alpha t}\psi(t)\end{aligned}$$

where we used successively that a monotone function is *ds-a.s.* continuous and that ψ is non-decreasing. $\diamond$

Now we are in position to prove the theorem. For the sake of simplicity, we will assume that $d = 1$, $p = 2$ and that the diffusion is homogenous with Lipschitz coefficients. Furthermore, we will not try dealing with the constants in an optimal way.

Doob's Inequality. (see *e.g.* [33]) Let $M = (M_t)_{t\ge 0}$ be a continuous time martingale. Then, for every $T > 0$,

$$\mathbb{E}\left(\sup_{t\in[0,T]} M_t^2\right) \le 4\,\mathbb{E}\, M_T^2 = 4\,\mathbb{E}\, <M>_T .$$

Proof of Theorem 5 (partial). Step 1. Let $\tau_L := \in \{t : |X_t - X_0| \ge L\}$, $L \in \mathbb{N}\setminus\{0\}$. It is a positive $\mathcal{F}$-stopping times and $|X_t^{\tau_L}| \le L + |X_0|$ for every $t \in [0,\infty)$. Then, using the local feature of standard and stochastic integral leads to

$$X_t^{\tau_L} = X_0 + \int_0^{t\wedge\tau_L} b(X_s^{\tau_L})ds + \int_0^{\tau_L}\sigma(X_s^{\tau_L})dW_s.$$

The continuous local martingale $M^{(n)}_t := \int_0^{t\wedge\tau_L} \sigma(X^{\tau_L}_s)X^{\tau_L}dW_s$ is a true square integrable martingale since

$$<M^{(L)}>_t = \int_0^{t\wedge\tau_L} \sigma^2(X^{\tau_L}_s)ds \le C(1+L^2+X_0^2)t \in L^1(\mathbb{P})$$

where we used that $|\sigma(x)| \le C_\sigma(1+|x|)$. Consequently, using that b also has at most linear growth, that $t\wedge\tau_L \le t$, one derives that

$$\sup_{s\in[0,t]}(X^{\tau_L}_s)^2 \le 3\left(X_0^2 + \left(\int_0^{t\wedge\tau_n} C_b(1+|X^{\tau_L}_s|)ds\right)^2 + \sup_{s\in[0,t]}|M^{(L)}_s|^2\right).$$

Consequently, using Schwarz inequality to make the square "shift down" inside the "time" integral and Doob Inequality for the stochastic integral yield

$$\mathbb{E}(\sup_{s\in[0,t]}(X^{\tau_L}_s)^2) \le C_{b,\sigma,T}\left(\mathbb{E}X_0^2 + \int_0^t(1+\mathbb{E}(\sup_{u\in[0,s]}(X^{\tau_L}_u)^2))ds + \mathbb{E}\int_0^{\tau_L\wedge t}(C_\sigma(1+|X^{\tau_L}_s|))^2ds\right).$$

This can be rewritten (using that $\tau_L\wedge t\le t$),

$$\mathbb{E}(\sup_{s\in[0,t]}(X^{\tau_L}_s)^2) \le C_{b,\sigma,T}\left(1+\mathbb{E}X_0^2 + \int_0^t \mathbb{E}(\sup_{u\in[0,s]}(X^{\tau_L}_u)^2)ds\right).$$

Gronwall Lemma implies that

$$\mathbb{E}(\sup_{s\in[0,t]}(X^{\tau_L}_s)^2) \le C_{b,\sigma,T}(1+\mathbb{E}X_0^2)e^{C_{b,\sigma,T}t}.$$

This holds or every $L\ge 1$ so that Fatou's lemma implies

$$\mathbb{E}(\sup_{s\in[0,T]}X_s^2) \le C_{b,\sigma,T}(1+\mathbb{E}X_0^2)e^{C_{b,\sigma,T}T} = C'_{b,\sigma,T}(1+\mathbb{E}X_0^2).$$

The same approach works for the Euler scheme (introduce$\bar\tau_L$ for $\bar X$ and note that $\sup_{u\in[0,s]}|\bar X_{\underline{u}}| \le \sup_{u\in[0,s]}|\bar X_u|$. This yields

$$\sup_{n\ge1}\mathbb{E}(\sup_{s\in[0,T]}(\bar X_s)^2) \le C_{b,\sigma,T}(1+\mathbb{E}X_0^2)e^{C_{b,\sigma,T}T}.$$

Step 2: Combining the equations satisfied by X and its (continuous)Euler scheme yields

$$X_t - \bar X_t \;=\; \int_0^t(b(X_s)-b(\bar X_{\underline{s}}))ds + \int_0^t(\sigma(X_s)-\sigma(\bar X_{\underline{s}}))dW_s$$

Consequently, using that b and $\sigma$are Lipschitz and Doob Inequality lead to

$$\begin{aligned}
\mathbb{E}\sup_{s\in[0,t]}|X_s-\bar X_s|^2 &\le 2\mathbb{E}\left(\int_0^t[b]_{Lip}|X_s-\bar X_{\underline{s}}|ds\right)^2 + 2\mathbb{E}\sup_{s\in[0,t]}\left(\int_0^s((\sigma(X_u)-\sigma(\bar X_{\underline{u}}))dW_u\right)^2\\
&\le 2\mathbb{E}\left(\int_0^t[b]_{Lip}|X_s-\bar X_{\underline{s}}|ds\right)^2 + 8\mathbb{E}\int_0^t((\sigma(X_u)-\sigma(\bar X_{\underline{u}}))^2du\\
&\le 2\mathbb{E}\left(\int_0^t[b]_{Lip}|X_s-\bar X_{\underline{s}}|ds\right)^2 + 8[\sigma]^2_{Lip}\int_0^t\mathbb{E}|X_u-\bar X_{\underline{u}}|^2du\\
&\le C_{b,\sigma,T}\int_0^t|X_s-\bar X_{\underline{s}}|^2ds\\
&\le C_{b,\sigma,T}\int_0^t\mathbb{E}\sup_{u\in[0,s]}|X_u-\bar X_u|^2ds + C_{b,\sigma,T}\int_0^t\mathbb{E}|\bar X_s-\bar X_{\underline{s}}|^2ds.
\end{aligned}$$

Consequently, it follows from Gronwall Lemma (at $t = T$) that

$$\mathbb{E} \sup_{s\in[0,T]} |X_s - \bar{X}_s|^2 \leq C_{b,\sigma,T} \int_0^T \mathbb{E}|\bar{X}_s - \bar{X}_{\underline{s}}|^2 ds\, e^{C_{b,\sigma,T}T}.$$

Now

$$\bar{X}_s - \bar{X}_{\underline{s}} = b(\bar{X}_{\underline{s}})(s - \underline{s}) + \sigma(\bar{X}_{\underline{s}})(W_s - W_{\underline{s}}) \tag{3.7}$$

so that, using Step 1, and the fact that $W_s - W_{\underline{s}}$ and $\bar{X}_{\underline{s}}$ are independent

$$\begin{aligned}
\mathbb{E}|\bar{X}_s - \bar{X}_{\underline{s}}|^2 &\leq C_{b,\sigma}(1 + \mathbb{E} \sup_{t\in[0,T]} |\bar{X}_t|^2) \left((T/n)^2 + \mathbb{E} \sup_{t_k^n \leq u \leq t_k^n + T/n} |W_u - W_{t_k^n}|^2 \right) \\
&= C_{b,\sigma}(1 + \mathbb{E} \sup_{t\in[0,T]} |\bar{X}_t|^2) \left((T/n)^2 + \mathbb{E}(\sup_{0\leq u \leq T/n} W_u^2) \right) \\
&= C_{b,\sigma}(1 + \mathbb{E} \sup_{t\in[0,T]} |\bar{X}_t|^2)((T/n)^2 + T/n) \\
&= C_{b,\sigma,T} T/n.
\end{aligned}$$

Step 3: Item (b) of the theorem follows from the following bound

$$\mathbb{E} \sup_{t\in[0,T]} |\bar{X}_t - \bar{X}_{\underline{t}}|^2 \leq C \frac{\log n}{n}.$$

It follows from (3.7) that

$$\sup_{t\in[0,T]} |\bar{X}_t - \bar{X}_{\underline{t}}|^2 \leq C_{b,\sigma,T}(1 + \sup_{t\in[0,T]} |\bar{X}_t|^2) \left((T/n)^2 + \sup_{t\in[0,T]} |W_t - W_{\underline{t}}|^2 \right)$$

so that, using Schwarz Inequality,

$$\| \sup_{t\in[0,T]} |\bar{X}_t - \bar{X}_{\underline{t}}| \|_2 \leq C_{b,\sigma,T}(T/n + \| \sup_{t\in[0,T]} |W_t - W_{\underline{t}}| \|_4).$$

Now, as already mentioned in the above remark that follows Theorem 5,

$$\| \sup_{t\in[0,T]} |W_t - W_{\underline{t}}| \|_4 \leq C \sqrt{\frac{T}{n}} \sqrt{\log n}$$

which completes the proof. $\diamond$

Remarks. • The proof in the general L^p framework follows exactly the same lines, except that one replaces Doob' Inequality for continuous (local) martingale $(M_t)_{t\geq 0}$ by the so-called Burkholder-Davis-Gundy Inequality (see *e.g.* [46]) which holds for every exponent $p > 0$ (in the continuous setting)

$$\| \sup_{s\in[0,t]} |M_s| \|_p \leq c_p \| <M>_t \|_{\frac{p}{2}}$$

• In some so-called mean-reverting situations one may even get boundedness over $t \in (0, \infty)$.

3.3 Path-dependent options II: Asian, Lookback and barrier options, first approach

Let

$$\mathbb{D}([0,T],\mathbb{R}^d) := \Big\{\xi : [0,T] \to \mathbb{R}^d,\ \text{càdlàg}\Big\}.$$

(cà dlàg is the French acronym for "right continuous with let limits"). The above result shows that if $F : \mathbb{D}([0,T]) \to \mathbb{R}$ is a Lipschitz functional for the sup norm *i.e.* satisfies

$$|F(\xi) - F(\xi')| \le C_F \sup_{t\in[0,T]} |\xi(t) - \xi'(t)|$$

then

$$|\mathbb{E}(F((X_t)_{t\in[0,T]}) - \mathbb{E}(F(\bar{X}_t)_{t\in[0,T]})| \le Cn^{-\frac{1}{2}}$$

and

$$|\mathbb{E}(F((X_t)_{t\in[0,T]}) - \mathbb{E}(F(\bar{X}_{\underline{t}})_{t\in[0,T]})| \le Cn^{-\frac{1}{2}}\sqrt{\log n}.$$

(keep in mind that $\widetilde{X}_t = \bar{X}_{\underline{t}}$).

Typical example in option pricing. Assume $X = (X_t)_{t\in[0,T}$ denotes the dynamics of a single risky asset.

– The Lookback and partial lookback options:

$$h_T := \left(X_T - \lambda \min_{t\in[0,T]} X_t\right)_+$$

where $\lambda = 1$ in the regular Lookback case and $\lambda > 1$ in the so-called "partial lookback" case.

– Vanilla Option on extrema (like Calls and Puts)

$$h_T = \varphi(\sup_{t\in[0,T]} X_t).$$

– Asian options of the form

$$h_T = \varphi\left(\frac{1}{T-T_0}\int_{T_0}^T X_s ds\right)$$

where φ is Lipschitz on $\mathbb{R}$. In fact such payoffs are continuous with respect to the pathwise L^2-norm *i.e.* $\|f\|_{L^1_T} := \sqrt{\int_0^T f^2(s)ds}$.

3.4 Milshtein scheme

In this section we deal with homogenous diffusion for notational convenience. However the extension to general non homogenous diffusions is straightforward (in particular it adds no further terms to the discretization scheme). The Milshtein scheme has been designed to produce a $O(1/n)$-error (in L^p) like standard schemes in a deterministic framework. This is a higher order scheme. In 1-dimension, its expression is simple and it can easily be implemented, provided b and σ have enough regularity. In higher dimension some theoretical and simulation problems make its use more questionable, especially when compared to the results about the weak error in the Euler scheme described below. The starting idea is the following. For small t, one has for the (homogenous) diffusion starting at x at time 0

$$X_t^x = x + \int_0^t b(X_s^x)ds + \int_0^t \sigma(X_s^x)dW_s$$

One has in mind that $\mathbb{E}W_t^2 = t$ *i.e.* $\|W_t\|_2 = \sqrt{t}$ so that in a scheme a Brownian term is somewhat equivalent to the square root of dt. As $t \to 0$, $\int_0^t b(X_s^x)ds$ is $O(t)$. Then by Itô's Lemma

$$\sigma(X_s^x) = \sigma(x) + \int_0^s (\sigma'(X_u^x)b(X_u^x) + \frac{1}{2}\sigma''(X_u^x)\sigma^2(X_u^x))du + \int_0^s \sigma(X_u^x)\sigma'(X_s^x)dW_u$$

so that

$$\begin{aligned}\int_0^t \sigma(X_s^x)dW_s &= \sigma(x)W_t + \int_0^t\int_0^s \sigma(X_u^x)\sigma'(X_s^x)dW_u dW_s + \text{“}o(t)\text{”}\\ &= \sigma(x)W_t + \sigma\sigma'(x)\int_0^t\int_0^s W_s dW_s + \text{“}o(t)\text{”}\\ &= \sigma(x)W_t + \frac{1}{2}\sigma\sigma'(x)(W_t^2 - t) + \text{“}o(t)\text{”}.\end{aligned}$$

Using the Markov property of the diffusion, one can reproduce this on each time step $[t_k^n, t_{k+1}^n]$ which leads to

$$\widetilde{X}^{mil}_{t^n_{k+1}} = \widetilde{X}^{mil}_{t^n_k} + \left(b(\widetilde{X}^{mil}_{t^n_k}) - \frac{1}{2}\sigma\sigma'(\widetilde{X}^{mil}_{t^n_k})\right)\frac{T}{n} + \sigma(\widetilde{X}^{mil}_{t^n_k})\sqrt{\frac{T}{n}}\,U_{k+1} + \frac{1}{2}\sigma\sigma'(\widetilde{X}^{mil}_{t^n_k})\frac{T}{n}U^2_{k+1},\ \widetilde{X}^{mil}_0 = X_0,$$

where $U_k = \sqrt{\frac{n}{T}}(W_{t^n_k} - W_{t^n_{k-1}})$, $k = 0, \ldots, n-1$. The following theorem gives the rate of strong pathwise convergence of the Milshtein scheme.

Theorem 6 *Assume b and σ are $\mathcal{C}^2$ on $\mathbb{R}$ with bounded derivatives. Then, for every $p \in (0, \infty)$ such that $X_0 \in L^p(\mathbb{P})$, one has*

$$\max_{0\le k\le n} \|\widetilde{X}^{mil}_{t^n_k} - X_{t^n_k}\|_p = O\left(\frac{T}{n}\right)$$

In higher dimension, *i.e.* when the Brownian motion W is q-dimensional, the same reasoning leads to the following scheme

$$\widetilde{X}^{mil}_{t^n_{k+1}} = \widetilde{X}^{mil}_{t^n_k} + b(\widetilde{X}^{mil}_{t^n_k})\frac{T}{n} + \sigma(\widetilde{X}^{mil}_{t^n_k})\Delta W_{t^n_{k+1}} + \sum_{1\le i,j\le q}\int_{t^n_k}^{t^n_{k+1}} (W^i_s - W^i_{t^n_k})dW^j_s\, \partial\sigma_{.i}\sigma_{.j}(\widetilde{X}^{mil}_{t^n_k}),\ \widetilde{X}_0 = X_0,$$

$k = 0, \ldots, n-1$ where

$$\partial\sigma_{.i}\,\sigma_{.j}(x) = \sum_{\ell=1}^{d}\frac{\partial\sigma_{.i}}{\partial x_\ell}(x)\sigma_{\ell j}(x).$$

If one has a look at this formula when $d = 1$ and $q = 2$, that simulating the Milshtein scheme in a general setting amounts to be able to simulate

$$\left(W^1_t, W^2_t, \int_0^t W^1_s dW^2_s\right).$$

(at time $t = t^n_1$). No convincing (*i.e.* efficient) method to achieve that is known so far.

One can of course notice that when the “rectangular” terms commute *i.e.*

$$\forall\, i \neq j, \qquad \frac{\partial\sigma_{.i}}{\partial x_\ell}\sigma_{\ell j} = \frac{\partial\sigma_{.j}}{\partial x_\ell}\sigma_{\ell i}$$

then the Milshtein scheme reduces to

$$\widetilde{X}_{t^n_{k+1}} = \widetilde{X}_{t^n_k} + \left(b(\widetilde{X}_{t^n_k}) - \frac{1}{2}\sum_{\ell=1}^{d} \partial\sigma_{.i}\sigma'_{.i}\widetilde{X}_{t^n_k})\right)\frac{T}{n} + \sigma(\widetilde{X}_{t^n_k})\Delta W_{t^n_{k+1}} \tag{3.8}$$
$$+\frac{1}{2}\sum_{1\le i,j\le q} \Delta W^i_{t^n_{k+1}}\Delta W^j_{t^n_{k+1}}\,\partial\sigma_{.i}\sigma_{.i}(\widetilde{X}_{t^n_k}),\ \widetilde{X}_0 = X_0,$$

since, if $i \neq j$,

$$\int_{t^n_k}^{t^n_{k+1}} (W^i_s - W^i_{t^n_k})dW^j_s + \int_{t^n_k}^{t^n_{k+1}} (W^j_s - W^j_{t^n_k})dW^i_s = \Delta W^i_{t^n_{k+1}}\Delta W^j_{t^n_{k+1}}$$

and

$$\int_{t^n_k}^{t^n_{k+1}} (W^i_s - W^i_{t^n_k})dW^i_s = \frac{1}{2}\left((\Delta W^i_{t^n_{k+1}})^2 - \frac{T}{n}\right).$$

The scheme (3.8) can easily be simulated since it only involves some Brownian increments.

3.5 Weak error for the Euler scheme

Usually, one introduces a discretization scheme $\bar{X} = (\bar{X}_t)_{t\in[0,T]}$ of a diffusion process $X = (X_t)_{t\in[0,T]}$ in order to compute by a Monte Carlo simulation an approximation $\mathbb{E}\,F(\bar{X})$ of $\mathbb{E}\,F(X)$. This suggests that the rates of strong convergence established above may turn to be inappropriate.

The special case of the approximation of $\mathbb{E}(f(X_T))$ by its counterpart with the Euler scheme has been extensively investigated in the literature (after being initiated by Talay-Tubaro in[48] and Bally-Talay in [3], etc), leading to an expansion of the time discretization error at an arbitrary accuracy. Then the same kind of question has been investigated with some functionals F of the path of X with some applications to path-dependent option pricing. These results show that the resulting *weak rate* is the same as the strong rate obtained with the Milshtein scheme.

As a second step, Romberg extrapolation methods provide an approach to take optimally advantage of this weak rate, including in their higher order form.

3.5.1 Main results for $\mathbb{E}(f(X_T))$: Talay-Tubaro Theorem, Bally-Talay Theorems

We adopt the notations of the former paragraph. For notational convenience we assume the diffusion is homogenous.

Theorem 7 *Assume b and σ are 4 times continuously differentiable with bounded existing partial derivatives (this implies that b and σ are Lipschitz with at most linear growth). Assume $f : \mathbb{R}^d \to \mathbb{R}$ is 4 times differentiable with polynomial growth (as well as its existing derivatives). Then*

$$\mathbb{E}\,f(X^x_T) - \mathbb{E}\,f(\bar{X}^x_T) = O\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty. \tag{3.9}$$

Sketch of proof. Assume $d = 1$ for notational convenience. We also assume $T = 1$, $b \equiv 0$ and f has bounded existing derivatives, for simplicity. The diffusion $(X^x_t)_{t\in[0,T],x\in\mathbb{R}^d}$ is a Markov process with transition semi-group (P_t) given by

$$P_t(g)(x) := \mathbb{E}\,g(X^x_t).$$

On the other hand, the Euler scheme $(\bar{X}^x_{t^n_k})_{0\le k\le n}$ is a discrete time homogenous Markov (indexed by k) chain with transition

$$\bar{P}g(x) = \mathbb{E}g(x + b(x)/n + \sigma(x)Z/\sqrt{n}), \qquad Z \stackrel{d}{=} \mathcal{N}(0;1).$$

Now

$$\mathbb{E}f(X_T^x) = P_T f(x) = P_{T/n}^n(f)(x)$$

and

$$\mathbb{E}f(\bar{X}_T^x) = \bar{P}^n(f)(x)$$

so that

$$\begin{aligned}
\mathbb{E}f(X_T^x) - \mathbb{E}f(\bar{X}_T^x) &= \sum_{k=1}^{n} P_{k/n}(\bar{P}^{n-k}f)(x) - (P_{(k-1)/n}(\bar{P}^{n-(k-1)}f)(x) \\
&= \sum_{k=1}^{n} P_{(k-1)/n}((\bar{P} - P_{T/n})(\bar{P}^{n-(k-1)}f))(x). \qquad (3.10)
\end{aligned}$$

Now, Let $g : \mathbb{R} \to \mathbb{R}$ be a four times differentiable function g with bounded existing derivatives. First, Itô's formula yields

$$P_t(g)(x) := g(x) + \underbrace{\mathbb{E}\int_0^t (g'\sigma)(X_s^x)dW_s}_{=0} + \frac{1}{2}\mathbb{E}\int_0^t g''(X_s^x)\sigma^2(X_s^x)ds$$

where we used that $g'\sigma$ has linear growth which ensures that the stochastic integral is a true martingale. A Taylor expansion then yields for the transition of the Euler scheme

$$\begin{aligned}
\bar{P}(g)(x) &= g(x) + \frac{1}{2n}(g"\sigma^2)(x) + \frac{\sigma^4(x)}{4!}\mathbb{E}(g^{(4)}(\xi)(Z/\sqrt{n})^4) \\
&= g(x) + \frac{1}{2n}(g"\sigma^2)(x) + \frac{\sigma^4(x)}{4!n^2}\|g^{(4)}\|_\infty \varepsilon_n(g)
\end{aligned}$$

with $|\varepsilon_n(g)| \le \mathbb{E}|Z^4|$, where we used that $\mathbb{E}Z = \mathbb{E}Z^3 = 0$. Consequently

$$P_{1/n}(g)(x) - \bar{P}(g)(x) = \frac{1}{2}\int_0^{\frac{1}{n}} \mathbb{E}((g"\sigma^2)(X_s^x) - (g"\sigma^2)(x))ds + \frac{\sigma^4(x)}{4!n^2}\|g^{(4)}\|_\infty \varepsilon_n(g).$$

Applying again Itô's formula to the $\mathcal{C}^2$ function $\gamma := g''\sigma^2$, yields

$$\mathbb{E}((g''\sigma^2)(X_s^x) - (g''\sigma^2)(x)) = \frac{1}{2}\varepsilon\int_0^s \gamma''(X_u^x)\sigma^2(X_u^x)du.$$

Now elementary computations show that

$$|\gamma''(y)| \le C\max(\|g^{(4)}\|_\infty, \|g^{(3)}\sigma'\|_\infty, \|g^{(2)}\sigma^{(2)}\|_\infty, \|g^{(2)}(\sigma')^2\|_\infty)(1 + |\sigma(x)|)$$

where C does not depend on g and σ. Since we know that $\sup_{t\in[0,T]}\mathbb{E}|X_t^x|^p \le C_p(1+|x|^p)e^{C_{p,\sigma}T}$, one derives (with $p = 3$) that

$$|P_{1/n}(g)(x) - \bar{P}(g)(x)| \le C_\sigma \max(\|g^{(4)}\|_\infty, \|g^{(3)}\|_\infty, \|g^{(2)}\|_\infty)\frac{1}{n^2}.$$

In order to plug this estimate in (3.10), we need now to control the first four derivatives of $\bar{P}^k f$. Let us consider again the generic function g and its fourth bounded derivatives.

$$(\bar{P}g)'(x) = \mathbb{E}(g'(x + \sigma(x)Z/\sqrt{n})(1 + \sigma'(x)Z/\sqrt{n})).$$

so that

$$
\begin{aligned}
|(\bar{P}g)'(x)| &\leq \|g'\|_\infty \|1+\sigma'(x)Z/\sqrt{n}\|_1 \\
&\leq \|g'\|_\infty \|1+\sigma'(x)Z/\sqrt{n}\|_2 \\
&= \|g'\|_\infty \sqrt{1+(\sigma')^2(x)/n} \\
&\leq \|g'\|_\infty (1+(\sigma')^2(x)/(2n)).
\end{aligned}
$$

Hence

$$|(\bar{P}^\ell f)'(x)| \leq \|g'\|_\infty (1+(\sigma')^2(x)/(2n))^\ell \leq \|g'\|_\infty e^{\|\sigma'\|_\infty^2}.$$

Then

$$(\bar{P}g)''(x) = (\bar{P})'(g')(x) + \mathbb{E}(g'(x+\sigma(x)Z/\sqrt{n})(\sigma\text{"}(x)Z/\sqrt{n}))$$

and

$$|\mathbb{E}(g'(x+\sigma(x)Z/\sqrt{n})(\sigma''(x)Z/\sqrt{n}))| \leq \|g\text{"}\|_\infty \|\sigma\sigma''\|_\infty \mathbb{E}(Z^2)/n$$

so that

$$|(\bar{P}g)''(x)| \leq \|g''\|_\infty (1+(\|\sigma\sigma''\|_\infty + \|\sigma')^2\|_\infty)/(2n)).$$

which implies the boundedness of $|(\bar{P}^\ell)'' f(x)|$. The same reasoning finally yields the bounded ness of $|(\bar{P}^k)(i)(g)|$, $i=1,2,3,4$, $k=0,\ldots,n$.

Plugging these estimates in each term of (3.10), finally yields

$$|\mathbb{E}f(X_T^x) - \mathbb{E}f(\bar{X}_T^x)| \leq C_{\sigma,f}\sum_{k=1}^n \frac{1}{n^2} \leq \frac{C_{\sigma,f}}{n}$$

which completes the proof. $\diamondsuit$

If one assumes more regularity on the coefficients or some uniform ellipticity on the diffusion coefficient σ it is possible to obtain an expansion of the error at any order.

Theorem 8 *(see [48])(a) Assume b and σ are infinitely differentiable with bounded partial derivatives. Assume $f:\mathbb{R}^d \to \mathbb{R}$ is infinitely differentiable with partial derivative having polynomial growth. Then, for every $R \geq 1$*

$$(\mathcal{E}_{R+1}) \;\equiv\; \mathbb{E}\, f(\bar{X}_T) - \mathbb{E}\, f(X_T) = \sum_{k=1}^R \frac{c_k}{n^k} + O(n^{-(R+1)}) \qquad \text{as } n\to\infty, \tag{3.11}$$

where the real coefficients c_k depend on f, T, b and σ.

(b)(see [3]) If b and σ are bounded, infinitely differentiable with bounded partial derivatives and if σ is uniformly elliptic i.e.

$$\forall\, x \in \mathbb{R}^d, \qquad \sigma\sigma^*(x) \geq \varepsilon_0 I_d \text{ for some } \varepsilon_0 > 0$$

then the conclusion of (a) holds true for any bounded Borel function.

One method of proof for (a) is to rely on the PDE method *i.e.* considering the solution of the parabolic equation

$$\left(\frac{\partial}{\partial t} + L\right)(u)(t,x) = 0, \quad u(T,.) = f$$

where $Lg = g'(x)b(x) + \frac{1}{2}g"(x)\sigma^2(x)$ denotes the infinitesimal generator of the diffusion. It follows from the Feynmann-Kac formula that (under some appropriate regularity assumptions)

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

so that

$$\begin{aligned}\mathbb{E}f(X_T) - \mathbb{E}\,f(\bar{X}_T^x) &= \mathbb{E}(u(0, x) - u(T, \bar{X}_T^x)) \\ &= \sum_{k=1}^{n} \mathbb{E}(u(t_k^n, \bar{X}_{t_k^n}^x) - u(t_{k-1}^n, \bar{X}_{t_{k-1}^n})).\end{aligned}$$

The proof consists in applying Itô's formula to show that

$$\mathbb{E}(u(t_k^n, \bar{X}_{t_k^n}^x) - u(t_{k-1}^n, \bar{X}_{t_{k-1}^n})) = \frac{\mathbb{E}\phi(t_k^n, X_{t_k^n}^x)}{n^2} + o(n^{-2}).$$

for some continuous function ϕ.

Remark. The last important information about weak error is that the weak error induced by the Milshtein scheme has exactly the same order as that of the Euler scheme *i.e.* $O(1/n)$. So the Milshtein scheme seems of little interest as long as one wishes to compute $\mathbb{E}(f(X_T))$ with a reasonable framework like the ones described in the theorem, since, even when it can be implemented without restriction, its complexity is higher than that of the standard Euler scheme. Furthermore, the next paragraph about Romberg extrapolation will show that its is possible to take advantage of the higher order time discretization error expansion which become dramatically faster than Milshtein scheme.

3.6 Standard Romberg extrapolation and multistep Romberg extrapolation

3.6.1 Application to Romberg extrapolation (with consistent increments)

Assume that $(\mathcal{E}_2^V)$ holds. Let $f \in V$ where V denotes a vector space of continuous functions with linear growth. The case of non continuous functions is investigated in the next section. For notational convenience we set $W^{(1)} = W$ and $X^{(1)} := X$. A regular Monte Carlo simulation based on M independent copies $(\bar{X}_T^{(1)})^m$, $m = 1, \ldots, M$, of the Euler scheme $\bar{X}_T^{(1)}$ with step T/n induces the following global (squared) quadratic error

$$\begin{aligned}\|\mathbb{E}(f(X_T)) - \frac{1}{M}\sum_{m=1}^{M} f((\bar{X}_T^{(1)})^m)\|_2^2 &= |\mathbb{E}(f(X_T)) - \mathbb{E}(f(\bar{X}_T^{(1)}))|^2 \\ &\quad + \|\mathbb{E}(f(\bar{X}_T^{(1)})) - \frac{1}{M}\sum_{m=1}^{M} f((\bar{X}_T^{(1)})^m)\|_2^2 \\ &= \frac{c_1^2}{n^2} + \frac{\mathrm{Var}(f(\bar{X}_T^{(1)}))}{M} + O(n^{-3}). \qquad (3.12)\end{aligned}$$

This quadratic error bound (3.12) does not take fully advantage of the above expansion $(\mathcal{E}_2)$. To take advantage of the expansion, one needs to make an Romberg extrapolation. In that framework (originally introduced in [48]) one considers a second Brownian Euler scheme, this time of the solution $X^{(2)}$ of a "copy" of Equation (3.4) *a priori* written with respect to a second Brownian motion $W^{(2)}$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In fact, one may chose this Brownian motion by enlarging Ω if necessary. This second Euler scheme has *a twice smaller* step $\frac{T}{2n}$ and is denoted $\bar{X}^{(2)}$. Then, assuming $(\mathcal{E}_2^V)$to be more precise,

$$\mathbb{E}(f(X_T)) = \mathbb{E}(2f(\bar{X}_T^{(2)}) - f(\bar{X}_T^{(1)})) - \frac{1}{2}\frac{c_2}{n^2} + O(n^{-3}).$$

Then, the new global (squared) quadratic error becomes

$$\|\mathbb{E}(f(X_T)) - \frac{1}{M}\sum_{m=1}^{M} 2f((\bar{X}_T^{(2)})^m) - f((\bar{X}_T)^m)\|_2^2 = \frac{c_2^2}{4n^4} + \frac{\mathrm{Var}(2f(\bar{X}_T^{(2)}) - f(\bar{X}_T^{(1)}))}{M} + O(n^{-5}). \quad (3.13)$$

The structure of this quadratic error suggests the following question: is it possible to reduce the (asymptotic) time discretization error *without increasing the Monte Carlo error? To what extend is it possible to control the variance term* $\mathrm{Var}(2f(\bar{X}_T^{(2)}) - f(\bar{X}_T^{(1)}))$*?*

It is shown in [39], that if $W^{(2)} = W^{(1)}$ $(= W)$ then

$$\mathrm{Var}(2f(\bar{X}_T^{(2)}) - f(\bar{X}_T^{(1)})) \overset{n\to\infty}{\longrightarrow} \mathrm{Var}(2f(X_T) - f(X_T)) = \mathrm{Var}(f(X_T))$$

and that this choice is optimal among all possible choice of correlated Brownian motions $W^{(1)}$ and $W^{(2)}$. This result can be extended to Borel functions f when the diffusion is uniformly elliptic (and b, σ bounded, infinitely differentiable with bounded partial derivatives).

From a practical viewpoint, one first simulates an Euler scheme with step $\frac{T}{2n}$ using a white Gaussian noise $(U_k^{(2)})_{k\geq 1}$, then one simulates an Euler scheme with step $\frac{T}{n}$ using the white Gaussian white noise

$$U_k^{(1)} = \frac{U_{2k}^{(2)} + U_{2k-1}^{(2)}}{\sqrt{2}}, \; k \geq 1.$$

Note that if one adopts the "lazy" approach based on independent Gaussian noises $U^{(1)}$ and $U^{(2)}$ the asymptotic variance is

$$\mathrm{Var}(2f(X_T^{(2)}) - f(X_T^{(1)})) = 4\mathrm{Var}(f(X_T^{(2)})) + \mathrm{Var}(f(X_T^{(1)})) = 5\mathrm{Var}(f(X_T^{(1)})).$$

3.6.2 Toward a multistep Romberg extrapolation

In [39], a more general approach to multistep Romberg extrapolation with consistent Brownian increments is developed. Given the state of the technology and the needs, it seems interesting up to $R = 4$. We will sketch the case $R = 3$ which may also work for path dependent options (see below). Set

$$\alpha_1 = \frac{1}{2}, \qquad \alpha_2 = -4, \qquad \alpha_3 = \frac{9}{2}.$$

(Note that $\sum_i \alpha_i^2 = \frac{73}{2}$ which would correspond to the variance term if the extrapolation is implemented with independent Brownian motions). Then, easy computations show that

$$\mathbb{E}\left(\alpha_1 f(\bar{X}_T^{(1)}) + \alpha_2 f(\bar{X}_T^{(2)}) + \alpha_3 f(\bar{X}_T^{(3)})\right) = \frac{c_3^{(3)}}{n^3} + O(n^{-4}).$$

where $\bar{X}^{(r)}$ denotes the Euler scheme with step $\frac{T}{rn}$, $r = 1, 2, 3$, *with respect to the same Brownian motion* W. Once again this choice induces a control of the variance of the estimator. However, this choice is theoretically no longer optimal although natural.

In practice the three Gaussian white noises can be simulated following the following consistency rules: We give below the most efficient way to simulate the three whites noises $(U_k^{(r)})_{1\le k\le r}$ on one time step T/n. Let U_1, U_2, U_3, U_4 be four i.i.d. copies of $\mathcal{N}(0; I_q)$. Set

$$U_1^{(3)} = U_1, \quad U_2^{(3)} = \frac{U_2+U_3}{\sqrt{2}}, \quad U_3^{(3)} = U_4,$$

$$U_1^{(2)} = \frac{\sqrt{2}\,U_1+U_2}{\sqrt{3}}, \quad U_2^{(2)} = \frac{U_3+\sqrt{2}\,U_4}{\sqrt{3}},$$

$$U_1^{(1)} = \frac{U_1^{(2)}+U_2^{(2)}}{\sqrt{2}}.$$

A general formula for the weights $\alpha_r^{(R)}$, $r=1,\dots,R$ as well as the consistent increments tables are provided in [39].

Guided by some complexity considerations,one shows that the parameters of this multistep Romberg extrapolation should satisfy some constraints. typically if M denotes the size of the MC simulation and n the discretization parameter, they should be chosen so that

$$M \propto n^{2R}.$$

For details we refer to [39]. The practical limitation of these results about Romberg extrapolation is that the control of the variance is only asymptotic (as $n\to\infty$) whereas the method is usually implemented for small values of n. However it is efficient up to $R=4$ when for Monte Carlo simulation of sizes $M=10^6$ to $M=10^8$.

Exercise: We consider the simplest option pricing model, the (risk-neutral) Black-Scholes dynamics, but with unusually high volatility. (We are aware that this model used to price Call options does not fulfill the theoretical assumptions made above). To be precise

$$dX_t = X_t\,(rdt+\sigma dW_t),$$

with the following values for the parameters

$$X_0 = 100,\ K=100,\ r=0.15,\ \sigma=1.0,\ T=1.$$

Note that a volatility $\sigma=100\%$ per year is equivalent to a 4 year maturity with volatility 50% (or 16 years with volatility 25%). The reference Black-Scholes premium is $C_0^{BS}=42.96$.

- We consider the Euler scheme with step T/n of this equation.

  $$\bar{X}_{t_{k+1}} = \bar{X}_{t_k}\left(1+r\frac{T}{n}+\sigma\sqrt{\frac{T}{n}}\,U_{k+1}\right), \qquad \bar{X}_0 = X_0,$$

  where $t_k=\frac{kT}{n}$, $k=0,\dots,n$. We want to price a vanilla Call option *i.e.* to compute

  $$C_0 = e^{-rT}\mathbb{E}((X_T-K)_+)$$

  using a Monte Carlo simulation with M sample paths, $M=10^4$, $M=10^6$, etc.

- Test now the standard Romberg extrapolation ($R=2$) based on Euler schemes with steps T/n and $T/(2n)$, $n=2\,4\,6,8,\ 10$, respectively with

  – independent Brownian increments

  – consistent Brownian increments

  Compute an estimator of the variance of the estimator.

3.7 Weak error for path-dependent functionals: the Brownian bridge method

In this section we will consider some path-dependent (European) options *i.e.* related to some payoffs $F((X_t)_{t\in[0,T]})$ where F is a functional defined on the set $\mathbb{D}([0,T],\mathbb{R}^d)$ of right continuous left-limited functions $x:[0,T]\to\mathbb{R}$. It is clear that all the asymptotic control of the variance obtained in the former section for the estimator $\sum_{r=1}^{R}\alpha_r f(\bar X_T^{(r)})$ of $\mathbb{E}(f(X_T))$ when f is continuous can be extended to functionals $F:I\!\!D([0,T],\mathbb{R}^d)\to\mathbb{R}$ which are $\mathbb{P}_X$-*a.s.* continuous with respect to the sup-norm defined by $\|x\|_{\sup}:=\sup_{t\in[0,T]}|x(t)|$ with polynomial growth (*i.e.* $|F(x)|=O(\|x\|_{\sup}^{\ell})$ for somme natural integer ℓ as $\|x\|_{\sup}\to\infty$). This simply follows from the fact that the (continuous) Euler scheme $\bar X$ (with step T/n) defined by

$$\forall\, t\in[0,T],\qquad \bar X_t = x_0+\int_0^t b(\bar X_{\underline{s}})ds+\int_0^t\sigma(\bar X_{\underline{s}})dW_s,\ \underline{s}=\lfloor ns/T\rfloor$$

converges for the sup-norm toward X in every $L^p(\mathbb{P})$.

Furthermore, this asymptotic control of the variance holds true with any R-tuple $\alpha=(\alpha_r)_{1\le r\le R}$ of weights coefficients satisfying $\sum_{1\le r\le R}\alpha_r=1$, so these coefficients can be adapted to the structure of the weak error expansion.

On the other hand, in the recent past years, several papers provided some weak rates of convergence for some families of functionals F. These works were essentially motivated by the pricing of path-dependent (European) options, like Asian, lookback or barrier options. This corresponds to functionals

$$F(x):=\Phi\Big(\int_0^T x(s)ds\Big),\quad F(x):=\Phi\big(x(T),\sup_{t\in[0,T]}x(t),\inf_{t\in[0,T]}x(t)\big),\quad F(x)=\Phi(x(T))\mathbf{1}_{\{\tau_D(x)\le T\}}$$

where Φ is usually at least Lipschitz and $\tau_D:=\inf\{s\in[0,T],\ x(s\pm)\in{}^cD\}$ is the hitting time of ${}^cD$ by x ([1]). Let us briefly mention two well-known examples:

– In [23], it is established that if the domain D has a smooth enough boundary, $b,\,\sigma\in\mathcal{C}^3(\mathbb{R}^d)$, σ uniformly elliptic on D, then for every Borel bounded function f vanishing in a neighbourhood of ∂D,

$$\mathbb{E}(f(\widetilde X_T)\mathbf{1}_{\{\tau(\widetilde X)>T\}})-\mathbb{E}(f(X_T)\mathbf{1}_{\{\tau(X)>T\}})=O\left(\frac{1}{\sqrt n}\right)\quad\text{as}\quad n\to\infty. \tag{3.14}$$

If furthermore, b and σ are $\mathcal{C}^5$, then

$$\mathbb{E}(f(\bar X_T)\mathbf{1}_{\{\tau(\bar X)>T\}})-\mathbb{E}(f(X_T)\mathbf{1}_{\{\tau(X)>T\}})=O\left(\frac{1}{n}\right)\quad\text{as}\quad n\to\infty. \tag{3.15}$$

Note however that these assumptions are not satisfied by usual barrier options (see below).

– it is suggested in [47] (including a rigorous proof when $X=W$) that if $b,\,\sigma\in\mathcal{C}_b^4(\mathbb{R})$, σ is uniformly elliptic and $\Phi\in\mathcal{C}^{4,2}(\mathbb{R}^2)$ (with some partial derivatives with polynomial growth), then

$$\mathbb{E}(\Phi(\widetilde X_T,\min_{0\le k\le n}\widetilde X_{t_k}))-\mathbb{E}(\Phi(X_T,\min_{t\in[0,T]}X_t))=O\left(\frac{1}{\sqrt n}\right)\quad\text{as}\quad n\to\infty. \tag{3.16}$$

A similar improvement – $O(\frac1n)$ rate – as above can be expected when replacing $\widetilde X$ by the continuous Euler scheme $\bar X$.

[1] when x is stepwise constant and càdlàg, one can write "s" instead of "$s\pm$".

3.7.1 The Brownian bridge and its application to simulation

To take advantage of the above rates (even in a crude simulation), one needs to simulate the the continuous Euler scheme or at least some quantities related to this continuous scheme between to discretization points t_k^n and t_{k+1}^n.

Lemma 3 *Let $W = (W_t)_{t\geq 0}$ be a standard Brownian motion.*
(a) Let $T > 0$. Then, the standard Brownian bridge on $[0,T]$ defined by

$$Y_t^{W,T} := W_t - \frac{t}{T} W_T \tag{3.17}$$

is a $(\mathcal{F}_t^W)_{0\leq t\leq T}$-measurable centered Gaussian process independent of $(W_{T+s})_{s\geq 0}$ characterized by its covariance

$$\mathbb{E}\, Y_s Y_t = \frac{(s\wedge t)(T - s\vee t)}{T}, \qquad 0 \leq s,\, t \leq T.$$

(b) Let $0 \leq T_0 < T_1 < +\infty$. Then

$$\mathcal{L}\left((W_t)_{t\in[T_0,T_1]} \,|\, W_s \notin (T_0,T_1)\right) = \mathcal{L}\left((W_t)_{t\in[T_0,T_1]} \,|\, W_{T_0}, W_{T_1}\right)$$

is independent of $\sigma(W_s,\, s \notin (T_0,T_1))$ and

$$\mathcal{L}\left((W_t)_{t\in[T_0,T_1]} \,|\, W_{T_0} = w_0, W_{T_1} = w_1\right) \stackrel{d}{=} w_0 + \frac{t - T_0}{T_1 - T_0}(w_1 - w_0) + (Y_{t-T_0}^{W^{(T_0)},T_1-T_0})_{t\in[T_0,T_1]}.$$

where $W_t^{(T)} := W_{T+t} - W_T$, $t \geq 0$ is a standard Brownian motion (independent of $\mathcal{F}_T^W$).

Proof. (a) Elementary computations based on $\mathbb{E} W_s W_t = s\wedge t$, having in mind that independence and non correlation are equivalent in a Gaussian space.

(b) This is a consequence of item (a). First note that for every $t \in [T_0,T_1]$,

$$W_t = W_{T_0} + \frac{t - T_0}{T_1 - T_0}(W_{T_1} - W_{T_0}) + Y_{t-T_0}^{W^{(T_0)},T_1-T_0}.$$

It follows from (a) that the process $\widetilde{Y} := (Y_{t-T_0}^{W^{(T_0)},T_1-T_0})_{t\in[T_0,T_1]}$ is a Gaussian process independent of $\mathcal{F}_{T_0}^W$ and of $W_{T_1-T_0+s}^{(T_0)}$, $s \geq 0$. Since $(W,\widetilde{Y})$ is a Gaussian process, this is equivalent to the independence of $\widetilde{Y}$ with $\sigma(W_s,\, s \notin (T_0,T_1))$ since $W_{T_1+s} = W_{T_1-T_0+s}^{(T_0)} + W_{T_0}$. $\diamond$

Proposition 7 *Assume that $\sigma(t,x) \neq 0$ for every $t \in [0,T]$, $x \in \mathbb{R}$.*

The processes $(\bar{X}_t)_{t\in[t_k^n,t_{k+1}^n]}$, $k = 0,\ldots,n-1$ are conditionally independent given the σ-field $\sigma(\{\bar{X}_{t_k^n} = x_k,\, k = 0,\ldots,n\})$.

Furthermore, the conditional distribution $\mathcal{L}\left((\bar{X}_t)_{t\in[t_k^n,t_{k+1}^n]} \,|\, \bar{X}_{t^n} = x_k, \bar{X}_{t_{k+1}^n} = x_{k+1}\right)$ is given by

$$x_k + \frac{t - t_k^n}{t_{k+1}^n - t_k^n}(x_{k+1} - x_k) + \sigma(t_k^n, x_k) Y_{t-t_k^n}^{W,T/n}$$

where $(Y_s^{W,T/n})_{s\in[0,T/n]}$ is a Brownian bridge as defined by (3.17). This formula is sometimes called a diffusion bridge.

Proof. Elementary computations show that for every $t\in[t^n_k,t^n_{k+1}]$,

$$\bar X_t=\bar X_{t^n_k}+\frac{t-t^n_k}{t^n_{k+1}-t^n_k}(\bar X_{t^n_{k+1}}-\bar X_{t^n_k})+\sigma(t^n_k,\bar X_{t^n_k})Y^{W^{(t^n_k)},T/n}_{t-t^n_k}$$

(with in mind that $t^n_{k+1}-t^n_k=T/n$) Consequently the conditional independence claim will follow if the processes $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$, $k=0,\ldots,n-1$, are independent given $\sigma(\bar X_{t^n_\ell},\,\ell=0,\ldots,n)$. Now, it follows from the assumption on σ that

$$\sigma(\bar X_{t^n_\ell},\,\ell=0,\ldots,n)=\sigma(X_0,\,W_{t^n_\ell},\,\ell=1,\ldots,n).$$

So we have to establish the conditional independence of the processes $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$, $k=0,\ldots,n-1$, given $\sigma(X_0,\,W_{t^n_k},\,k=1,\ldots,n)$ or equivalently given $\sigma(W_{t^n_k},\,k=1,\ldots,n)$ since X_0 and W are independent (note all the above bridges are $\mathcal{F}^W_T$-measurable). First note that all the bridges $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$, $k=0,\ldots,n-1$ and W live in a Gaussian space.

We know from Lemma 3(a) that each bridge $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$ is independent of both $\mathcal{F}_{t^n_k}$ and $\sigma(W_{t^n_{k+1}+s}-W_{t^n_k},s\ge 0)$ hence in particular of all $\sigma(\{W_{t^n_\ell},\,\ell=1,\ldots,n\})$ (we use here a specificity of Gaussian processes). On the other hand, all bridges are independent since they are built from independent Brownian motions $(W^{(t^n_k)}_t)_{t\in[0,T/n]}$. Hence, the bridges $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$, $k=0,\ldots,n-1$ are i.i.d. and independent of $\sigma(W_{t^n_k},\,k=1,\ldots,n)$.

Now $\bar X_{t^n_k}$ is $\sigma(\{W_{t^n_\ell},\,\ell=1,\ldots,k\})$-measurable consequently $\sigma(X_{t^n_k},W_{t^n_k},\bar X_{t^n_{k+1}})\subset\sigma(\{W_{t^n_\ell},\,\le=1,\ldots,n\})$ so that $(Y^{W^{(t^n_k)},T/n}_t)_{t\in[0,T/n]}$ is independent of $(X_{t^n_k},W_{t^n_k},\bar X_{t^n_{k+1}})$. The conclusion follows. $\diamond$

Proposition 8 *The distribution of the supremum of the Brownian bridge arriving at y at time T, defined by $Y^{(0,y)}_t=\frac{t}{T}y+W_t-\frac{t}{T}W_T$ on $[0,T]$ is given by*

$$\forall\,u>y,\quad \mathbb{P}(\sup_{t\in[0,T]}Y^{(0,y)}_t\le u)=1-\exp\Big(-\frac{2}{T}u(u-y)\Big).$$

Proof. The key is to have in mind that $Y^{(0,y)}$ is the conditional distribution of W given $W_T=y$. So, we compute we can derive the result from an expression of the joint distribution of $(\sup_{t\in[0,T]}W_t,W_T)$, *e.g.* from

$$\mathbb{P}(\sup_{t\in[0,T]}W_t\ge u,\,W_T\le y).$$

It is well-known from the symmetry principle that

$$\mathbb{P}(\sup_{t\in[0,T]}W_t\ge u,\,W_T\le y)=\mathbb{P}(W_T\ge 2u-y).$$

One introduces the hitting time $\tau_u:=\inf\{s>0\,|\,W_s=u\}$. Then τ_u is *a.s.* finite, $W_{\tau_u}=u$ *a.s.* and $W_{\tau_u+t}-W_{\tau_u}$ is independent of $\mathcal{F}_{\tau_u}$. As a consequence

$$\begin{aligned}\mathbb{P}(\sup_{t\in[0,T]}W_t\ge u,\,W_T\le y)&=\mathbb{P}(\tau_u\le T,W_T-W_{\tau_u}\le y-u)\\&=\mathbb{P}(\tau_u\le T,-(W_T-W_{\tau_u})\le y-u)\\&=\mathbb{P}(\tau_u\le T,\,W_T\ge 2u-y)\qquad\text{since } u>y\\&=\mathbb{P}(W_T\ge 2u-y).\end{aligned}$$

Hence, since the involved functions are regular, one has

$$\begin{aligned}\mathbb{P}(\sup_{t\in[0,T]} W_t \geq u \,|\, W_T = y) &= \lim_{\eta\to 0} \frac{(\mathbb{P}(W_T \geq 2u-(y+\eta)) - \mathbb{P}(W_T \geq 2u-y))/\eta}{(\mathbb{P}(W_T \leq y+\eta) - \mathbb{P}(W_T \leq y))/\eta} \\ &= -\frac{\frac{\partial \mathbb{P}(W_T \geq 2u-y)}{\partial y}}{\frac{\partial \mathbb{P}(W_T\leq y)}{\partial y}}.\end{aligned}$$

Elementary computations yield the result. $\diamond$.

Simulation of the maximum of the continuous Euler scheme and application

First we wish to simulate the distribution of the supremum over $[0,T/n]$ of some bridges staring at x and arriving at y, *i.e.* of the form $x + \frac{t}{(T/n)}(y-x)Y_t^{W,T/n}$. In view of the above proposition 8, we will use the distribution function inverse method to proceed. Elementary computations show that

$$\mathcal{L}\left(\sup_{t\in[0,T/n]}\left(x+\frac{t}{(T/n)}(y-x)Y_t^{W,T/n}\right)\right) \stackrel{d}{=} G_{x,y}^{-1}(U), \quad U \stackrel{d}{=} U([0,1])$$

where

$$G_{x,y}^{-1}(u) = \frac{1}{2}\left(x+y+\sqrt{(x-y)^2 - 2T\sigma(x)\log(u)/n}\right).$$

Now, we derive from Proposition 7 that

$$\mathcal{L}(\max_{t\in[0,T]} \bar{X}_t \,|\, \{\bar{X}_{t_k} = x_k,\ k=0,\ldots,n\}) = \mathcal{L}(\max_{0\leq k\leq n-1} G_{x_k,x_{k+1}}^{-1}(U_k))$$

where $(U_k)_{0\leq k\leq n-1}$ are *i.i.d.* uniformly distributed random variables over the unit interval.

Once one can simulate $\sup_{t\in[0,T]} \bar{X}_t$ (and its minimum, see exercise below), it is easy to price by simulation the exotic options mentioned in the former section (lookback, options on maximum) but also the barrier options since one can decide whether or not the *continuous* Euler scheme strikes or not a barrier (up or down). Brownian bridge is also involved in the methods designed for pricing Asian options.

Exercise Using the symmetry $W \stackrel{d}{=} -W$, one derives a similar formula holds for the minimum using now the inverse distribution function

$$F_{x,y}^{-1}(u) = \frac{1}{2}\left(x+y-\sqrt{(x-y)^2 - 2T\sigma(x)\log(u)/n}\right).$$

3.7.2 Weak errors and Romberg extrapolation for path-dependent options: results and experiments

For both classes of functionals (with D as a half-line in 1-dimension in the first setting), the practical implementation of the continuous Euler scheme is known as the *Brownian bridge method*.

At this stage there are two ways to implement the (multistep) Romberg extrapolation with consistent Brownian increments in order to improve the performances of the original (stepwise constant or continuous) Euler schemes. Both rely on natural conjectures about the existence of a higher order expansion of the time discretization error suggested by the above rates of convergence (3.14), (3.15) and (3.16).

• *Stepwise constant Euler scheme:* As concerns the standard Euler scheme, this means the existence of a vector space V (stable by product) of admissible functionals satisfying

$$(\mathcal{E}_R^{\frac{1}{2},V}) \qquad \equiv \qquad \forall\, F\in V, \qquad \mathbb{E}(F(X)) = \mathbb{E}(F(\widetilde{X})) + \sum_{k=1}^{R-1}\frac{c_k}{n^{\frac{k}{2}}} + O(n^{-\frac{R}{2}}). \tag{3.18}$$

still for some real constant c_k depending on b, σ, F, etc For small values of R, one checks that

$$R=2: \quad \alpha_1^{(\frac12)} = -(1+\sqrt{2}), \qquad \alpha_2^{(\frac12)} = \sqrt{2}(1+\sqrt{2}).$$

$$R=3: \quad \alpha_1^{(\frac12)} = \frac{\sqrt{3}-\sqrt{2}}{2\sqrt{2}-\sqrt{3}-1}, \qquad \alpha_2^{(\frac12)} = -2\frac{\sqrt{3}-1}{2\sqrt{2}-\sqrt{3}-1}, \qquad \alpha_3^{(\frac12)} = 3\frac{\sqrt{2}-1}{2\sqrt{2}-\sqrt{3}-1}.$$

Note that these coefficients have greater absolute values than in the standard case. Thus if $R=4$, $\sum_{1\le r\le 4}(\alpha_r^{(\frac12)})^2 \approx 10\,900!$ which induces an increase of the variance term for too small values of the time discretization parameters n even when increments are consistently generated. The complexity computations of the procedure needs to be updated but *grosso modo* the optimal choice for the time discretization parameter n as a function of the MC size M is

$$M \propto n^R.$$

• *The continuous Euler scheme:* The conjecture is simply to assume that the expansion $(\mathcal{E}_R)$ now holds for a vector space V of *functionals* F (with polynomial growth with respect to the sup-norm). The increase of the complexity induced by the Brownian bridge method is difficult to quantize: it amounts to computing $\log(U_k)$ and the inverse distribution functions $F^{-1}_{x,y}$ and $G^{-1}_{x,y}$.

The second difficulty is that simulating (the extrema of) some of continuous Euler schemes using the Brownian bridge in a consistent way is not straightforward at all. However, one can reasonably expect that using independent Brownian bridges "relying" on stepwise constant Euler schemes with consistent Brownian increments will have a small impact on the global variance (although slightly increasing it).

To illustrate and compare these approaches we carried some numerical tests on partial lookback and barrier options in the Black-Scholes model presented in the previous section.

$\triangleright$ Partial lookback options: The partial lookback Call option is defined by its payoff functional

$$F(x) = e^{-rT}\left(x(T) - \lambda \min_{s\in[0,T]} x(s)\right)_+, \quad x\in\mathcal{C}([0,T],\mathbb{R}),$$

where $\lambda>0$ (if $\lambda\le 1$, the $(\,.\,)_+$ can be dropped). The premium

$$\mathrm{Call}_0^{Lkb} = e^{-rT}\mathbb{E}((X_T - \lambda \min_{t\in[0,T]} X_t)_+)$$

is given by

$$\mathrm{Call}_0^{Lkb} = X_0\mathrm{Call}^{BS}(1,\lambda,\sigma,r,T) + \lambda\frac{\sigma^2}{2r}X_0\mathrm{Put}^{BS}\left(\lambda^{\frac{2r}{\sigma^2}}, 1, \frac{2r}{\sigma}, r, T\right).$$

We took the same values for the B-S parameters as in the former section and set the coefficient λ at $\lambda=1.1$. For this set of parameters $\mathrm{Call}_0^{Lkb} = 57.475$.

As concerns the MC simulation size, we still set $M=10^6$. We compared the following three methods for every choice of n:

– A 3-step Romberg extrapolation ($R=3$) of the stepwise constant Euler scheme (for which a $O(n^{-\frac32})$-rate can be expected from the conjecture).

– A 3-step Romberg extrapolation ($R=3$) based on the continuous Euler scheme(Brownian bridge method) for which a $O(\frac{1}{n^3})$-rate can be conjectured (see [23]).

– A continuous Euler scheme (Brownian bridge method) of equivalent complexity *i.e.* with discretization parameter $6n$ for which a $O(\frac{1}{n})$-rate can be expected (see [23]).

The three procedures have the same complexity if one neglects the cost of the bridge simulation with respect to that of the diffusion coefficients (note this is very conservative in favour of "bridged schemes").

We do not reproduce the results obtained for the standard stepwise constant Euler scheme which are clearly out of the game (as already emphasized in [23]). In Figure 3.7.2, the abscissas represent the size of Euler scheme with equivalent complexity (*i.e.* $6n$, $n = 2, 4, 6, 8, 10$). Figure 3.7.2(a) (left) shows that both 3-step Romberg extrapolation methods converge significantly faster than the "bridged" Euler scheme with equivalent complexity in this high volatility framework. The standard deviations depicted in Figure 3.7.2(a) (right) show that the 3-step Romberg extrapolation of the Brownian bridge is controlled even for small values of n. This is not the case with the 3-step Romberg extrapolation method of the stepwise constant Euler scheme. Other simulations – not reproduced here – show this is already true for the standard Romberg extrapolation and the bridged Euler scheme. In any case the multistep Romberg extrapolation with $R = 3$ significantly outperforms the bridged Euler scheme.

When $M = 10^8$, one verifies (see Figure 3.7.2(b)) that the time discretization error of the 3-step Romberg extrapolation vanishes like for the partial lookback option. In fact for $n = 10$ the 3-step bridged Euler scheme yields a premium equal to 57.480 which corresponds to less than half a cent error, *i.e.* 0.05 % accuracy! This result being obtained without any control variate variable.

The Romberg extrapolation of the standard Euler scheme also provides excellent results. In fact it seems difficult to discriminate them with those obtained with the bridged schemes, which is slightly unexpected if one think about the natural conjecture about the time discretization error expansion.

As a theoretical conclusion, these results strongly support both conjectures about the existence of expansion for the weak error in the $(n^{-p/2})_{p\geq 1}$ and $(n^{-p})_{p\geq 1}$ scales respectively.

$\triangleright$ Up & out Call option: Let $0 \leq K \leq L$. The Up-and-Out Call option with strike K and barrier L is defined by its payoff functional

$$F(x) = e^{-rT}\,(x(T) - K)_+\,\mathbf{1}_{\{\max_{s\in[0,T]} x(s)\leq L\}}, \quad x\in\mathcal{C}([0,T],\mathbb{R}).$$

It is again classical background, that in a B-S model

$$\begin{aligned}\mathrm{Call}^{U\&O}(X_0,r,\sigma,T) &= \mathrm{Call}^{BS}(X_0,K,r,\sigma,T) - \mathrm{Call}^{BS}(X_0,L,r,\sigma,T) - e^{-rT}(L-K)\Phi(d^-(L))\\ &\quad - \left(\frac{L}{X_0}\right)^{1+\mu}\Big(\mathrm{Call}^{BS}(X_0,K',r,\sigma,T)-\mathrm{Call}^{BS}(X_0,L',r,\sigma,T)-e^{-rT}(L'-K')\Phi(d^-(L'))\Big)\end{aligned}$$

with

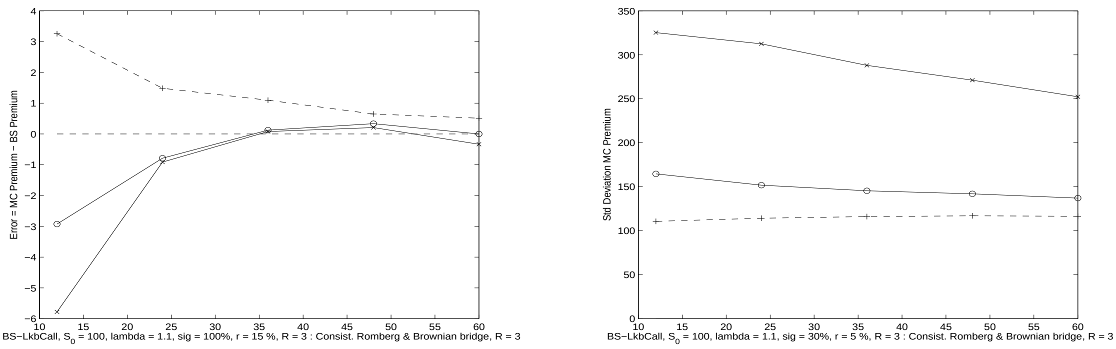
$$K' = K\left(\frac{X_0}{L}\right)^2,\; L' = L\left(\frac{X_0}{L}\right)^2,\; d^-(L) = \frac{\log(X_0/L) + (r-\frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \text{ and } \Phi(x) := \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}}\frac{d\xi}{\sqrt{2\pi}}$$

and $\mu = \dfrac{2r}{\sigma^2}$.

We took again the same values for the B-S parameters as for the vanilla call. We set the barrier value at $L = 300$. For this set of parameters $C_0^{UO} = 8.54$. We tested the same three schemes. The numerical results are depicted in Figure 3.7.2.

The conclusion (see Figure 3.7.2(a) (left)) is that, at this very high level of volatility, when $M = 10^6$ (which is a standard size given the high volatility setting) the (quasi-)consistent 3-step

(a)



(b)

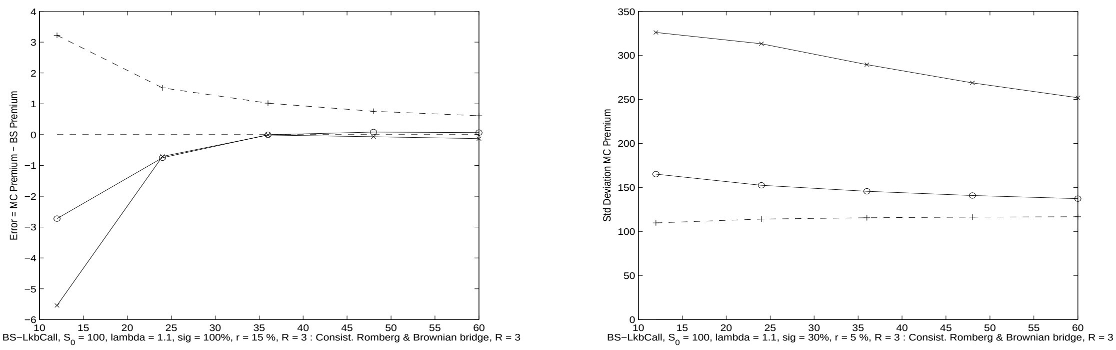
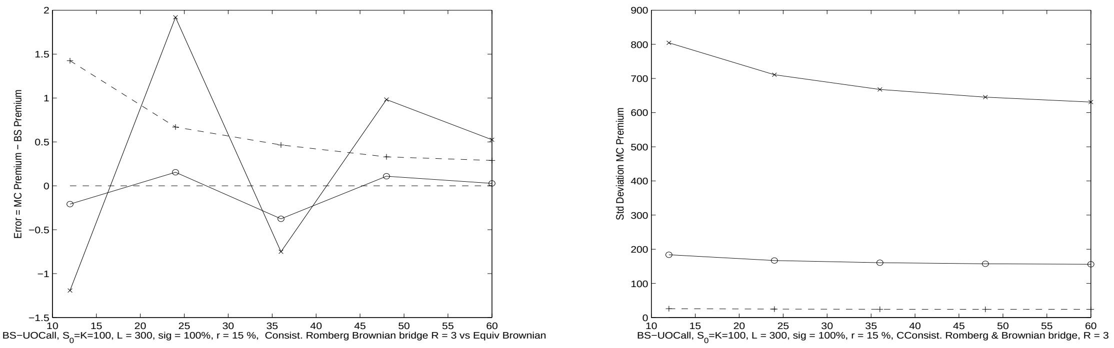


Figure 7: *B-S* Euro Partial Lookback Call option. (*a*) $M = 10^6$. *Romberg extrapolation (*$R = 3$*) of the Euler scheme with Brownian bridge:* o—o—o. *Consistent Romberg extrapolation (*$R = 3$*):* $\times\!\!-\!\!\times\!\!-\!\!\times$. *Euler scheme with Brownian bridge with equivalent complexity:* $+ - - + - - +$. $X_0 = 100$, $\sigma = 100\%$, $r = 15\,\%$, $\lambda = 1.1$. *Abscissas:* $6n$, $n = 2, 4, 6, 8, 10$. Left: *Premia.* Right: *Standard Deviations.* (*b*) *Idem with* $M = 10^8$.

(a)



(b)

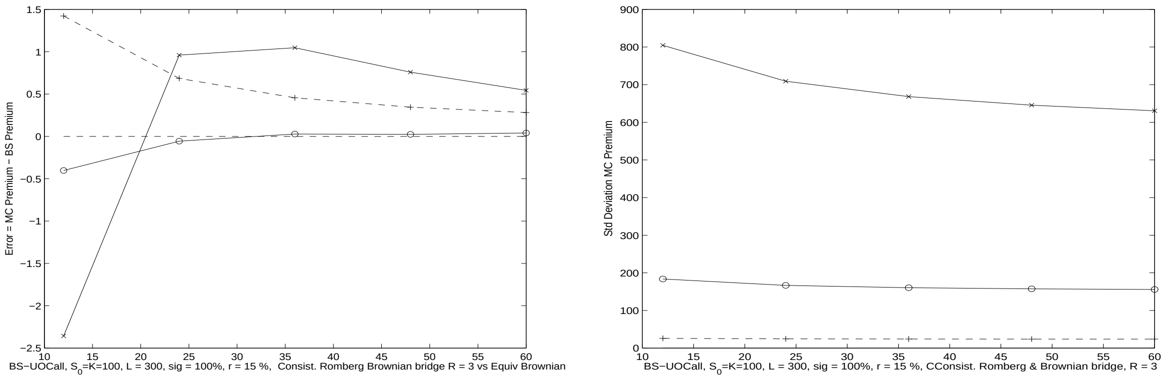


Figure 8: B-S Euro up-&-out Call option. (a) $M=10^6$. *Romberg extrapolation* $(R=3)$ *of the Euler scheme with Brownian bridge:* o—o—o. *Consistent Romberg extrapolation* $(R=3)$*:* —×—×—×—. *Euler scheme with Brownian bridge and equivalent complexity:* $+--+--+$. $X_0=K=100$, $L=300$, $\sigma=100\%$, $r=15\%$. *Abscissas:* $6n$, $n=2,4,6,8,10$. Left: *Premia.* Right: *Standard Deviations.* (b) *Idem with* $M=10^8$.

Romberg extrapolation with Brownian bridge clearly outperforms the continuous Euler scheme (Brownian bridge) of equivalent complexity while the 3-step Romberg extrapolation based on the stepwise constant Euler schemes with consistent Brownian increments is not competitive at all: it suffers from both a too high variance (see Figure 3.7.2(a) (right)) for the considered sizes of the Monte Carlo simulation and from its too slow rate of convergence in time.

When $M=10^8$ (see Figure 3.7.2(b) (left)), one verifies again that the time discretization error of the 3-step Romberg extrapolation almost vanishes like for the partial lookback option. This no longer the case with the 3-step Romberg extrapolation of stepwise constant Euler schemes. It seems clear that the discretization time error is more prominent for the barrier option: thus with $n=10$, the relative error is $\frac{9.09-8.54}{8.54}\approx 6.5\%$ by this first Romberg extrapolation whereas, the 3-step Romberg method based on the quasi-consistent "bridged" method yields a an approximate premium of 8.58 corresponding to a relative error of $\frac{8.58-8.54}{8.54}\approx 0.4\%$. These specific results (obtained without any control variate) are representative of the global behaviour of the methods as emphasized by Figure 3.7.2(b)(left).

3.7.3 The case of Asian options

The family of Asian options is related to the payoffs of the form

$$g\Big(\int_0^T X_s ds\Big).$$

This kind of options need some specific treatments to improve the rate of convergence of its time discretization. This is due to the continuity of the functional $f \mapsto \int_0^T f(s)ds$ in $L^1([0,T], dt)$.

This problem has been extensively investigated (essentially for a Black-Scholes dynamics) by E. Temam in his PHD thesis (see [34]). What follows comes from this work.

Let

$$X_t^x = x \exp(\mu t + \sigma W_t), \qquad \mu = r - \frac{\sigma^2}{2}.$$

$$\begin{aligned}
\int_0^T X_s^x ds &= \sum_{k=0}^{n-1} \int_{t_k^n}^{t_{k+1}^n} X_s^x ds \\
&= \sum_{k=0}^{n-1} X_{t_k^n}^x \int_0^{T/n} \exp(\mu s + \sigma W_s^{(t_k^n)}) ds
\end{aligned}$$

Roughly speaking $W_s^{(t_k^n)}$ is proportional to $\sqrt{T/n}$ so that

$$\exp(\mu s + \sigma W_s^{(t_k^n)}) = 1 + \mu s + \sigma W_s^{(t_k^n)} + \frac{\sigma^2}{2}(W_s^{(t_k^n)})^2 + \text{“}O(n^{-\frac{3}{2}})\text{”}.$$

Hence

$$\begin{aligned}
\int_0^{T/n} \exp(\mu s + \sigma W_s^{(t_k^n)}) ds &= \frac{T}{n} + \frac{\mu T^2}{2n^2} + \sigma \int_0^{T/n} W_s^{(t_k^n)} ds + \frac{\sigma^2}{2}\frac{T^2}{2n^2} + \frac{\sigma^2}{2}\int_0^{T/n}((W_s^{(t_k^n)})^2 - s)ds + \text{“}O(n^{-\frac{5}{2}})\text{”} \\
&= \frac{T}{n} + \frac{rT^2}{2n^2} + \sigma \int_0^{T/n} W_s^{(t_k^n)} ds + \text{“}O(n^{-2})\text{”}
\end{aligned}$$

since $\int_0^{T/n}((W_s^{(t_k^n)})^2 - s)ds$ is a centered with a quadratic norm proportional to n^{-2} [this is of course heuristic, but can be made rigorous, see [34]].

Now, using the results about Brownian bridges, we know (see Lemma (3.17)) that the $(W_t^{(t_k^n)})_{t\in[t_k^n, t_{k+1}^n]}$, $k = 0, \ldots, n-1$, are independent processes given $\sigma(\{W_{t_k^n},\, k = 1, \ldots, n\})$ and, owing to Lemma 3, the random vectors $\int_{t_k^n}^{t_{k+1}^n}(W_s - W_{t_k^n})ds$ are conditionally i.i.d. given $\{W_{t_k^n} = w_k,\, k = 1, \ldots, n\}$ with a conditional Gaussian distribution given by

$$\mathcal{L}\left(\int_0^{T/n} W_s\, ds \,|\, W_{\frac{T}{n}} = y\right) = \mathcal{N}\left(\frac{T}{2n} y; \frac{T^3}{12n^3}\right).$$

The parameters of this normal distribution can be computed using again Lemma 3:

$$\begin{aligned}
\mathbb{E}\left(\int_0^{T/n} W_s ds \,|\, W_{\frac{T}{n}}\right) &= \frac{n}{T}\int_0^{\frac{T}{n}} t\, dt\, W_{\frac{T}{n}} + \mathbb{E}\left(\int_0^{T/n} Y_s^{W,T/n} ds \,|\, W_{\frac{T}{n}}\right) \\
&= \frac{T}{2n} W_{\frac{T}{n}} + \int_0^{T/n} \mathbb{E}(Y_s^{W,T/n}) ds \\
&= \frac{T}{2n} W_{\frac{T}{n}}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(\left(\int_0^{T/n} W_s ds - \frac{T}{2n} W_{\frac{T}{n}}\right)^2 \mid W_{\frac{T}{n}}\right) &= \mathbb{E}\left(\left(\int_0^{T/n} Y_s^{W,T/n} ds\right)^2\right) \\
&= \int_{[0,T/n]^2} \mathbb{E}(Y_s^{W,T/n} Y_t^{W,T/n})\, ds\, dt \\
&= \frac{n}{T} \int_{[0,T/n]^2} (s \wedge t)(\frac{T}{n} - (s \vee t))\, ds\, dt \\
&= (\frac{T}{n})^3 \int_{[0,1]^2(} u \wedge v)(1 - (u \vee v)) du dv \\
&= \frac{1}{12} \left(\frac{T}{n}\right)^3.
\end{aligned}$$

Exercise. Use stochastic calculus to show directly that

$$\mathbb{E}\left(\int_0^T W_s ds - \frac{T}{2} W_T\right)^2 = \int_0^T (\frac{T}{2} - s)^2 ds = \frac{T^3}{12}.$$

Using that scheme, one derives the following proposition (we refer to the original paper for a detailed proof).

Proposition 9 *(cf. [34]) If g is Lipschitz continuous, then*

$$\|g(A_T) - g(\bar{A}_T)\|_p = O(n^{-\frac{3}{2}}).$$

More generally, one has for real valued Lipschitz functions G on $\mathbb{R}^2$

$$\|G(X^x{}_T, A_{T)} - g(G(X_T^x, \bar{A}_T)\|_p = O(n^{-\frac{3}{2}}).$$

4 Back to sensibility computation : the tangent processes

In a general setting, no closed form are available to compute sensibility. One approach is to rely on the tangent processes (when dealing with the δ and the γ). The main result on which we rely is due to Kunita (see [32])

Theorem 9 *Let b, $\sigma \in \mathcal{C}_b^1$ (continuously differentiable on with bounded partial derivatives). Let X^x denote the unique strong solution of the SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad X_0 = x \in \mathbb{R}^d. \tag{4.19}$$

where W is q-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then at every $t \in \mathbb{R}_+$, the mapping $x \mapsto X_t^x$ is a.s. continuously differentiable and its gradient $\nabla_x X_t^x$ satisfies the linear equation

$$\nabla_x X_t^x = I_d + \int_0^t \nabla b(X_s).\nabla_x X_s^x ds + \int_0^t (\nabla \sigma(X_s^x) \nabla_x X_s^x).dW_s$$

Remark. This embodies the non homogenous case by considering (t, X_t) instead of (X_t).

Example. If $q = d = 1$, then an elementary computation shows that

$$\nabla_x X_t^x = \exp\left(b'(X_s^x) - (\sigma'(X_s))^2/2\right)ds + \int_0^t \sigma'(X_s^x) dW_s$$

so that, in the Black-Scholes model ($b(x) = rx$, $\sigma(x) = \sigma x$), one retrieves that

$$\frac{d}{dx} X_t^x = \frac{X_t^x}{x}.$$

4.1 Application to sensibility computation

The tangent process and the δ are closely related. Assume a basket is made up of d risky asset with dynamics given by $(X_t^x)_{t\in[0,T]}$ with starting value $x \in (0,+\infty)^d$ at time 0, solution to (4.19). Assume the interest rate is 0 for simplicity. Then the premium of the payoff $g(X_T^x)$ on the basket is given by

$$v(x) := \mathbb{E}(g(X_T^x)).$$

The δ of this option (at time 0) is given by $v'(x)$. Under natural (domination) assumption on g, one shows that

$$\nabla v(x) = \mathbb{E}(\nabla g(X_T^x)\nabla X_T^x)$$

One can also consider a forward start payoffs $g(X_{T_1},\dots,X_{T_n})$. Then, its premium $v(x)$ is differentiable and

$$\nabla v(x) = \sum_{i=1}^{n} \mathbb{E}(\nabla g(X_{T_i}^x)\nabla X_{T_i}^x).$$

4.2 Extension to a parameter θ

Assume $b = b(\theta,.)$ and $\sigma = \sigma(\theta,.)$ and the initial value $x = x(\theta)$, $\theta\Theta \subset \mathbb{R}^q$. One can also differentiate a SDE with respect to this parameter θ, namely

$$(D_\theta X_t(\theta)) = (D_\theta x(\theta)) + \int_0^t \nabla_\theta b(X(\theta)_s)D_\theta(X(\theta)_s)ds + \int_0^t (\nabla_\theta\sigma(X(\theta)_s)D_\theta(X(\theta)_s)).dW_s.$$

This yields some expressions for the vega, etc.

4.3 Computation by simulation

One uses these formulae to compute some sensibility by Monte Carlo simulations: it suffices to consider *the Euler scheme of the couple* made up by the SDE and its tangent (or pseudo-tangent) processes.

5 Multi-asset American/Bermuda Options

in this section we will shift to slightly different notations: S_t will denote the price of (a vector of) asset(s) at time t and $X = (X_k)$ will denote a discrete time “structure” Markov process.

$\triangleright$ *d Traded risky assets:* $S_t = (S_t^1,\dots,S_t^d)$ $\quad t \in [0,T]$ with natural (augmented...) filtration $\mathcal{F}^S = (\mathcal{F}_t^S)_{t\in[0,T]}$.

$\triangleright$ *Discounted price:* $\widetilde{S}_t^i = \frac{S_t^i}{S_t^0} = e^{-rt}S_t^i$, $i = 1,\dots,d$, is a $(\mathbb{P},\mathcal{F}^S)$-martingale under the risk-neutral probability (if AOA holds) where r is a *riskless asset and Mathematical interest rate.*

$\triangleright$ *American Payoff process:* $(h_t)_{t\in[0,T]}$ is a nonnegative, $\mathcal{F}^S$-adapted process.

$\triangleright$ *American option on* $(h_t)_{t\in[0,T]}$*:*

$$\boxed{\text{Choose to receive } h_t \text{ once within 0 and } T}$$

$\triangleright$ *Bermuda option on* $(h_t)_{t\in[0,T]}$*:*

$$\boxed{\text{Choose to receive } h_{t_k} \text{ once, } k = 0,\dots,n.}$$

usually with $t_k = \dfrac{kT}{n}$, $k = 0, \ldots, n$.

Examples:

$\triangleright$ *Call/Put Option*: $\boxed{\text{Right to buy/sell once the asset } S \text{ at } \textit{the strike price } K}$

American: once at $t \in [0, T]$ *vs* Bermuda: once at a time $t = t_k = \dfrac{kT}{n}$, $k = 0, \ldots, n$.

$h_t = (S_t^1 - K)_+$ or $h_t = (K - S_t^1)_+$.

$\triangleright$ *"Vanilla" American Options*:

$$\boxed{\text{Right to receive once } h_t = h(t, S_t) \geq 0 \text{ within time 0 and } T}$$

vs Bermuda: once at a time $t = t_k = \dfrac{kT}{n}$, $k = 0, \ldots, n$.

Example: Exchange American/Bermuda options (Villeneuve):

$$h_t = (S_t^1 - \lambda\, S_t^2)_+.$$

$\triangleright$ *"Exotic" American/Bermuda Options*: $h_t \neq h(t, S_t)$.
/ms Example: American/Bermuda Asian options: $h_t = \left(\frac{1}{t}\int_0^t S_s ds - K\right)_+$.

American/Bermuda Lookback options, etc.

$\triangleright$ *"Shout" Options*:

$$\boxed{\text{Right to "shout" once within time 0 and } T}$$

vs Shout Bermuda: once at a time $t = t_k = \dfrac{kT}{n}$, $k = 0, \ldots, n$.

to receive (a non adapted) h_t at T.

5.1 Pricing Bermuda options: the dynamical programming principle

5.2 Markov structure process

(Replace $t_k = \frac{kT}{n}$ by k) Let $(X_k)_{0 \leq k \leq n}$ be a Markov structure process.
with transition $P_{k-1,k}(g)(x) = \mathbb{E}\left(g(X_{k+1} \,|\, X_k = x\right)$ such that

– $\mathcal{F}_k^X = \mathcal{F}_{t_k}^S$

– Risky asset vector satisfies

$$S_{t_k} = (S_{t_k}^1, \ldots, S_{t_k}^d) = G(X_k)$$

– Payoff process satisfies

$$h_{t_k} = h(k, X_k).$$

– Simulability: $(X_k)_{0 \leq k \leq n}$ can be simulated (at a reasonable cost).

• Typical structure processes (for American/Bermuda "Vanilla" options) :

$$X_k := \begin{cases} S_{t_k} & (Ex : X_k = W_{t_k} \text{the multi-dim } B\text{-}S \text{ model}) \\ \log(S_{t_k}) & \\ \bar{S}_{t_k} & \text{(Euler scheme)} \end{cases}$$

• For path-dependent options (Asian, lookback, etc)

$$X_k := \begin{cases} (S_{t_k}, \frac{1}{t_k}(S_0 + \cdots + S_{t_k})) \\ (\bar{S}_{t_k}, \frac{1}{t_k}(\bar{S}_0 + \cdots + \bar{S}_{t_k})), \\ (S_{t_k}, \max_{0\le i\le k} S_{t_i}), \\ \text{etc.} \end{cases}$$

5.3 Arbitrage and value function

$\boxed{\text{Step 1}}$

$$\begin{cases} \mathcal{V}_n & := & h(n, X_n) \\ & & \\ \mathcal{V}_k & := & \max\Big(h(k, X_k), \mathbb{E}(\mathcal{V}_{k+1}|\mathcal{F}_k^X)\Big). \end{cases}$$

$\boxed{\text{Step 2}}$ Backward induction based on the Markov property

Markov $\Longrightarrow$ Conditioning given $\mathcal{F}_k^X$ = Conditioning given X_k.

$$\boxed{\mathcal{V}_k = v_k(X_k), \qquad k = 0, \ldots, n.}$$

5.4 Arbitrage and optimal exercise

• Set for convenience $Z_k = \varphi(k, X_k)$ ("the obstacle")

$\boxed{\text{Step 1}}$ Backward induction on "local" optimal stopping times

$$\begin{cases} \tau_n & := & n \\ \tau_k & := & k\mathbf{1}_{\{Z_k > \mathbb{E}(Z_{\tau_{k+1}}\,|\,\mathcal{F}_k^X\}} + \tau_{k+1}\mathbf{1}_{\{Z_k \le \mathbb{E}(Z_{\tau_{k+1}}\,|\,\mathcal{F}_k^X)\}} \end{cases}$$

Markov $\Longrightarrow$ Conditioning given $\mathcal{F}_k^X$ = Conditioning given X_k.

$\boxed{\text{Step 2}}$ Fundamental theorem of Optimal Stopping Theory says

$$\mathcal{V}_0 = \mathbb{E}(Z_{\tau_0}) \quad i.e. \quad \mathcal{V}(0, X_0) = \mathbb{E}(\varphi(\tau_0, X_{\tau_0})).$$

6 Toward Numerical methods

6.1 Regression method (Longstaff-Schwarz, 1999)

$\boxed{\text{Approximation 1:}}$ Dimension Truncation

▷ At time k, one considers "a" basis

$$(e_1(X_k), e_2(X_k), \ldots, e_N(X_k), \ldots) \text{ of } L^2_{\mathbb{R}}(\Omega, \sigma(X_k), \mathbb{P}).$$

▷ Truncate at level N

$$e^{[N]}(X_k) := (e_1(X_k), e_2(X_k), \ldots, e_N(X_k)).$$

and set

$\triangleright \quad \tau_n^{[N]} := n,$

$\triangleright \quad \alpha_k^{[N]} := \text{argmin}\left\{\mathbb{E}(Z_{\tau_{k+1}^{[N]}} - \alpha.e^{[N]}(X_k))^2, \quad \alpha \in \mathbb{R}^N\right\}$, *i.e.*

$$\alpha_k^{[N]} = (\text{Gram}(e^{[N]}(X_k))^{-1}\left(< Z_{\tau_{k+1}^{[N]}} \,|\, e_\ell^{[N]}(X_k) >\right)_{1\le \ell\le N}$$

where $< . \,|\, . >$ denotes the canonical inner product in $L^2_{\mathbb{R}}(\Omega, \sigma(X_k), \mathbb{P})$ and $\text{Gram}(e^{[N]}(X_k))$ denotes the so-called Gram matrix of $e^{[N]}(X_k)$ defined by

$$\text{Gram}(e^{[N]}(X_k)) = \left[< e_\ell^{[N]}(X_k) \,|\, e_{\ell'}^{[N]}(X_k) >\right]_{1\le \ell,\ell'\le N}$$

$\triangleright \quad \tau_k^{[N]} := k\mathbf{1}_{\{Z_k > \alpha_k^{[N]}.e^{[N]}(X_k)\}} + \tau_{k+1}^{[N]}\mathbf{1}_{\{Z_k \le \alpha_k^{[N]}.e^{[N]}(X_k)\}}.$

$\boxed{\textsc{Approximation 2 :}}$

Backward recursive approximation of Forward MC simulation of $(X_k)_{0\le k\le n}$ and backward computation of $\tau_0^{[N]}$:

$\triangleright$ Simulate M independent copies $X^{(1)}, \ldots, X^{(m)}, \ldots, X^{(M)}$ of the structure process $X = (X_k)_{0\le k\le n}$.

$\triangleright$ For every $m \in \{1, \ldots, M\}$,

$$\tau_n^{[N,m,M]} := n$$

$\triangleright \quad \alpha_k^{[N],M} := \text{argmin}_{\alpha\in\mathbb{R}^N}\left(\frac{1}{M}\sum_{m=1}^M Z^{(m)}_{\tau_{k+1}^{[N],m,M}} - \alpha.e^{[N]}(X^{(m)})\right)^2$ so that

$$\alpha_k^{[N],M} = (\text{Gram}(e^{[N]}(X_k))^{-1}\left(\frac{1}{M}\sum_{m=1}^M Z^{(m)}_{\tau_{k+1}^{[N],m,M}} e_\ell^{[N]}(X_k^{(m)})\right)_{1\le \ell\le N}$$

$$\boxed{\triangleright \quad \tau_{k+1}^{[N],m,N} := k\mathbf{1}_{\{Z_k^{(m)} > \alpha_k^{[N],M}.e^{[N]}(X_k)\}} + \tau_{k+1}^{[N],m,N}\mathbf{1}_{\{Z_k^{(m)} \le \alpha_k^{[N],M}.e^{[N]}(X_k)\}}. \qquad (*)}$$

$\triangleright \quad \boxed{\mathcal{V}_0 \approx \mathbb{E}(Z_{\tau_0^{[N]}}) \approx \frac{1}{M}\sum_{m=1}^M Z_{\tau_0^{[N],m,M}}.}$

Theorem 10 *(Clément-Lamberton-Protter (2003), see [15]) The Monte Carlo approximation satisfies a CLT*

$$\left(\frac{1}{\sqrt{M}}\sum_{m=1}^M Z^{(m)}_{\tau_k^{[N],m,M}} - \mathbb{E}(Z_{\tau_k^{[N]}}), \sqrt{M}(\alpha_k^{[N],M} - \alpha_k^{[N]})\right)_{0\le k\le n-1} \xrightarrow{\mathcal{L}} \mathcal{N}(0;\Sigma).$$

$\triangleright$ Pros & Cons of the regression method:

• The method is "natural" : Approximation of conditional expectation by (affine) regression operator on a truncated basis of $L^2(\sigma(X_k), \mathbb{P})$.

• But :

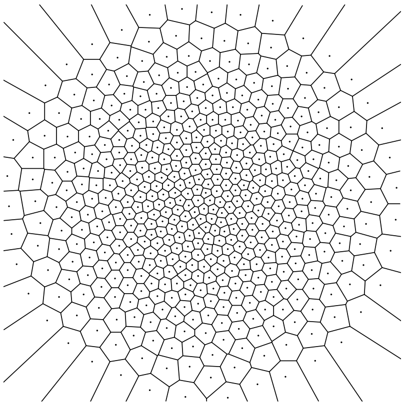


Figure 9: *An "optimal" quantization of the bi-variate normal distribution with size* $N = 500$

– Almost all computations are made *on-line*. However, note that the Gram matrix of $(e_\ell^{[N]}(X_k))_{1\le \ell\le N}$ can be computed off-line since it only depends on the structure process.

– The **choice of the functions** $e_k(x)$ $k \ge 1$ is crucial and needs much care and intuition. In practical implementations it may vary at every times step. Furthermore, it may have a biased effect in and out of the money.

A natural idea can be to consider an orthonormal basis for the underlying Markov structure process like *e.g.* the Hermite polynomials for the Brownian motion.

– Huge need of RAM (induces swapping).

• Furthermore : Strongly *pay-off* dependent.

6.2 Vector Quantization approach

Based on the value function.

APPROXIMATION 1: QUANTIZATION

Substitution by the nearest neighbour projection on grids Γ_k:

$$\widehat{X}_k = \pi_k(X_k) \longleftarrow X_k$$

where $\pi_k : \mathbb{R}^d \to \Gamma_k$, Γ_k is a grid of size N_k, $\Gamma_k = \{x_k^1, \dots, x_k^{N_k}\} \subset \mathbb{R}^d$.

But $\boxed{\text{loss of the Markov property}\dots}$

APPROXIMATION 2: MARKOV APPROXIMATION

$$\text{Quantized obstacle} : h(k, \widehat{X}_k),\ k = 0\dots, n.$$

The Markov property is forced: one defines $\widehat{V}_k$ by a backward induction

$$(\text{QDPP-I}) \equiv \left\{\begin{array}{lll} \widehat{V}_n & := & h(n, \widehat{X}_n) \\ \widehat{V}_k & := & \max(h(k, \widehat{X}_k), \mathbb{E}(\widehat{V}_{k+1} \,|\, \widehat{X}_k)), \quad k = 0, \dots, n-1. \end{array}\right.$$

Again a Backward induction

$$\widehat{V}_k = \widehat{v}_k(\widehat{X}_k), \quad k = 0, \dots, n.$$

where

$$
\text{(QDPP-II)} \equiv \begin{cases} \widehat{v}_n(x_n^i) &= h(n, x_n^i), \qquad i = 1, \ldots, N_n \\ \widehat{v}_k(x_k^i) &= \max\left(h(k, x_k^i), \sum_{j=1}^{N_k} \widehat{p}_k^{ij}\, \widehat{v}_{k+1}(x_{k+1}^j)\right), \quad i = 1, \ldots, N_k \\ & \qquad\qquad\qquad\qquad k = 1, \ldots, n-1. \end{cases}
$$

$\boxed{\textsc{Numerical Task(s)}}$ Optimize and Compute *off-line*

– Task 1: (good) grids Γ_k including the quantization error.

and

– Task 2: (accurate) quantized transitions $\widehat{p}_k^{ij} := \dfrac{\mathbb{P}(\widehat{X}_{k+1} = x_{k+1}^j,\, \widehat{X}_k = x_k^i)}{\mathbb{P}(\widehat{X}_k = x_k^i)}$.

$\boxed{\textsc{Conclusion}}$

(QDPP-II) is instantaneous for the *on line* computation of any portfolio of options.

$\boxed{\textsc{Interpretation}}$ Global Transition operators approximation

$$\text{Grids } \Gamma_k + \text{ quantized transitions } \widehat{p}_k^{ij}$$

$$\Downarrow$$

$$\boxed{\widehat{P}_{k-1,k}(x_k^i, dy) = \textstyle\sum_j \widehat{p}_k^{ij} \delta_{x_k^j}}$$

with

$$\widehat{P}_{k-1,k}(x_k^i, dy) \approx P_{k-1,k}(x, dy), \qquad k = 1, \ldots, n.$$

6.3 Quantization tree (I)

• For every $k \in \{0, \ldots, n\}$, $|\Gamma_k| = N_k$.

  • Theoretical complexity of a tree descent: $\quad \kappa \displaystyle\sum_{k=0}^{n-1} N_k\, N_{k+1}$.

  • Global size of the tree (constraint) : $\quad \displaystyle\sum_{k=0}^{n} N_k = N$.

The theoretical complexity is minimal when (Schwarz Inequality)

$$N_k = \frac{N}{n+1}$$

with complexity $\dfrac{n}{(n+1)^2} N^2$. Not so important in practise since

$$\text{Most connections } \widehat{p}_k^{ij} \text{ are negligible} \Longrightarrow \text{pruning...}$$

Theorem 11 (a) *(Bally-Pagès, 2001 (MCMA) to 2005 (Math.Fin.)) Scheme of order* 0 *(described above, to be compared to* non conformal finite elements of order 0*). If $h(k,.)$ are Lipschitz, the transitions $P_{k,k-1}$ are Lipschitz, the*

$$\|\mathcal{V}_0 - \widehat{v}_0(\widehat{X}_0)\|_2 \le C_{X,\varphi} \sum_{k=0}^{n} \|X_k - \widehat{X}_k^{\Gamma_k}\|_2.$$

(b) *(Bally-Pagès-Printems, (Math.Fin.), 2003) Scheme of order* 1 *(to be compared to* non conformal finite elements of order 1*). If (...)*

$$\|\mathcal{V}_0 - \widehat{v}_0(\widehat{X}_0)\|_2 \le C_{X,\varphi} \sum_{k=0}^{n} \|X_k - \widehat{X}_k^{\Gamma_k}\|_2^2.$$

These results show the interest to have access to optimal quantization grids *i.e.*, k being fixed, a grids Γ_k^* which minimizes the induced quantization error $\|X_k - \widehat{X}_k^{\Gamma_k^*}\|_2^2$.

6.4 Optimal quantization

6.4.1 Existence and rate of vector quantization

Let X temporarily denote a single random vector taking its values in $\mathbb{R}^d$. The fact that for every size N tehre exist a grid $\Gamma^{*,N}$ with at most N elements which minimizes

$$\|X - \widehat{X}^{\Gamma^{*,N}}\|_2 = \min_{|\Gamma|\le N} \|X - \widehat{X}^{\Gamma}\|_2$$

is true as soon as $X \in L^2(\mathbb{P})$ (see [38, 25]). Then it is rather simple to show that $N \mapsto \|X - \widehat{X}^{\Gamma^{*,N}}\|_2$ is a non-increasing sequence that goes to 0 as N goes to ∞. The rate of convergence to 0 is a much more challenging problem. An answer is provided by the so-called Zador Theorem stated below.

This theorem was first stated and established for distributions with compact supports by Zador (see [50, 51]). The first mathematically rigorous proof can be found in [25], and relies on a random quantization argument (Pierce Lemma).

Theorem 12 (a) Sharp rate. *Let $r > 0$ and $X \in L^{r+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi)\,d\xi \overset{\perp}{+} \nu(d\xi)$ be the canonical decomposition of the distribution of X (ν and the Lebesgue measure are singular). Then (if $\varphi \not\equiv 0$),*

$$e_{N,r}(X, \mathbb{R}^d) \sim \widetilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+r}}(u)\,du\right)^{\frac{1}{d}+\frac{1}{r}} \times N^{-\frac{1}{d}} \quad as \quad N \to +\infty. \tag{6.20}$$

where $\widetilde{J}_{r,d} \in (0,\infty)$.

(b) Non asymptotic upper bound *(Luschgy-P. 06). Let $d \ge 1$. There exists $C_{d,r,\eta} \in (0,\infty)$ such that, for every $\mathbb{R}^d$-valued random vector X,*

$$\forall\, N \ge 1, \qquad e_{N,r}(X, \mathbb{R}^d) \le C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}.$$

Remarks. • The real constant $\widetilde{J}_{r,d}$ clearly corresponds to the case of the uniform distribution over the unit hypercube $[0,1]^d$ for which a slightly more precise statement, namely

$$\lim_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \inf_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \widetilde{J}_{r,d}.$$

The proof is based on a self-similarity argument. The value of $\widetilde{J}_{r,d}$ depends on the reference norm on $\mathbb{R}^d$. When $d = 1$, elementary computations show that $\widetilde{J}_{r,1} = (r+1)^{-\frac{1}{r}}/2$. When $d = 2$, with the canonical Euclidean norm, one shows (see [36] for a proof, see also[25]) that $\widetilde{J}_{2,d} = \sqrt{\frac{5}{18\sqrt{3}}}$. Its exact value is unknown for $d \geq 3$ but, still for the canonical Euclidean norm, one has (see [25]) using some random quantization arguments,

$$\widetilde{J}_{2,d} \sim \sqrt{\frac{d}{2\pi e}} \approx \sqrt{\frac{d}{17,08}} \quad \text{as} \quad d \to +\infty.$$

6.4.2 Numerical aspects

The procedures that minimizes the quantization error are usually stochastic (except in 1-dimension). The most famous ones are undoubtedly the so-called *Competitive Leaning Vector Quantization* algorithm (see [38, 43] or [41]) and the Lloyd's I procedure (see [43, 40, 19]

Some algorithmic details are also available on the website

www.quantize.maths-fi.com

On this website are also available some (free access) grids for the d-variate normal distribution.

6.5 Optimal design of the quantization tree

Idea: optimal integral allocation problem

From item (a) of the theorem & non asympotic Zador's Theorem

$$\begin{aligned}
\|\mathcal{V}_0 - \widehat{v}_0(\widehat{X}_0)\|_2 &\leq C_{X,\varphi} \sum_{k=0}^{n} \|X_k - \widehat{X}_k^{\Gamma_k}\|_2 \\
&\leq C_{X,\varphi} C_\delta \sum_{k=0}^{n} \|X_k\|_{2+\delta} |\Gamma_k|^{-\frac{1}{d}} \\
&= C_{X,\varphi} C_\delta \sum_{k=0}^{n} \|X_k\|_{2+\delta} N_k^{-\frac{1}{d}}
\end{aligned}$$

Amounts to solving the

$$\min_{N_0+\cdots+N_n=N} \sum_{k=0}^{n} \|X_k\|_{2+\delta} N_k^{-\frac{1}{d}}$$

i.e. denoting the (upper) integral part of x by $\lceil x \rceil$,

$$N_k = \left\lceil \frac{(\|X_k\|_{2+\delta})^{\frac{d}{d+1}}}{\sum_{0\leq \ell \leq n} (\|X_\ell\|_{2+\delta})^{\frac{d}{d+1}}} N \right\rceil, \qquad k = 0, \ldots, n$$

so that

$$\boxed{\|\mathcal{V}_0 - \widehat{v}_0(\widehat{X}_0)\|_2 \leq C_{X,\varphi} C_\delta \left(\sum_{k=0}^{n} (\|X_k\|_{2+\delta})^{\frac{d}{d+1}} \right)^{1-\frac{1}{d}} \widetilde{N}^{-\frac{1}{d}}.}$$

with $\widetilde{N} = N_0 + \cdots + N_n$ (usually $> N$).

Examples:

• Brownian motion $X_k = W_{t_k}$: Then $\widehat{W}_0 = 0$ and

$$\|W_{t_k}\|_{2+\delta} = C_\delta\sqrt{t_k}, \qquad k = 0,\ldots,n.$$

Hence $N_0 = 1$ and

$$N_k \approx \frac{2(d+1)}{d+2}\left(\frac{k}{n}\right)^{\frac{d}{2(d+1)}} N, \qquad k = 1,\ldots,n;$$

$$|\mathcal{V}_0 - \widehat{v}_0(0)| \le C_{W,\delta}\left(\frac{2(d+1)}{d+2}\right)^{1-\frac{1}{d}} \frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}} = O(\frac{n}{\bar{N}^{\frac{1}{d}}})$$

with $\bar{N} = \frac{N}{n}$. Useful but theoretically not crucial. Numerically......
• Stationary process (ex: $X_k = OU_{t_k}$):

– Essential feature: Only needs

ONE OPTIMAL GRID ... and ONE QUANTIZED TRANSITION MATRIX

– $\|X_k\|_{2+\delta} = \|X_0\|_{2+\delta}$
hence

$$N_k = \left\lceil \frac{N}{n+1} \right\rceil, \quad k = 0,\ldots,n.$$

$$\|\mathcal{V}_0 - \widehat{v}_0(\widehat{X}_0)\|_2 \le C_{X,\delta}\frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}} = C_{X,\delta}\frac{n}{\bar{N}^{\frac{1}{d}}}$$

with $\bar{N} = \frac{N}{n}$.

6.6 Computing the quantized transitions $\widehat{p}_k^{ij}$

6.6.1 Standard Monte Carlo estimation

• As a companion procedure of grid updating:

– Nearest neighbour search at every time step to update the grid $\Gamma_k \subset \mathbb{R}^d$ *via* $CLVQ$ and the transition frequency estimators.

– or "batch" estimation *via* randomized Lloyd's I procedure.

• Freeze the grids and carry on the MC estimation of the transitions.

– M independent copies $X^m = (X_0^m, X_1^m, \ldots, X_n^m)$, $m = 1,\ldots,M$ "passing through" the quantization tree.

6.6.2 Alternative methods

• Fast tree quantization for Gaussian structure processes ([11] for swing options, see further on).

• The "spray" method ("gerbes" in French) ([42] for filtering by optimal quantization).

6.7 δ-Hedging, higher order schemes...

6.7.1 Computing the δ-hedge, $X_k = S_{t_k}$ (B-S) or $\bar{S}_{t_k}$ (local vol).

• Quantized δ-Hedging:

• $\widehat{\zeta}_k^n := \dfrac{n}{Tc^2(\widehat{S}_{t_k})}\widehat{\mathbb{E}}_k\left((\widehat{v}_{k+1}^n(\widehat{S}_{t_{k+1}}) - \widehat{v}_k^n(\widehat{S}_{t_k}))(\widehat{S}_{t_{k+1}} - \widehat{S}_{t_k})\right).$

• Similar formulae for the Euler scheme... $(\mathcal{H}) \equiv (i)\ \sigma \in C_b^\infty(\mathbb{R}^d), \quad (ii)\ \sigma\sigma^* \geq \varepsilon_0 I_d, \quad (iii)\ \|x\sigma'(x)\|_\infty < +\infty.$

• Bermuda Error:

$$\mathbb{E}\int_0^T |c^*(S_u)(Z_u - \zeta_u^n)|^2\, ds \leq C_{h,\sigma}\frac{(1+|s_0|)^q}{\varepsilon_0}\frac{1}{n^{\frac{1}{6}}}.$$

• Quantization Error:

$$\mathbb{E}\int_0^T |\zeta_u^n - \widehat{\zeta}_u^n| du \leq C(1+|s_0|)|\frac{n^{\frac{3}{2}}}{(N/n)^{\frac{1}{d}}}.$$

6.8 Numerical experiments (with quantization)

6.8.1 Numerical experiments I: Exchange geometric options

• Exchange American options on geometric assets.

• Reference: [49] Villeneuve-Zanette, 1998 *Finite differences for 2-Dim exchange American options with dividends.*

• Model: Standard $2d$-dim (B & S) model with *non correlated* Brownian Motions (The most "hostile" to quantization...).

• Maturity: $T = 1$ year. Volatility : $\sigma_i = \dfrac{20\%}{\sqrt{d}}$, $i = 1, \ldots, d$.

• $2d$-dim pay-Off: $$h(t,x) = \left(\prod_{i=1}^d e^{-\mu_i t} S_t^i - \prod_{i=d+1}^{2d} e^{-\mu_i t} S_t^i\right)_+.$$

• Initial values: $\prod_{i=1}^d S_0^i = 40$, $\prod_{i=d+1}^{2d} S_0^i = 36$ (in-the money), $\mu_1 := 5\,\%$, $\mu_2 = 0, \ldots$
$\prod_{i=1}^d S_0^i = 36$, $\prod_{i=d+1}^{2d} S_0^i = 40$ (out-of-the money), $\mu_{d+1} := 0\,\%, \ldots$

6.8.2 Results: Premium and δ-hedge: 0-order scheme

6.8.3 0-order scheme *vs* 1-order scheme

Computation velocity: Pentium II, 800 MHz, 500 MO RAM [2003...]

$$d = 5 \qquad N = 2.10^4 \qquad n = 10$$

• Design of the quantization tree (grid/weights) : 3 seconds;

• (Premium+ δ-Hedge) (QBDPP): 3 per second.

Maturity	3 months		6 months		9 months		12 months	
AM_{ref}	**4.4110**		**4.8969**		**5.2823**		**5.6501**	
	Price	Error (%)	Price	Error (%)	Price	Error (%)	Price	Error (%)
$d = 2$	4.4111	0.0023	4.8971	0.0041	5.2826	0.0057	5.6505	0.0071
$d = 4$	4.4076	0.08	4.9169	0.34	5.3284	0.82	5.7366	1.39
$d = 6$	4.4156	0.1	4.9276	0.63	5.3550	1.38	5.7834	2.20
$d = 10$	4.4317	0.47	4.9945	2.00	5.4350	2.89	5.8496	3.53

Table 1: American Premium & Relative error. Different maturities and dimensions.

(a) (b)

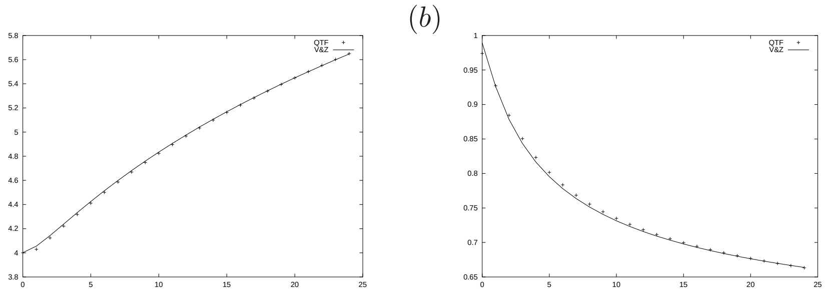


Figure 10: $d = 2$, $n = 25$ and $\bar{N} = 300$. (a) American premium as a function of the maturity. (b) Hedging strategy on the first asset. The cross + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z).

(a) (b)

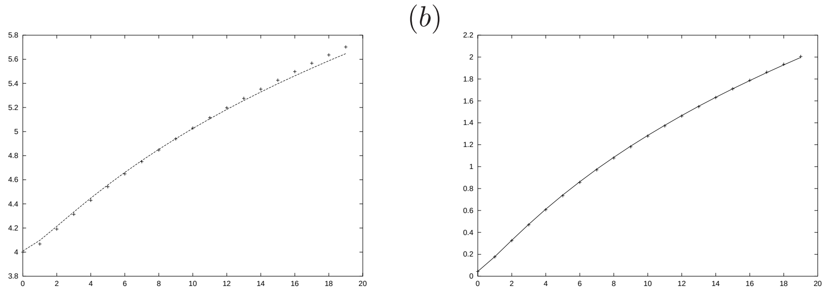


Figure 11: $d = 4$. American premium as a function of the maturity. *(a) In-the-money. (b) Out-of-the-money. + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z).*

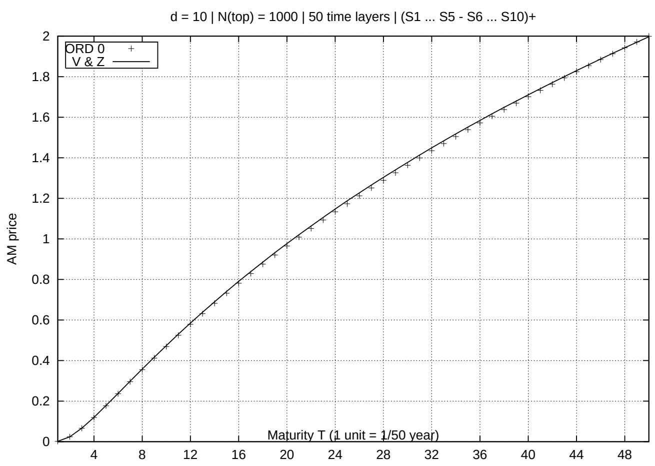


Figure 12: Exchange option $10D$ $(S^1 \cdots S^5 - S^6 \cdots S^{10})_+$: out-of-the-money

Maturity	3 months		6 months		9 months		12 months	
AM_{ref}	**4.4110**		**4.8969**		**5.2823**		**5.6501**	
	Price	Error (%)	Price	Error (%)	Price	Error (%)	Price	Error (%)
$d = 4$								
AM_0	4.4076	0.08	4.9169	0.34	5.3284	0.82	5.7366	1.39
AM_1	4.4058	0.1	4.8991	0.04	5.2881	0.08	5.6592	0.13
$d = 6$								
AM_0	4.4156	0.1	4.9276	0.63	5.3550	1.38	5.7834	2.20
AM_1	4.4099	0.02	4.8975	0.01	5.3004	0.34	5.6557	0.10
$d = 10$								
AM_0	4.4317	0.47	4.9945	2.00	5.4350	2.89	5.8496	3.53
AM_1	4.4194	0.19	4.8936	0.07	5.1990	1.58	5.4486	3.56

Table 2: Relative errors of AM_0 and AM_1 with respect to a reference price for different maturities and dimensions.

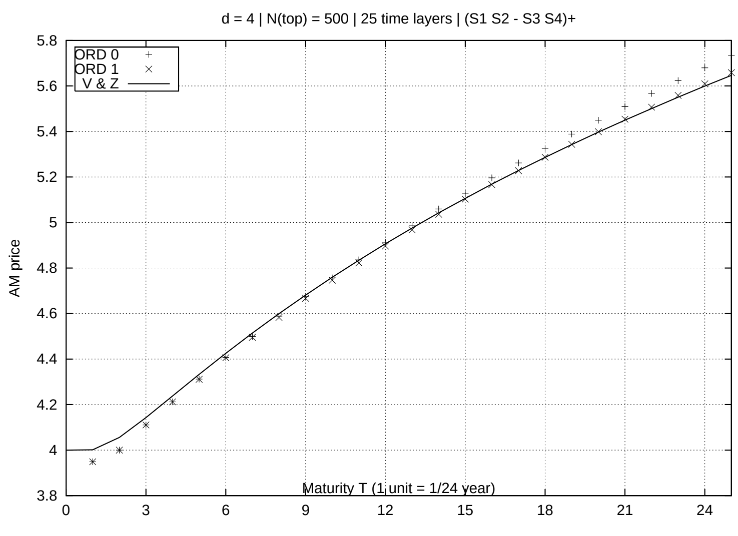


Figure 13: Exchange option 4D $(S^1S^2 - S^3S^4)_+$: In-the-money. Dimension $d = 4$, $n = 25$ and $N_{25} = 500$. American option function of the maturity T. The crosses denote the quantized version with order 0 (+) and order 1 (×)

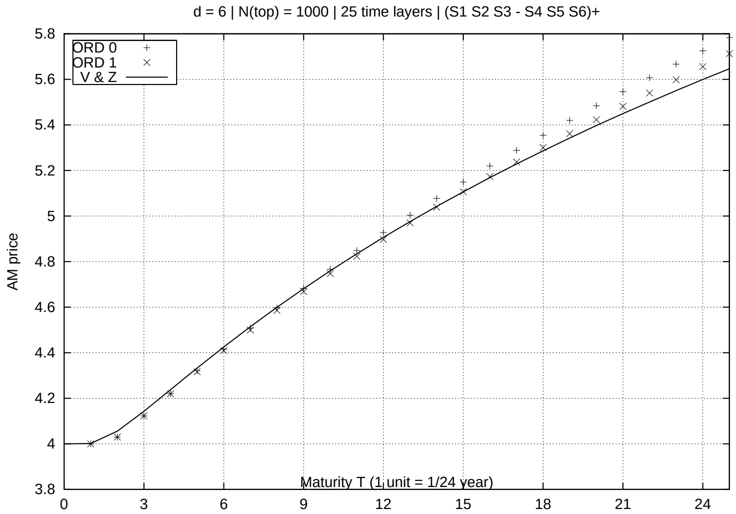


Figure 14: Quantized version order 0 (+), order 1 (×). (a) Dimension $d = 6$, $n = 25$, $N_{25} = 1000$, In-the-money case. Value of the American option function of the maturity T.

6.9 Swing Options (a word about)

Take or Pay contract on gas (with firm constraints)

– Spot or day-ahead delivery contract S_{t_k} *assumed to Markov (for convenience) i.e.*

$$X_k = S_{t_k}$$

– *Local volume constraints*: Buy daily $q_{t_k} \in [q_{\min}, q_{\max}]$ m^3 of natural gas at price K_k

– *Global volume constraints* $Q_{\min} \leq q_0 + q_{t_1} + \cdots + q_{t_{n-1}} \leq Q_{\max}$.

$$P(Q_{\min}, Q_{\max}, s_0) = \sup_{(q_{t_k})_{0\leq k\leq n-1}\in\mathcal{A}_{Q_{\min},Q_{\max}}} \mathbb{E}\left(\sum_{k=0}^{n-1} q_{t_k} e^{-r(T-t_k)}(S_{t_k} - K_k)\right)$$

where the set of admissible daily purchased quantities is given by

$$\mathcal{A}_{Q_{\min},Q_{\max}} = \left\{(q_{t_k})_{0\leq k\leq n-1},\ q_{t_k} \underline{\in} \mathcal{F}^S_{t_k},\ 0 \leq q_{t_k} \leq 1,\ Q_{\min} \leq \sum_{0\leq k\leq n-1} q_{t_k} \leq Q_{\max}\right\}$$

6.9.1 Pricing swing by (optimal) Quantization

It is a stochastic control problem.

▷ Dynamic programming principle on the price $P(t_k, S_{t_k}, Q_{t_k})$

$$\begin{aligned} P(t_k, S_{t_k}, Q_{t_k}) = \max\{& q(S_{t_k} - K) + \mathbb{E}(P(t_{k+1}, S_{t_{k+1}}, Q_{t_k} + q)|S_{t_k}), \\ & q \in [q_{\min}, q_{\max}],\ Q_{t_k} + q \in [(Q_{\min} - (n-k)q_{\max})_+, (Q_{\max} - (n-k)q_{\min})_+]\}\,. \end{aligned} \tag{6.21}$$

▷ Bang-bang control (see [11]).

If $\dfrac{Q_{\max} - Q_{\min}}{q_{\max} - q_{\min}} \in \mathbb{N}$, then the optimal control is bang-bang *i.e.* $\{q_{\min}, q_{\max}\}$-valued.

▷ Quantized Dynamic programming principle Let $\widehat{S}_{t_k}$ be an (optimal) quantization of S_{t_k} taking values in $\Gamma_k := \{s_k^1, \ldots, s_k^{N_k}\}$, $k = 0, \ldots, n$.

$$\begin{cases} P(t_k, s_k^i, \hat{Q}_{t_k}) = \max\limits_{q \in \mathcal{A}_k^{\hat{Q}_{t_k}}} [q(s_k^i - K) + \mathbb{E}(P(t_{k+1}, \hat{S}_{t_{k+1}}, \hat{Q}_{t_k} + q) | \hat{S}_{t_k} = s_k^i)] \\ \mathcal{A}_k^{\hat{Q}_{t_k}} = \{q \in \{q_{\min}, q_{\max}\},\ \hat{Q}_{t_k} + q \in [(Q_{\min} - (n-k)q_{\max})_+, (Q_{\max} - (n-k)q_{\min})_+]\} \\ \qquad\qquad i = 1, \ldots, N_k, \\ P(T, s_T^i, \hat{Q}_T) = P_T(s_T^i, \hat{Q}_T),\ i = 1, \ldots, N_n. \end{cases} \tag{6.22}$$

Since $\hat{S}_{t_k}$ takes its values in Γ_k, we can rewrite the conditional expectation as:

$$\mathbb{E}(P(t_{k+1}, \hat{S}_{t_{k+1}}, Q) | \hat{S}_{t_k} = s_k^i) = \sum_{j=1}^{N_{k+1}} P(t_{k+1}, s_{k+1}^j, Q) \widehat{p}_k^{ij}$$

where

$$\widehat{p}_k^{ij} = \mathbb{P}(\hat{S}_{t_{k+1}} = s_{k+1}^j | \hat{S}_{t_k} = s_k^i)$$

- Technical Parameters:
  - ◦ Optimal quantization approach $n = 365$ (1 year), $N_k = \bar{N} = 100$
  - ◦ Longstaff-Schwartz approach (following [10]: MC size, $M = 1000$).
- Processor: Céléron, CPU 2,4 GHz. RAM 1,5 Go

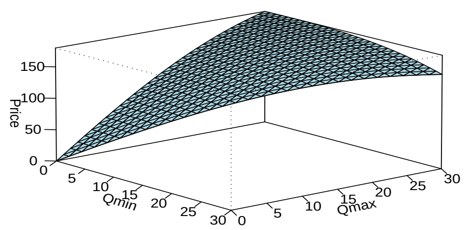


Figure 15: *Price Surface (as a function of global constraints) by Optimal Quantization*

6.10 Quantization *vs* *L-S* for Swing options

- First results:
  - ◦ 1 contract:

L-S	Quantization: Transitions + pricing	Quantization: Pricing alone
160 sec	65 sec	5 sec

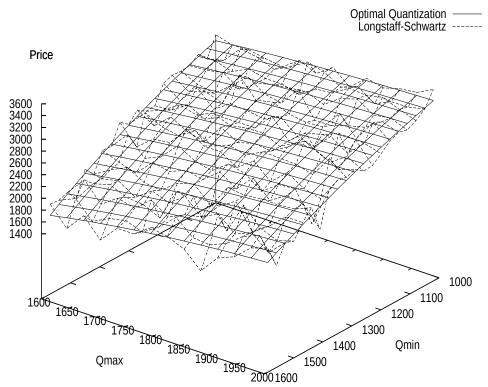


Figure 16: *Price Surface by L-S like procedure (dotted lines) and by Optimal Quantization (solid lines)*

◦ 10 contracts:

L-S	Quantization
1600 sec	110 sec

- If less RAM available:
  - ◦ Quantization is unchanged
  - ◦ *L-S* slows down because the computers “swaps”...

References

[1] B. Arouna (2004). Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.* , **10**(1), 1-24.

[2] J. Barraquand, D. Martineau (1995). Numerical valuation of high dimensional multivariate American securities, *Journal of Finance and Quantitative Analysis*, **30**.

[3] V. Bally, D. Talay (1996). The distribution of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function, *Probab. Theory Related Fields*, **104**(1), 43-60 (1996).

[4] V. Bally, D. Talay (1996). The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density, *Monte Carlo Methods Appl.*, **2**(2), 93-128.

[5] V. Bally, G. Pagès et J. Printems (2001). A Stochastic quantization method for nonlinear problems, *Monte Carlo Methods and Appl.*, 7(1), 21-34.

[6] V. Bally, G. Pagès (2003). A quantization algorithm for solving discrete time multidimensional optimal stopping problems, *Bernoulli*, 9(6), 1003-1049.

[7] V. Bally, G. Pagès (2003). Error analysis of the quantization algorithm for obstacle problems, *Stochastic Processes & Their Applications*, 106(1), 1-40.

[8] V. Bally, G. Pagès, J. Printems (2003). First order schemes in the numerical quantization method, *Mathematical Finance* 13(1), 1-16.

[9] V. Bally, G. Pagès, J. Printems (2005). A quantization tree method for pricing and hedging multidimensional American options, *Mathematical Finance*, 15(1), 119-168.

[10] C. Barrera-Esteve and F. Bergeret and C. Dossal and E. Gobet and A. Meziou and R. Munos and D. Reboul-Salze 2006. Numerical methods for the pricing of Swing options: a stochastic control approach, *Methodology and Computing in Applied Probability.*

[11] O. Bardou, S. Bouthemy, G. Pagès (2006) Pricing swing options using Optimal Quantization, pre-print, submitted.

[12] O. Bardou, S. Bouthemy, G. Pagès (2006) When are Swing options bang-bang and how to use it?, pre-print, submitted.

[13] A. Beskos, G.O. Roberts (2005). Exact simulation of diffusions, *Ann. appl. Prob.*, **15**(4), 2422-2444.

[14] N. Bouleau, D. Lépingle (1994). :em Numerical methods for stochastic processes , Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 359 pp. ISBN: 0-471-54641-0.

[15] E. Clément, A. Kohatsu-Higa, D. Lamberton (2006). A duality approach for the weak approximation of Stochastic Differential Equations, *Annals of Applied Probability*, **16**(3), 449-471.

[16] É. Clément, P. Protter, D. Lamberton (2002). An analysis of a least squares regression method for American option pricing, *Finance & Stochastics*, **6**, n^0 2, 449-47.

[17] L. Devroye, *Non uniform random variate generation*, Springer. Available as a pdf file on Luc Devroye's web page webpage at `cg.scs.carleton.ca/ luc/books-luc.html`.

[18] A. Friedman (1975). *Stochastic differential equations and applications. Vol. 1-2. Probability and Mathematical Statistics*, Vol. 28. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 528p.

[19] A. Gersho and R.M. Gray (1992). *Vector Quantization and Signal Compression.* Kluwer, Boston.

[20] P. Glasserman (2003). Monte Carlo Methods in Financial Engineering, Springer-Verlag, New York, 596p.

[21] E. Gobet, G. Pagès, H. Pham, J. Printems (2005). Discretization and simulation for a class of SPDE's with applications to Zakai and McKean-Vlasov equations, to appear in *SIAM Journal on Numerical Analysis.*

[22] E. Gobet, S. Menozzi (2004): Exact approximation rate of killed hypo-elliptic diffusions using the discrete Euler scheme, *Stochastic Process. Appl.*, **112**(2), 201-223.

[23] E. Gobet (2000): Weak approximation of killed diffusion using Euler schemes, *Stoch. Proc. and their Appl.*, **87**, 167-197.

[24] E. Gobet, R. Munos (2005). Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to optimal control, *SIAM J. Control Optim.*, **43**(5), 1676-1713.

[25] S. Graf, H. Luschgy (2000). *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics $n^0$1730, Springer, 2000, 230p.

[26] J. Guyon (2006). Euler scheme and tempered distributions, *Stochastic Processes and their Applications*, **116**(6), 877-904.

[27] S.L. Heston (1993). A closed-form solution for options with stochastic volatility with applications to bon and currency options, *The review of Financial Studies*, **6**, $n^0 2$, 327-343.

[28] I. Karatzas, S.E. Schreve (1998). *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer-Verlag, New York, 470p.

[29] A.G. Kemna, A.C. Vorst (1990). A pricing method for options based on average asset value, *J. Banking Finan.*, **14**, 113-129.

[30] P.E. Kloeden, E. Platten (1992). *Numerical solution of stochastic differential equations. Applications of Mathematics*, **23**, Springer-Verlag, Berlin (New York), 632 pp. ISBN: 3-540-54062-8.

[31] A. Kohatsu-Higa, R. Pettersson (2002): Variance reduction methods for simulation of densities on Wiener space, *SIAM J. Numer. Anal.*, **40**(2), 431-450.

[32] H. Kunita (1982): Stochastic differential equations and stochastic flows of diffeomorphisms, *Cours d'école d'été de Saint-Flour*, LN-1097, Springer-Verlag, 1982.

[33] D. Lamberton, B. Lapeyre (1996). Introduction to stochastic calculus applied to finance. Chapman & Hall, London, 185 pp. ISBN: 0-412-71800-6.

[34] B. Lapeyre, E. Temam (2001): Competitive Monte Carlo methods for the pricing of Asian Options, *J. of Computational Finance*, **5**(1), 39-59.

[35] F.A. Longstaff, E.S. Schwarz (2001) Valuing American options by simulation: a simple least-squares approach, *Review of Financial Studies*, **14**, 113-148.

[36] D.J. Newman (1982). The Hexagon Theorem. *IEEE Trans. Inform. Theory*, **28**, 137-138.

[37] Numerical Recipes: Textbook available at `www.ma.utexas.edu/documentation/nr/bookcpdf.html`. See also [45].

[38] G. Pagès (1998). A space vector quantization method for numerical integration, *J. Computational and Applied Mathematics*, 89, 1-38.

[39] G. Pagès (2007). Multistep Richardson-Romberg extrapolation: controlling variance and complexity, to appear in *Monte Carlo Methods and Applications J.*.

[40] G. Pagès (2007). Quadratic optimal functional quantization of stochastic processes and numerical applications, to appear in*Proceedings of MCQMC 06', Ulm*, Springer, pre-prtint LPMA.

[41] G. Pagès, H. Pham, J. Printems (2004). Optimal quantization methods and applications to numerical problems in finance, *Handbook on Numerical Methods in Finance* (S. Rachev, ed.), Birkhauser, Boston, 253-298.

[42] G. Pagès, H. Pham, J. Printems (2004). An Optimal Markovian Quantization Algorithm for Multidimensional Stochastic Control Problems, Stochastics and Dynamics, 4(4), 501-545.

[43] G. Pagès, J. Printems (2003). Optimal quadratic quantization for numerics: the Gaussian case, *Monte Carlo Methods and Appl.* 9(2), 135-165.

[44] G. Pagès, J. Printems (2005). `www.quantize.maths-fi.com`, website devoted to optimal quantization.

[45] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery (2002). Numerical recipes in C++. The art of scientific computing. Second edition, updated for C++. Cambridge University Press, Cambridge, 2002. 1002p. ISBN: 0-521-75033-4 65-04.

[46] D. Revuz, M. Yor (1999). *Continuous martingales and Brownian motion.* Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **293**, Springer-Verlag, Berlin, 602 p. ISBN: 3-540-64325-7 60G44 (60G07 60H05)

[47] P. Seumen-Tonou (1997). *Méthodes numériques probabilistes pour la résolution d'équations du transport et pour l'évaluation d'options exotiques*, Theèse de l'Université de Provence (Marseille, France), 116p.

[48] D. Talay, L. Tubaro (1990). Expansion of the global error for numerical schemes solving stochastic differential equations, *Stoch. Anal. Appl.*, **8**, 94-120.

[49] S. Villeneuve, A. Zanette (2002). Parabolic A.D.I. methods for pricing american option on two stocks, *Mathematics of Operation Research*, **27**, n^{o}1, 121-149.

[50] P.L. Zador (1963). Development and evaluation of procedures for quantizing multivariate distributions. Ph.D. dissertation, Stanford Univ.

[51] P.L. Zador (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory*, **28**(2), 139-149.