2.1.3 Vanilla Option Pricing in a Black–Scholes Model: The Premium

For the sake of simplicity, we consider a 2-dimensional risk-neutral correlated Black–Scholes model for two risky assets X^1 and X^2 under its unique risk neutral probability (but a general d-dimensional model can be defined likewise):

$$dX_t^0 = rX_t^0 dt, \;\; X_0^0 = 1,$$
$$dX_t^1 = X_t^1\big(rdt + \sigma_1 dW_t^1\big), X_0^1 = x_0^1, \;\; dX_t^2 = X_t^2\big(rdt + \sigma_2 dW_t^2\big), X_0^2 = x_0^2, \tag{2.2}$$

with the usual notations (r interest rate, $\sigma_i > 0$ volatility of X^i). In particular, $W = (W^1, W^2)$ denotes a *correlated* bi-dimensional Brownian motion such that

$$\langle W^1, W^2\rangle_t = \rho t, \;\; \rho \in [-1, 1].$$

This implies that W^2 can be decomposed as $W_t^2 = \rho W_t^1 + \sqrt{1-\rho^2}\,\widetilde{W}_t^2$, where $(W^1, \widetilde{W}^2)$ is a standard 2-dimensional Brownian motion. The filtration $(\mathcal{F}_t)_{t\in[0,T]}$ of this market is the augmented filtration of W, *i.e.* $\mathcal{F}_t = \mathcal{F}_t^W := \sigma(W_s, \; 0 \le s \le t, \; \mathcal{N}_{\mathbb{P}})$ where $\mathcal{N}_{\mathbb{P}}$ denotes the family of $\mathbb{P}$-negligible sets of $\mathcal{A}$ ([5]). By "filtration of the market", we mean that $(\mathcal{F}_t)_{t\in[0,T]}$ is the smallest filtration satisfying the usual conditions to which the process $(X_t)_{t\in[0,T]}$ is adapted. By "risk-neutral", we mean that $e^{-rt}X_t$ is a $\big(\mathbb{P}, (\mathcal{F}_t)_t\big)$-martingale. We will not go further into financial modeling at this stage, for which we refer *e.g.* to [185] or [163], but focus instead on numerical aspects.

For every $t \in [0, T]$, we have

$$X_t^0 = e^{rt}, \quad X_t^i = x_0^i e^{(r-\frac{\sigma_i^2}{2})t + \sigma_i W_t^i}, \;\; i = 1, 2.$$

(One easily verifies using Itô's Lemma, see Sect. 12.8, that X_t thus defined satisfies (2.2); formally finding the solution by applying Itô's Lemma to $\log X_t^i$, $i = 1, 2$, assuming *a priori* that the solutions of (2.2) are positive).

When $r = 0$, X^i is called a geometric Brownian motion associated to W^i with volatility $\sigma_i > 0$.

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

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

$/T$ (e^{-rT})

[5] One shows that, owing to the 0-1 Kolmogorov law, this filtration is right continuous, *i.e.* $\mathcal{F}_t = \bigcap_{s>t}\mathcal{F}_s$. A right continuous filtration which contains the $\mathbb{P}$-negligible sets satisfies the so-called "usual conditions".

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

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

The fact that W has independent stationary increments implies that X^1 and X^2 have independent stationary ratios, *i.e.*

$$\left(\frac{X_T^i}{X_t^i}\right)_{i=1,2} \stackrel{d}{=} \left(\frac{X_{T-t}^i}{x_0^i}\right)_{i=1,2} \text{ is independent of } \mathcal{F}_t.$$

As a consequence, if we define for every $T > 0$, $x_0 = (x_0^1, x_0^2) \in (0, +\infty)^2$,

$\times e^{-rT}$

$$v(x_0, T) = e^{-rT} \mathbb{E}\, h(X_T),$$

then

$$\begin{aligned}
V_t &= e^{-r(T-t)} \mathbb{E}\left(h(X_T) \mid \mathcal{F}_t\right) \\
&= e^{-r(T-t)} \mathbb{E}\left(h\left(X_t^i \times \left(\frac{X_T^i}{X_t^i}\right)_{i=1,2}\right) \Big| \mathcal{F}_t\right) \\
&= e^{-r(T-t)} \left[\mathbb{E}\, h\left(\left(x^i \frac{X_{T-t}^i}{x_0^i}\right)_{i=1,2}\right)\right]_{|x^i = X_t^i,\, i=1,2} \qquad \text{by independence} \\
&= v(X_t, T-t).
\end{aligned}$$

⊳ **Examples. 1.** *Vanilla call* with strike price K:

$$h(x^1, x^2) = (x^1 - K)_+.$$

There is a closed form for such a *call* option – the celebrated *Black–Scholes formula* for option on stock (without dividend) – given by

$\times e^{-rT}$

$$\text{Call}_0^{BS} = C(x_0, K, r, \sigma_1, T) = x_0 \Phi_0(d_1) - e^{-rT} K \Phi_0(d_2) \tag{2.3}$$

$$\text{with } d_1 = \frac{\log(x_0/K) + (r + \frac{\sigma_1^2}{2})T}{\sigma_1 \sqrt{T}}, \quad d_2 = d_1 - \sigma_1 \sqrt{T}, \tag{2.4}$$

where Φ_0 denotes the c.d.f. of the $\mathcal{N}(0; 1)$-distribution.

2. *Best-of-call* with strike price K:

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

A quasi-closed form is available involving the distribution function of the bi-variate (correlated) normal distribution. Laziness may lead to price it by Monte Carlo simulation (a *PDE* approach is also appropriate but needs more care) as detailed below.

3. *Exchange Call Spread* with strike price K:

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

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

We will illustrate below the regular Monte Carlo procedure on the example of a *Best-of-Call* which is traded on an organized market, unlike its cousin the *Exchange Call Spread*.

Pricing a *best-of-call* by a monte carlo simulation

To implement a (crude) Monte Carlo simulation we need to write the payoff as a function of independent uniformly distributed random variables, or, equivalently, as a tractable function of such random variables. In our case, we write it as a function of two standard normal variables, *i.e.* a bi-variate standard normal distribution (Z^1, Z^2), namely

$$e^{-r\cancel{t}}h_T \stackrel{d}{=} \varphi(Z^1, Z^2)$$

$$:= \left(\max\left(x_0^1 \exp\left(-\frac{\sigma_1^2}{2}T + \sigma_1\sqrt{T}Z^1\right), x_0^2 \exp\left(-\frac{\sigma_2^2}{2}T + \sigma_2\sqrt{T}\big(\rho Z^1 + \sqrt{1-\rho^2}Z^2\big)\right)\right) - Ke^{-rt}\right)_+$$

/T

(e^{-rT})

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–Muller method yields the estimate

$$\begin{aligned} Best\text{-}of\text{-}Call_0 &= e^{-r\cancel{t}}\mathbb{E}\Big(\big(\max(X_T^1, X_T^2) - K\big)_+\Big) \\ &= \mathbb{E}\,\varphi(Z^1, Z^2) \\ &\simeq \overline{\varphi}_M := \frac{1}{M}\sum_{m=1}^{M}\varphi(Z_m). \end{aligned}$$

/T

(e^{-rT})

One computes an estimate for the variance using the same sample

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

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

1. *Application to the computation of the γ (i.e. $\Phi''(x)$).* Show that, if φ is differentiable with a derivative having polynomial growth,

$$\Phi''(x) := \frac{1}{x^2\sigma T}\mathbb{E}\Big(\big(\varphi'(X_T^x)X_T^x - \varphi(X_T^x)\big)W_T\Big)$$

and that, if φ is continuous with compact support,

$$\Phi''(x) := \frac{1}{x^2\sigma T}\mathbb{E}\left(\varphi(X_T^x)\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\right).$$

Extend this identity to the case where φ is simply Borel with polynomial growth. Note that a (somewhat simpler) formula also exists when the function φ is itself twice differentiable, but such a smoothness assumption is not realistic, at least for financial applications.

2. *Variance reduction for the δ* ([7]). The above formulas are clearly not the unique representations of the δ as an expectation: using that $\mathbb{E}\,W_T = 0$ and $\mathbb{E}\,X_T^x = xe^{rT}$, one derives immediately that

$$\Phi'(x) = \varphi'(xe^{rT})e^{rT} + \mathbb{E}\left(\big(\varphi'(X_T^x) - \varphi'(xe^{rT})\big)\frac{X_T^x}{x}\right)$$

as soon as φ is differentiable at xe^{rt}. When φ is simply Borel

$$\Phi'(x) = \frac{1}{x\,\sigma\,T}\mathbb{E}\Big(\big(\varphi(X_T^x) - \varphi(xe^{rT})\big)W_T\Big).$$

3. *Variance reduction for the γ.* Show that

$$\Phi''(x) = \frac{1}{x^2\sigma T}\mathbb{E}\Big(\big(\varphi'(X_T^x)X_T^x - \varphi(X_T^x) - xe^{rT}\varphi'(xe^{rT}) + \varphi(xe^{rT})\big)W_T\Big).$$

4. *Testing the variance reduction, if any.* Although the former two exercises are entitled "variance reduction" the above formulas do not guarantee a variance reduction at a fixed time T. It seems intuitive that they do only when the maturity T is small. Perform some numerical experiments to test whether or not the above formulas induce some variance reduction.

As the maturity increases, test whether or not the regression method introduced in Sect. 3.2 works with these "control variates".

5. *Computation of the vega* ([8]). Show likewise that $\mathbb{E}\,\varphi(X_T^x)$ is differentiable with respect to the volatility parameter σ under the same assumptions on φ, namely

[7] In this exercise we slightly anticipate the next chapter, which is entirely devoted to variance reduction.

[8] Which is not a greek letter…

[illegible] de cette page les équations

$$\frac{\partial}{\partial\sigma}\mathbb{E}\,\varphi(X_T^x) = \mathbb{E}\Big(\varphi'(X_T^x)X_T^x\big(W_T - \sigma T\big)\Big)$$

ferentiable with a derivative having polynomial growth. Derive without any mputations – but with the help of the previous exercises – that

$$\frac{\partial}{\partial\sigma}\mathbb{E}\,\varphi(X_T^x) = \mathbb{E}\left(\varphi(X_T^x)\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\right)$$

ply Borel with polynomial growth. [Hint: use the former exercises.] erivative is known (up to an appropriate discounting) as the *vega* of the ated to the payoff $\varphi(X_T^x)$. Note that the γ and the *vega* of a Call satisfy ounting by e^{-rt})

$$vega(x, K, r, \sigma, T) = x^2\sigma T\gamma(x, K, r, \sigma, T),$$

e key of the tracking error formula.
the beginning of this section can be seen as an introduction to the so-called cess method (see Sect. 2.2.4 at the end of this chapter and Sect. 10.2.2).

irect Differentiation on the State Space: The g-Likelihood Method

formulas established in the former section for the Black–Scholes model ined by working directly on the state space $(0, +\infty)$, taking advantage of X_T^x has a smooth and explicit probability density $p_T(x, y)$ with respect to e measure on $(0, +\infty)$, *which is known explicitly* since it is a log-normal .

bability density also depends on the other parameters of the model like y, σ, the interest rate r and the maturity T. Let us denote by θ one of eters which is assumed to lie in a parameter set Θ. More generally, nagine that $X_T^x(\theta)$ is an $\mathbb{R}^d$-valued solution at time T to a stochastic equation whose coefficients $b(\theta, x)$ and $\sigma(\theta, x)$ depend on a parameter An important result of stochastic analysis for Brownian diffusions is that, rm ellipticity assumptions (or the less stringent "parabolic Hörmander sumptions", see [24, 139]), combined with smoothness assumptions on t and the diffusion coefficient, such a solution of an *SDE* does have a ity $p_T(\theta, x, y)$ – at least in (x, y) – with respect to the Lebesgue measure nore details, we refer to [25] or [11, 98]. Formally, we then get

$$\Phi(\theta) = \mathbb{E}\,\varphi(X_T^x(\theta)) = \int_{\mathbb{R}^d}\varphi(y)p_T(\theta, x, y)\mu(dy)$$

395

partir de cette page les equations

ın be viewed as ancestors of Malliavin calculus, provide the δ-hedge for ptions in local volatility models (see Theorem 10.2 and the application that

ises. 1. Provide simple assumptions to justify the above formal computations at some point θ_0 or for all θ running over a non-empty open interval Θ of ımain of $\mathbb{R}^d$ if θ is vector valued). [Hint: use the remark directly below ı 2.2.]

ute the probability density $p_T(\sigma, x, y)$ of $X_T^{x,\sigma}$ in a Black–Scholes model tands for the volatility parameter).

tablish all the sensitivity formulas established in the former Sect. 2.2.2 g the exercises at the end of the section) using this approach.

these formulas to the case $\varphi(x) := e^{-r}(x - K)_+$ and retrieve the classical ıns for the greeks in a Black–Scholes model: the δ, the γ and the *vega*. /T

; section we focused on the case of the marginal $X_T^x(\theta)$ at time T of a Brow-ısion as encountered in local volatility models viewed as a generalization ıck–Scholes models investigated in the former section. In fact, this method, ; the log-likelihood method, has a much wider range of application since it r any family $\big(X(\theta)\big)_{\theta\in\Theta}$ of $\mathbb{R}^d$-valued vectors, ($\Theta \subset \mathbb{R}^q$) such that, for every e distribution of $X(\theta)$ has a probability density $p(\theta, y)$ with respect to a : measure μ on $\mathbb{R}^d$, usually the Lebesgue measure.

The Tangent Process Method

hen both the payoff function/functional and the coefficients of the *SDE* are ıough, one can differentiate the function/functional of the process directly ect to a given parameter. The former Sect. 2.2.2 was a special case of this or vanilla payoffs in a Black–Scholes model. We refer to Sect. 10.2.2 for ıiled developments.

Var(X)

where $\overline{V}_M \simeq \sqrt{\mathrm{Var}(X)}$ is an estimator of the variance (see (2.1)) and $\Phi_0(x) = \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} \frac{d\xi}{\sqrt{2\pi}}$ is the c.d.f. of the normal distribution.

In numerical probability, we adopt the following *reverse point of view* based on the *target* or *prescribed accuracy* $\varepsilon > 0$: to make $\overline{X}_M$ enter a confidence interval $[m - \varepsilon, m + \varepsilon]$ with a confidence level $\alpha := 2\Phi_0(q_\alpha) - 1$, one needs to perform a Monte Carlo simulation of size

$$M \geq M^X(\varepsilon, \alpha) = \frac{q_\alpha^2 \mathrm{Var}(X)}{\varepsilon^2}. \tag{3.1}$$

In practice, of course, the estimator $\overline{V}_M$ is computed on-line to estimate the variance $\mathrm{Var}(X)$ as presented in the previous chapter. This estimate need not to be as sharp as the estimation of m, so it can be processed at the beginning of the simulation on a smaller sample size.

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

Variance reduction: (not so) naive approach

Assume now that we know *two* random variables $X,\ X' \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ satisfying

$$m = \mathbb{E}\,X = \mathbb{E}\,X' \in \mathbb{R}, \quad \mathrm{Var}(X),\ \mathrm{Var}(X'),\ \mathrm{Var}(X - X') = \mathbb{E}\,(X - X')^2 > 0$$

(the last condition only says that X and X' are not $a.s.$ equal).

Question: Which random vector (distribution…) is more appropriate?

... already been pointed out in the previous

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

Then, the random variable

$$X' = X -$$

can be simulated at the same cost as X,

$$\mathbb{E}\,X' = \mathbb{E}\,X = m \quad \text{and} \quad \mathrm{Var}($$

Definition 3.1 *A random variable* Ξ *sa* variate *for* X.

▶ **Exercise.** Show that if the simulation κ and κ' respectively, then (iii) becomes

$$(iii)' \quad \kappa' \mathrm{Var}(X -$$

The product of the variance of a random called the *effort*. It will be a central not Multilevel methods in Chap. 9.

Toy-example. In the previous chapter, w neutral Black–Scholes model $X_t^x = x\,e^{(r}$ payoff function is differentiable outside tinuous with polynomial growth at infin differentiable on $(0, +\infty)$ and

395 partir de cette page (b) et du (c) est ... rapport à ma version

$$X_t^{i,x_i} = x_i \exp\left(\left(r - \frac{\sigma_{i.}^2}{2}\right)t + \sum_{j=1}^{q} \sigma_{ij} W_t^j\right), \quad t \in [0, T], \quad x_i > 0,\ i = 1, \ldots, d,$$

where

$$\sigma_{i.}^2 = \sum_{j=1}^{q} \sigma_{ij}^2, \quad i = 1, \ldots, d.$$

▶ **Exercise.** Show that if the matrix $\sigma\sigma^*$ is positive definite (then $q \geq d$) and one may assume without modifying the model that X^{i,x_i} only depends on the first i components of a d-dimensional standard Brownian motion. [Hint: consider the Cholesky decomposition in Sect. 1.6.2.]

Now, let us describe the two phases of the variance reduction procedure:

– PHASE I: $\Xi = e^{-\cancel{rt}} k_T$ *as a pseudo-control variate and computation of its expectation* $\mathbb{E}\,\Xi$. / rT

The vanilla Call option has a closed form in a Black–Scholes model and elementary computations show that e^{-rT}

$$\sum_{1\leq i\leq d} \alpha_i \log(X_T^{i,x_i}/x_i) \stackrel{d}{=} \mathcal{N}\left(\left(r - \frac{1}{2}\sum_{1\leq i\leq d} \alpha_i \sigma_{i.}^2\right)T;\ \alpha^*\sigma\sigma^*\alpha\, T\right)$$

where α is the column vector with components α_i, $i = 1, \ldots, d$.

Consequently, the premium at the origin $e^{-rt}\mathbb{E}\,k_T$ admits a closed form (see Sect. 12.2 in the Miscellany Chapter) given by

$$e^{-\cancel{rt}}\mathbb{E}\,k_T = \text{Call}_{BS}\left(\prod_{i=1}^{d} x_i^{\alpha_i} e^{-\frac{1}{2}\left(\sum_{1\leq i\leq d}\alpha_i\sigma_{i.}^2 - \alpha^*\sigma\sigma^*\alpha\right)T},\ K, r, \sqrt{\alpha^*\sigma\sigma^*\alpha},\ T\right).$$

rT (e^{-rT})

– PHASE II: *Joint simulation of the pair* (h_T, k_T).

We need to simulate M independent copies of the pair (h_T, k_T) or, to be more precise of the quantity

$$e^{-rt}(h_T - k_T) = e^{-rt}\left(\left(\sum_{i=1}^{d}\alpha_i X_T^{i,x_i} - K\right)_+ - \left(e^{\sum_{1\leq i\leq d}\alpha_i \log(X_T^{i,x_i})} - K\right)_+\right).$$

e^{-rT} (twice)

This task clearly amounts to simulating M independent copies, of the q-dimensional standard Brownian motion W at time T, namely

$$W_T^{(m)} = (W_T^{1,(m)}, \ldots, W_T^{q,(m)}),\ m = 1, \ldots, M,$$

i.e. M independent copies $Z^{(m)} = (Z_1^{(m)}, \ldots, Z_q^{(m)})$ of the $\mathcal{N}(0; I_q)$ distribution in order to set $W_T^{(m)} \stackrel{d}{=} \sqrt{T} Z^{(m)}$, $m = 1, \ldots, M$.

The resulting pointwise estimator of the premium is given, with obvious notations, by

$$\frac{e^{-rt}}{M} \sum_{m=1}^{M} (h_T^{(m)} - k_T^{(m)}) + \mathrm{Call}_{BS}\left(\prod_{i=1}^{d} x_i^{\alpha_i} e^{-\frac{1}{2}\left(\sum_{1\le i\le d} \alpha_i \sigma_{i.}^2 - \alpha^* \sigma\sigma^* \alpha\right)T}, K, r, \sqrt{\alpha^*\sigma\sigma^*\alpha}, T\right).$$

Remark. The extension to more general payoffs of the form $\varphi\Big(\sum_{1\le i\le d} \alpha_i X_T^{i,x_i}\Big)$ is straightforward provided φ is non-decreasing and a closed form exists for the vanilla option with payoff $\varphi\Big(e^{\sum_{1\le i\le d} \alpha_i \log(X_T^{i,x_i})}\Big)$.

▶ **Exercise.** Other ways to take advantage of the convexity of the exponential function can be explored: thus one can start from

$$\sum_{1\le i\le d} \alpha_i X_T^{i,x_i} = \left(\sum_{1\le i\le d} \alpha_i x_i\right) \sum_{1\le i\le d} \widetilde{\alpha}_i \frac{X_T^{i,x_i}}{x_i},$$

where $\widetilde{\alpha}_i = \dfrac{\alpha_i x_i}{\sum_{1\le k\le d} \alpha_k x_k}$, $i = 1, \ldots, d$. Compare on simulations the respective performances of these different approaches.

2. Asian options and the Kemna–Vorst control variate in a Black–Scholes model (see [166]). Let

$$h_T = \varphi\left(\frac{1}{T}\int_0^T X_t^x dt\right)$$

be a generic *Asian* payoff where φ is non-negative, non-decreasing function defined on $\mathbb{R}_+$ and let

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

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

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

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

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

This suggests to rewrite the right-hand side of the above inequality in a "Black–Scholes asset" style, namely:

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

This naturally leads us to introduce the so-called *Kemna–Vorst* (pseudo-)control variate

$$k_T^{KV} := \varphi\left(x\,e^{-(\frac{r}{2}+\frac{\sigma^2}{12})T}\exp\left(\left(r-\frac{1}{2}\frac{\sigma^2}{3}\right)T + \sigma\frac{1}{T}\int_0^T W_t\,dt\right)\right)$$

which is clearly of Black–Scholes type and moreover satisfies

$$h_T \geq k_T^{KV}.$$

– PHASE I: The random variable k_T^{KV} is an admissible control variate *as soon as the vanilla option related to the payoff* $\varphi(X_T^x)$ *has a closed form*. Indeed, if a vanilla option related to the payoff $\varphi(X_T^x)$ has a closed form

$$e^{-r\cancel{T}}\mathbb{E}\,\varphi(X_T^x) = \text{Premium}_{BS}^{\varphi}(x,r,\sigma,T), \qquad rT \quad e^{-rT}$$

then, one has

$$e^{-r\cancel{T}}\mathbb{E}\,k_T^{KV} = \text{Premium}_{BS}^{\varphi}\left(x\,e^{-(\frac{r}{2}+\frac{\sigma^2}{12})T}, r, \frac{\sigma}{\sqrt{3}}, T\right).$$

– PHASE II: One has to simulate independent copies of $h_T - k_T^{KV}$, *i.e.* in practice, independent copies of the pair (h_T, k_T^{KV}). Theoretically speaking, this requires us to know how to simulate paths of the standard Brownian motion $(W_t)_{t\in[0,T]}$ exactly and, moreover, to compute with an infinite accuracy integrals of the form $\frac{1}{T}\int_0^T f(t)dt$.

In practice these two tasks are clearly impossible (one cannot even compute a real-valued function $f(t)$ *at every* $t \in [0,T]$ with a computer). In fact, one relies on quadrature formulas to approximate the time integrals in both payoffs which makes this simulation possible since only finitely many random marginals of the Brownian motion, say $W_{t_1},\ldots,W_{t_n}$, are necessary, which is then quite realistic. Typically, one uses a mid-point quadrature formula

3.3 Application to Option Pricing: Using Parity Equations to Produce Control Variates

The variance reduction by regression introduced in the former section still relies on the fact that $\kappa_X \simeq \kappa_{X-\lambda\Xi}$ or, equivalently, that the additional complexity induced by the simulation of Ξ given that of X is negligible. This condition may look demanding but we will see that in the framework of derivative pricing this requirement is always fulfilled as soon as the payoff of interest satisfies a so-called *parity equation*, *i.e.* that the original payoff can be *duplicated* by a "synthetic" version.

Furthermore, these *parity equations* are *model free* so they can be applied for various specifications of the dynamics of the underlying asset.

In this section, we denote by $(S_t)_{t\geq 0}$ the risky asset (with $S_0 = s_0 > 0$) and set $S_t^0 = e^{rt}$, the riskless asset. We work under the risk-neutral *risk-neutral* probability $\mathbb{P}$ (supposed to exist), which means that

$$\left(e^{-rt}S_t\right)_{t\in[0,T]} \quad \text{is a martingale on the scenarii space } (\Omega, \mathcal{A}, \mathbb{P})$$

(with respect to the augmented filtration of $(S_t)_{t\in[0,T]}$). Furthermore, to comply with usual assumptions of AOA theory, we will assume that this risk neutral probability is unique (complete market) to justify that we may price any derivative under this probability. However this has no real impact on what follows.

Vanilla Call-Put parity ($d = 1$) $/t \to T$

We consider a Call and a Put with common maturity T and strike K. We denote by

$$\text{Call}_0(K,T) = e^{-r\cancel{t}}\mathbb{E}\left((S_T - K)_+\right) \qquad \text{and} \qquad \text{Put}_0(K,T) = e^{-r\cancel{t}}\mathbb{E}\left((K - S_T)_+\right)$$

the premium of this Call and this Put option, respectively. Since

$$(S_T - K)_+ - (K - S_T)_+ = S_T - K$$

and $\left(e^{-rt}S_t\right)_{t\in[0,T]}$ is a martingale, one derives the classical Call-Put parity equation:

$$\text{Call}_0(K,T) - \text{Put}_0(K,T) = s_0 - e^{-r\cancel{t}}K$$

so that $\text{Call}_0(K,T) = \mathbb{E}(X) = \mathbb{E}(X')$ with

$$X := e^{-r\cancel{t}}(S_T - K)_+ \quad \text{and} \quad X' := e^{-r\cancel{t}}(K - S_T)_+ + s_0 - e^{-r\cancel{t}}K.$$

As a result one sets

$$\Xi = X - X' = e^{-r\cancel{t}}S_T - s_0,$$

which turns out to be the terminal value of a martingale null at time 0 (this is in fact the generic situation of application of this parity method).

ote that the simulation of X involves that of S_T so that the additional cost of the ılation of Ξ is definitely negligible.

ın Call-Put parity

consider an Asian Call and an Asian Put with common maturity T, strike K and ·aging period $[T_0, T]$, $0 \leq T_0 < T$.

$$\mathrm{Call}_0^{As} = e^{-rt}\mathbb{E}\left[\left(\tfrac{1}{T-T_0}\int_{T_0}^{T} S_t\,dt - K\right)_+\right]$$

$e^{rt} \to e^{rT}$

6 times

$$\mathrm{Put}_0^{As} = e^{-rt}\mathbb{E}\left[\left(K - \tfrac{1}{T-T_0}\int_{T_0}^{T} S_t\,dt\right)_+\right].$$

Still using that $\widetilde{S}_t = e^{-rt}S_t$ is a $\mathbb{P}$-martingale and, this time, the Fubini–Tonelli eorem yield

$$\mathrm{Call}_0^{As} - \mathrm{Put}_0^{As} = s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rt}K$$

that

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

ith

$$X := e^{-rt}\left(\frac{1}{T-T_0}\int_{T_0}^{T} S_t\,dt - K\right)_+$$

$$X' := s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rt}K + e^{-rt}\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} S_t\,dt\right)_+.$$

his leads to

$$\Xi = e^{-rt}\frac{1}{T-T_0}\int_{T_0}^{T} S_t dt - s_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)}.$$

Remark. In both cases, the parity equation directly follows from the $\mathbb{P}$*-martingale property* of $\widetilde{S}_t = e^{-rt}S_t$.

3.3.1 *Complexity Aspects in the General Case*

In practical implementations, one often neglects the cost of the computation of $\lambda_{\min}$ since only a rough estimate is computed: this leads us to stop its computation after the first 5% or 10% of the simulation.

– However, one must be aware that the case of the existence of parity equations is quite specific since the random variable Ξ is involved in the simulation of X, so the complexity of the simulation process *is not* increased: thus in the recursive approach the updating of λ_M and of (the empirical mean) $\overline{X}_M$ is (almost) costless. Similar observations can be made to some extent on batch approaches. As a consequence, in that specific setting, the complexity of the adaptive linear regression procedure and the original one are (almost) the same!

– **Warning!** This is no longer true in general...and in a general setting the complexity of the simulation of X and X' is double that of X itself. Then the regression method is efficient if and only if

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

(provided one neglects the cost of the estimation of the coefficient $\lambda_{\min}$).

The exercise below shows the connection with antithetic variables which then appears as a special case of regression methods.

▶ **Exercise** (*Connection with the antithetic variable method*). Let $X, X' \in L^2(\mathbb{P})$ such that $\mathbb{E}\,X = \mathbb{E}\,X' = m$ and $\mathrm{Var}(X) = \mathrm{Var}(X')$.

(a) Show that $\lambda_{\min} = \frac{1}{2}$.

(b) Show that $X^{\lambda_{\min}} = \dfrac{X + X'}{2}$ and $\mathrm{Var}\left(\dfrac{X+X'}{2}\right) = \dfrac{1}{2}\big(\mathrm{Var}(X) + \mathrm{Cov}(X, X')\big)$.

Characterize the pairs (X, X') for which the regression method does reduce the variance. Make the connection with the antithetic method.

3.3.2 *Examples of Numerical Simulations*

Vanilla B-S **Calls** (See Figs. 3.2, 3.3 and 3.4)

The model parameters are specified as follows

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

The simulation size is set at $M = 10^6$.

$r = 5\%,\ \sigma = 20\%$

ls in a Heston model (See Figs. 3.5, 3.6 **and** 3.7)

nics of the risky asset is this time a stochastic volatility model, namely ı model, defined as follows. Let ϑ, k, a such that $\vartheta^2/(2ak) \leq 1$ (so that v_t .s. positive, see [183], Proposition 6.2.4, p. 130).

$$= S_t\big(r\,dt + \sqrt{v_t}\,dW_t^1\big), \qquad s_0 = x_0 > 0,\ t \in [0, T], \quad \text{(risky asset)}$$
$$= k(a - v_t)dt + \vartheta\sqrt{v_t}\,dW_t^2,\ v_0 > 0$$

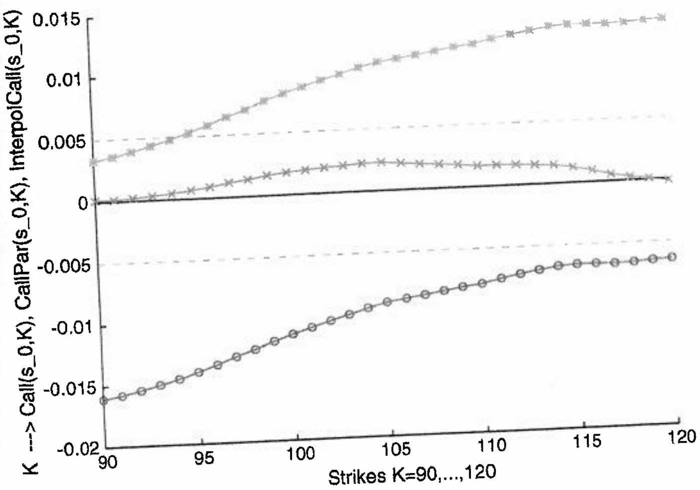
h $\langle W^1, W^2\rangle_t = \rho t,\ \rho \in [-1, 1],\ t \in [0, T]$.

ıff is an *Asian call* with strike price K

$$\mathrm{AsCall}^{Hest} = e^{-r}\mathbb{E}\left[\left(\frac{1}{T}\int_0^T S_s ds - K\right)_+\right].$$

/T (e^{-rT})

lly, no closed forms are available for Asian payoffs, even in the Black– model, and this is also the case in the Heston model. Note however that closed forms do exist for vanilla European options in this model (see [150]), s the origin of its success. The simulation has been carried out by replacing ve diffusion by an Euler scheme (see Chap. 7 for an introduction to the Euler ıcretization scheme). In fact, the dynamics of the stochastic volatility process ıt fulfill the standard Lipschitz continuous assumptions required to make the :heme converge, at least at its usual rate. In the present case it is even difficult ıe this scheme because of the term $\sqrt{v_t}$. Since our purpose here is to illustrate



ı.2 Black–Scholes Calls: Error = Reference BS-(MC Premium). $K = 90, \ldots, 120$. -o- Crude Call. -*-*-*- Synthetic Parity Call. -×××- Interpolated Synthetic Call

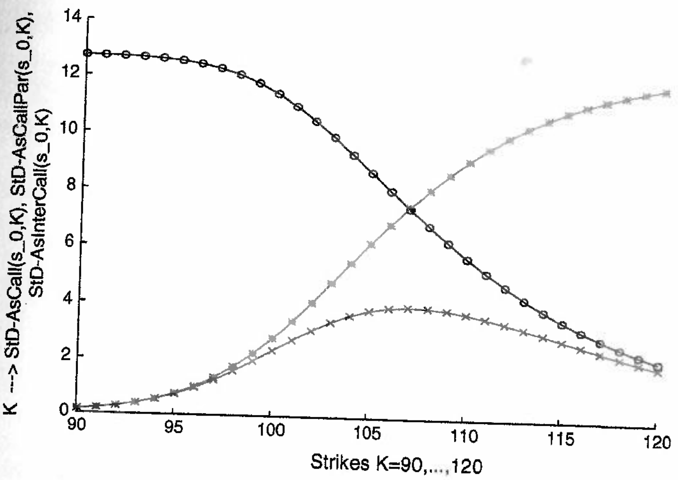


Fig. 3.5 HESTON ASIAN CALLS. *Standard Deviation (MC Premium).* $K = 90, \ldots, 120$. *-o-o-o- Crude Call. -*-*-*- Synthetic Parity Call. -xxx- Interpolated Synthetic Call*

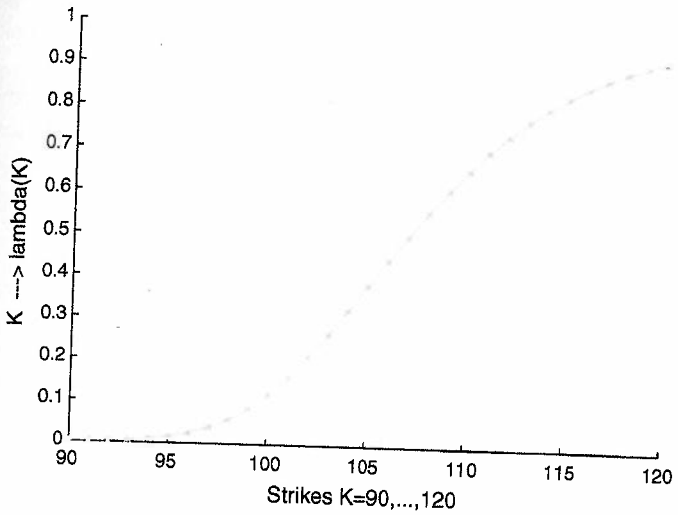


Fig. 3.6 HESTON ASIAN CALLS. $K \mapsto 1 - \lambda_{\min}(K)$, $K = 90, \ldots, 120$, *for the Interpolated Synthetic Asian Call*

$$
\begin{aligned}
\bar{S}_{\frac{kT}{n}} &= \bar{S}_{\frac{(k-1)T}{n}}\Big(1 + \frac{rT}{n} + \sqrt{|\bar{v}_{\frac{(k-1)T}{n}}|}\sqrt{\frac{T}{n}}\big(\rho Z_k^2 + \sqrt{1-\rho^2}Z_k^1\big)\Big),\\
\bar{S}_0 &= s_0 > 0,\\
\bar{v}_{\frac{kT}{n}} &= k\big(a - \bar{v}_{\frac{(k-1)T}{n}}\big)\frac{T}{n} + \vartheta\sqrt{|\bar{v}_{\frac{(k-1)T}{n}}|}\,Z_k^2, \quad \bar{v}_0 = v_0 > 0,
\end{aligned}
$$

$\wedge\ \bar{v}_{\frac{(k-1)T}{n}} +$

$= \bar{v}_{\frac{(k-1)T}{n}} + k(a - \ldots)\ldots$

rapport à ma version.

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

Then one can write

$$(W_T^1, W_T^2) = \sqrt{T}\big(\sqrt{1-\rho^2}Z_1 + \rho Z_2, Z_2\big),$$

where $Z = (Z_1, Z_2)$ is an $\mathcal{N}(0; I_2)$-distributed random vector. Then, see *e.g.* Sect. 12.2 in the Miscellany Chapter),

$$\begin{aligned} e^{-r\cancel{t}}\mathbb{E}\,(h_T \mid Z_2) &= e^{-r\cancel{t}}\Big[\mathbb{E}\Big(\Big(x_1 e^{(r-\frac{\sigma_1^2}{2})T+\sigma_1\sqrt{T}(\sqrt{1-\rho^2}\,Z_1+\rho z_2)} \\ &\qquad -x_2 e^{(r-\frac{\sigma_2^2}{2})T+\sigma_2\sqrt{T}z_2} - K\Big)_+\Big)\Big]_{|z_2=Z_2} \\ &= \mathrm{Call}_{BS}\Big(x_1 e^{-\frac{\rho^2\sigma_1^2 T}{2}+\sigma_1\rho\sqrt{T}Z_2}, x_2 e^{(r-\frac{\sigma_2^2}{2})T+\sigma_2\sqrt{T}Z_2} \\ &\qquad +K, r, \sigma_1\sqrt{1-\rho^2}, T\Big). \end{aligned}$$

Finally, one takes advantage of the closed form available for vanilla Call options in a Black–Scholes model to compute

$$\mathrm{Premium}_{BS}(x_1, x_2, K, \sigma_1, \sigma_2, r, T) = \mathbb{E}\Big(\mathbb{E}\,(e^{-rt}h_T \mid Z_2)\Big)$$

with a smaller variance than with the original payoff.

2. *Barrier options*. This example will be detailed in Sect. 8.2.3 devoted to the pricing (of some classes) of barrier options in a general model using the simulation of a continuous Euler scheme (using the so-called Brownian bridge method).

3.5 Stratified Sampling

The starting idea of stratification is to localize the Monte Carlo method on the elements of a measurable partition of the state space E of a random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$.

Let $(A_i)_{i\in I}$ be a finite $\mathcal{E}$-measurable partition of the state space E. The A_i's are called *strata* and $(A_i)_{i\in I}$ a *stratification* of E. Assume that the weights

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

are *known*, (strictly) *positive* and that, still for every $i \in I$,

$$\mathcal{L}(X \mid X \in A_i) \stackrel{d}{=} \varphi_i(U),$$

Of course there is no reason why the solution to the above problem should be θ_0 (if so, such a parametric model is inappropriate). At this stage one can follow two strategies:

– Try to solve by numerical means the above minimization problem.

– Use one's intuition to select *a priori* a good (though sub-optimal) $\theta \in \Theta$ by applying the heuristic principle: "focus light where needed".

$\triangleright$ **Example** *(The Cameron–Martin formula and Importance Sampling by mean translation).* This example takes place in a Gaussian framework. We consider (as a starting motivation) a one dimensional Black–Scholes model defined by

$$X_T^x = x e^{\mu T + \sigma W_T} = x e^{\mu T + \sigma\sqrt{T}Z}, \quad Z \stackrel{d}{=} \mathcal{N}(0;1),$$

with $x > 0$, $\sigma > 0$ and $\mu = r - \frac{\sigma^2}{2}$. Then, the premium of an option with payoff $h : (0, +\infty) \to (0, +\infty)$ reads

$$e^{-rT}\mathbb{E}\, h(X_T^x) = \mathbb{E}\,\varphi(Z) = \int_{\mathbb{R}} \varphi(z) e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}},$$

where $\varphi(z) = e^{-rT} h\left(x e^{\mu T + \sigma\sqrt{T}z}\right)$, $z \in \mathbb{R}$.

From now on, we forget about the financial framework and deal with

$$\mathbb{E}\,\varphi(Z) = \int_{\mathbb{R}} \varphi(z) g_0(z) dz \quad \text{where} \quad g_0(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$

and the random variable Z plays the role of X in the above theoretical part. The idea is to introduce the parametric family

$$Y_\theta = Z + \theta, \quad \theta \in \Theta := \mathbb{R}.$$

We consider the Lebesgue measure on the real line λ_1 as a reference measure, so that

$$g_\theta(y) = \frac{e^{-\frac{(y-\theta)^2}{2}}}{\sqrt{2\pi}}, \quad y \in \mathbb{R}, \ \theta \in \Theta := \mathbb{R}.$$

Elementary computations show that

$$\frac{g_0}{g_\theta}(y) = e^{-\theta y + \frac{\theta^2}{2}}, \quad y \in \mathbb{R}, \ \theta \in \Theta := \mathbb{R}.$$

Hence, we derive the *Cameron–Martin formula*

$$\begin{aligned}\mathbb{E}\,\varphi(Z) &= e^{\frac{\theta^2}{2}}\mathbb{E}\Big(\varphi(Y_\theta)e^{-\theta Y_\theta}\Big)\\ &= e^{\frac{\theta^2}{2}}\mathbb{E}\Big(\varphi(Z+\theta)e^{-\theta(Z+\theta)}\Big) = e^{-\frac{\theta^2}{2}}\mathbb{E}\Big(\varphi(Z+\theta)e^{-\theta Z}\Big).\end{aligned}$$

Remark. In fact, a *standard change of variable* based on the invariance of the Lebesgue measure by translation yields the same result in a much more straightforward way: setting $z = u + \theta$ shows that

$$\begin{aligned}\mathbb{E}\,\varphi(Z) &= \int_{\mathbb{R}} \varphi(u+\theta)e^{-\frac{\theta^2}{2}-\theta u-\frac{u^2}{2}}\frac{du}{\sqrt{2\pi}} = e^{-\frac{\theta^2}{2}}\,\mathbb{E}\left(e^{-\theta Z}\varphi(Z+\theta)\right)\\ &= e^{\frac{\theta^2}{2}}\mathbb{E}\Big(\varphi(Z+\theta)e^{-\theta(Z+\theta)}\Big).\end{aligned}$$

It is to be noticed again that there is no need to account for the normalization constants to compute the ratio $\frac{g_0}{g_\theta}$.

The next step is to choose a "good" θ which significantly reduces the variance, *i.e.* following Condition (3.13) (using the formulation involving "$Y_\theta = Z + \theta$"), such that

$$\mathbb{E}\left(e^{\frac{\theta^2}{2}}\varphi(Z+\theta)e^{-\theta(Z+\theta)}\right)^2 < \mathbb{E}\,\varphi^2(Z),$$

i.e.

$$e^{-\theta^2}\mathbb{E}\Big(\varphi^2(Z+\theta)e^{-2\theta Z}\Big) < \mathbb{E}\,\varphi^2(Z)$$

or, equivalently, if one uses the formulation of (3.13) based on the original random variable (here Z),

$$\mathbb{E}\Big(\varphi^2(Z)e^{\frac{\theta^2}{2}-\theta Z}\Big) < \mathbb{E}\,\varphi^2(Z).$$

Consequently the variance minimization amounts to the following problem

$$\min_{\theta\in\mathbb{R}}\left[e^{\frac{\theta^2}{2}}\mathbb{E}\left(\varphi^2(Z)e^{-\theta Z}\right) = e^{-\theta^2}\mathbb{E}\left(\varphi^2(Z+\theta)e^{-2\theta Z}\right)\right].$$

It is clear that the solution of this optimization problem and the resulting choice of θ highly depends on the function h.

– *Optimization approach*: When h is smooth enough, an approach based on large deviation estimates has been proposed by Glasserman et al. (see [115]). We propose a simple recursive/adaptive approach in Sect. 6.3.1 of Chap. 6 based on Stochastic Approximation which does not depend upon the regularity of the function h (see also [12] for a pioneering work in that direction).

– *Heuristic suboptimal approach*: Let us temporarily return to our pricing problem involving the specified function $\varphi(z) = e^{-rT}\big(x\exp(\mu T + \sigma\sqrt{T}z) - K\big)_+$, $z \in \mathbb{R}$. When $x \ll K$ (deep-out-of-the-money option), most simulations of $\varphi(Z)$ will pro-

Show that the function f defined on $[0,1]^2$ by

$$f(x^1,x^2) := (x^1+x^2)\wedge 1, \quad (x^1,x^2)\in[0,1]^2$$

is finite variation in the measure sense [Hint: consider the distribution of $(U, 1-U)$, $\stackrel{d}{=} \mathcal{U}([0,1])$].

please, skip line

For the class of functions with finite variation, the Koksma–Hlawka Inequal- y provides an error bound for $\frac{1}{n}\sum_{k=1}^n f(\xi_k) - \int_{[0,1]^d} f(x)dx$ based on the star screpancy.

roposition 4.3 (Koksma–Hlawka Inequality (1943 when $d=1$)) *Let $(\xi_1,\ldots,\xi_n)$ an n-tuple of $[0,1]^d$-valued vectors and let $f:[0,1]^d\to\mathbb{R}$ be a function with ite variation (in the measure sense). Then*

$$\left|\frac{1}{n}\sum_{k=1}^n f(\xi_k) - \int_{[0,1]^d} f(x)\lambda_d(dx)\right| \le V(f)D_n^*\big((\xi_1,\ldots,\xi_n)\big).$$

oof. Set $\widetilde{\mu}_n = \frac{1}{n}\sum_{k=1}^n \delta_{\xi_k} - \lambda_{d|[0,1]^d}$. It is a signed measure with 0-mass. Then, if has finite variation with respect to a signed measure ν,

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^n f(\xi_k) - \int_{[0,1]^d} f(x)\lambda_d(dx) &= \int f(x)\widetilde{\mu}_n(dx)\\
&= f(\mathbf{1})\widetilde{\mu}_n([0,1]^d) + \int_{[0,1]^d}\nu([\![0,\mathbf{1}-x]\!])\widetilde{\mu}_n(dx)\\
&= 0 + \int_{[0,1]^d}\left(\int \mathbf{1}_{\{v\le 1-x\}}\nu(dv)\right)\widetilde{\mu}_n(dx)\\
&= \int_{[0,1]^d}\widetilde{\mu}_n([\![0,\mathbf{1}-v]\!])\nu(dv),
\end{aligned}$$

here we used Fubini's Theorem to interchange the integration order (which is ssible since $|\nu|\otimes|\widetilde{\mu}_n|$ is a finite measure). Finally, using the extended triangle equality for integrals with respect to signed measures, This second lower bound can be derived from the first one, using the Hammersley procedure introduced and analyzed in the next section (see the exercise at the end of Sect. 4.3.4).

On the other hand, there exists (see Sect. 4.3.3 that follows) sequences for which

$$\forall n \geq 1, \quad D_n^*(\xi) = C(\xi)\frac{(\log n)^d}{n} \quad \text{where} \quad C(\xi) < +\infty.$$

Based on this, one can derive from the Hammersley procedure (see again Sect. 4.3.4 below) the existence of a real constant $C_d \in (0, +\infty)$ such that $C_d \frac{1+(\log n)^{d-1}}{n}$

$$\forall n \geq 1,\ \exists\, (\xi_1, \ldots, \xi_n) \in \left([0,1]^d\right)^n,\ D_n^*(\xi_1, \ldots, \xi_n) \leq C_d \frac{1+(\log n)^{d-1}}{n}.$$

In spite of more than fifty years of investigation, the gap between these asymptotic lower and upper-bounds have not been significantly reduced: it has still not been proved whether there exists a sequence for which $C(\xi) = 0$, *i.e.* for which the rate $\frac{(\log n)^d}{n}$ would not be optimal.

In fact, it is widely shared in the *QMC* community that in the above lower bounds, $\frac{d-1}{2}$ can be replaced by $d-1$ in (4.9) and $\frac{d}{2}$ by d in (4.10) so that the rate $O\left(\frac{(\log n)^d}{n}\right)$ is commonly considered as the lowest possible rate of convergence to 0 for the star discrepancy of a uniformly distributed sequence. When $d = 1$, Schmidt proved that this conjecture is true.

This leads to a more convincing definition of a *sequence with low discrepancy.*

Definition 4.8 *A $[0,1]^d$-valued sequence $(\xi_n)_{n \geq 1}$ is a sequence with low discrepancy if*

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

For more insight about the other measures of uniform distribution (L^p-discrepancy $D_n^{(p)}(\xi)$, diaphony, etc), we refer *e.g.* to [46, 219].

4.3.3 Examples of Sequences

▷ Van der Corput and Halton sequences

Let $p_1, \ldots, p_d$ be the first d prime numbers. The d-dimensional Halton sequence is defined, for every $n \geq 1$, by:

$$\xi_n = \left(\Phi_{p_1}(n), \ldots, \Phi_{p_d}(n)\right) \tag{4.11}$$

The only natural upper-bound for the left-hand side of this inequality is

$$\begin{aligned}\sigma_n^2(f,\xi) &= \int_{[0,1]^d}\left(\frac{1}{n}\sum_{k=1}^{n} f(\{u+\xi_k\}) - \int_{[0,1]^d} f\,d\lambda_d\right)^2 du\\ &\le \sup_{u\in[0,1]^d}\left|\frac{1}{n}\sum_{k=1}^{n} f(\{u+\xi_k\}) - \int_{[0,1]^d} f\,d\lambda_d\right|^2.\end{aligned}$$

One can show that $f_u : v \mapsto f(\{u+v\})$ has finite variation on $[0,1]^d$ as soon as f has (in the same sense) and that $\sup_{u\in[0,1]^d} V(f_u) < +\infty$ (more precise results can be established). Consequently

$$\sigma_n^2(f,\xi) \le \sup_{u\in[0,1]^d} V(f_u)^2 D_n^*(\xi_1,\ldots,\xi_n)^2$$

so that, if $\xi = (\xi_n)_{n\ge1}$ is a sequence with low discrepancy (say Faure, Halton, Kakutani or Sobol', etc),

× $C_{d,\xi}\,\frac{1+(\log n)^{2d}}{n^{2}}$

$$\sigma_n^2(f,\xi) \le C_{f,\xi}^2\frac{(\log n)^{2d}}{n^2},\ n\ge1.$$

Consequently, in that case, it is clear that randomized *QMC* provides a very significant variance reduction (for the same complexity) of a magnitude proportional to $\frac{(\log n)^{2d}}{n}$ (with an impact of magnitude $\frac{(\log n)^d}{\sqrt{n}}$ on the confidence interval). But one must bear in mind once again that such functions with finite variations become dramatically sparse among Riemann integrable functions as d increases.

In fact, an even better bound of the form $\sigma_n^2(f,\xi) \le \frac{C_{f,\xi}^2}{n^2}$ can be obtained for some classes of functions as emphasized in the Pros part of Sect. 4.3.5: when the sequence $(\xi_n)_{n\ge1}$ is the orbit of a (uniquely) ergodic transform and f is a coboundary for this transform. But of course this class is even sparser. For sequences obtained by iterating rotations – of the torus or of the Kakutani adding machine – some criteria can be obtained involving the rate of decay of the Fourier coefficients $c_p(f)$, $p=(p^1,\ldots,p^d)\in\mathbb{Z}^d$, of f as $\|p\| := p^1\times\cdots\times p^d$ goes to infinity since, in that case, one has $\sigma_n^2(f,\xi) \le \frac{C_{f,\xi}^2}{n^2}$. Hence, the gain in terms of variance becomes proportional to $\frac{1}{n}$ for such functions (a global budget/complexity being prescribed for the simulation).

By contrast, if we consider Lipschitz continuous functions, things go radically differently: assume that $f:[0,1]^d\to\mathbb{R}$ is Lipschitz continuous and isotropically periodic, *i.e.* for every $x\in[0,1]^d$ and every vector $e_i = (\delta_{ij})_{1\le j\le d}$, $i=1,\ldots,d$ of the canonical basis of $\mathbb{R}^d$ (δ_{ij} stands for the Kronecker symbol) $f(x+e_i) = f(x)$ as soon as $x+e_i\in[0,1]^d$, then f can be extended as a Lipschitz continuous function on whole $\mathbb{R}^d$ with the same Lipschitz coefficient, say $[f]_{\text{Lip}}$. Furthermore, it satisfies $(x) = f(\{x\})$ for every $x \in \mathbb{R}^d$. Then, it follows from Proïnov's Theorem 4.3 that

$$\sup_{u\in[0,1]^d}\left|\frac{1}{n}\sum_{k=1}^{n} f(\{u+\xi_k\}) - \int_{[0,1]^d} f\,d\lambda_d\right|^2 \le [f]_{\text{Lip}}^2 D_n^*(\xi_1,\dots,\xi_n)^{\frac{2}{d}}$$

$$\le C_d^2[f]_{\text{Lip}}^2 C_\xi^{\frac{2}{d}} \frac{(\log n)^2}{n^{\frac{2}{d}}}$$

$\frac{1+(\log n)^\ell}{n^{\frac{2}{d}}}$

(where C_d is Proïnov's constant). This time, still for a prescribed budget, the "gain" factor in terms of variance is proportional to $n^{1-\frac{2}{d}}(\log n)^2$, which is no longer a gain... but a loss as soon as $d \ge 2$!

$\frac{1+(\log n)^\ell}{n^{\frac{2}{d}}}$

For more results and details, we refer to the survey [271] on randomized *QMC* and the references therein.

Finally, randomized *QMC* is a specific (and not so easy to handle) variance reduction method, not a *QMC* speeding up method. It suffers from one drawback shared by all *QMC*-based simulation methods: the sparsity of the class of functions with finite variation and the difficulty for identifying them in practice when $d > 1$.

4.4.2 *Scrambled (Randomized) QMC*

If the principle of mixing randomness and the Quasi-Monte Carlo method is undoubtedly a way to improve rates of convergence of numerical integration over unit hypercubes, the approach based on randomly "shifted" sequences with low discrepancy (or nets) described in the former section turned out to be not completely satisfactory and it is no longer considered as the most efficient way to proceed by the *QMC* community.

A new idea emerged at the very end of the 20th century inspired by the pioneering work by A. Owen (see [221]): to break the undesired regularity which appears even in the most popular sequences with low discrepancy (like Sobol' sequences), he proposed to *scramble* them in an i.i.d. random way so that these regularity features disappear while preserving the quality, in terms of discrepancy, of these resulting sequences (or nets).

The underlying principle – or constraint – was to preserve their "geometric-combinatorial" properties. Typically, if a sequence shares the (s, d)-property in a given base (or the (s, m, d)-property for a net), its scrambled version should share it too. Several attempts to produce efficient deterministic scrambling procedures have been made as well, but of course the most radical way to get rid of regularity features was to consider a kind of i.i.d. scrambling as originally developed in [221]. This has been successfully applied to the "best" Sobol' sequences by various authors.

risk neutral probability)

$$S_t^i = s_0^i \exp\left(\left(r - \frac{\sigma_i^2}{2}\right)t + \sigma_i\sqrt{t}\,Z^{i,t}\right), \quad i = 1, \ldots, d,$$

where $Z^{i,t} = W_t^i$, $W = (W^1, \ldots, W^d)$ is a d-dimensional standard Brownian motion. Independence is unrealistic but corresponds to the most unfavorable case for numerical experiments. We also assume that $s_0^i = s_0 > 0$, $i = 1, \ldots, d$, and that the d assets share the same volatility $\sigma_i = \sigma > 0$. One considers the geometric index $I_t = \left(S_t^1 \ldots S_t^d\right)^{\frac{1}{d}}$. One shows that $e^{-\frac{\sigma^2}{2}(\frac{1}{d}-1)t} I_t$ itself has a risk neutral Black–Scholes dynamics. We want to test the *regularized Put Spread option on this geometric index* with strikes $K_1 < K_2$ (at time $T/2$). Let $\psi(s_0, K_1, K_2, r, \sigma, T)$ denote the premium at time 0 of a Put Spread on any of the assets S^i. We have

$$\begin{aligned}
\psi(x, K_1, K_2, r, \sigma, T) &= \pi(x, K_2, r, \sigma, T) - \pi(x, K_1, r, \sigma, T), \\
\pi(x, K, r, \sigma, T) &= K e^{-rt}\Phi_0(-d_2) - x\,\Phi_0(-d_1), \\
d_1 &= \frac{\log(x/K) + (r + \frac{\sigma^2}{2d})T}{\sigma\sqrt{T/d}}, \quad d_2 = d_1 - \sigma\sqrt{T/d}.
\end{aligned}$$

/rT

(e^{-rT})

Using the martingale property of the discounted value of the premium of a European option yields that the premium $e^{-rt}\mathbb{E}\left((K_1 - I_T)_+ - (K_2 - I_T)_+\right)$ of the Put Spread option on the index I satisfies on the one hand

$$e^{-rt}\mathbb{E}\left((K_1 - I_T)_+ - (K_2 - I_T)_+\right) = \psi\left(s_0 e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)T}, K_1, K_2, r, \sigma/\sqrt{d}, T\right)$$

/rT

and, one the other hand,

$$e^{-rt}\mathbb{E}\left((K_1 - I_T)_+ - (K_2 - I_T)_+\right) = \mathbb{E}\,g(Z),$$

/T

where

$$g(Z) = e^{-rT/2}\psi\left(e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)\frac{T}{2}} I_{\frac{T}{2}}, K_1, K_2, r, \sigma, T/2\right)$$

and $Z = (Z^{1,\frac{T}{2}}, \ldots, Z^{d,\frac{T}{2}}) \stackrel{d}{=} \mathcal{N}(0; I_d)$. The numerical specifications of the function g are as follows:

$$s_0 = 100, \qquad K_1 = 98, \qquad K_2 = 102, \qquad r = 5\%, \qquad \sigma = 20\%, \qquad T = 2.$$

The results are displayed in Fig. 5.5 in a log-log-scale for the dimensions $d = 4, 6, 8, 10$.

First, we recover theoretical rates (namely $-2/d$) of convergence for the error bounds. Indeed, some slopes $\beta(d)$ can be derived (using a regression) for the quantization errors and we found $\beta(4) = -0.48$, $\beta(6) = -0.33$, $\beta(8) = -0.25$ and $\beta(10) = -0.23$ for $d = 10$ (see Fig. 5.5). These rates plead for the implementation of the Richardson–Romberg extrapolation. Also note that, as already reported in [231], ow that quantization-based stratification has a uniform efficiency among the class of pschitz continuous functions. The first step is the universal stratified sampling for pschitz continuous functions detailed in the simple proposition below, where we use e notations introduced in Sect. 3.5. Also keep in mind that for a random vector $Y \in L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$, $\|Y\|_2 = \left(\mathbb{E}\,|Y|^2\right)^{1/2}$ where $|\cdot|$ denotes the canonical Euclidean norm.

roposition 5.4 (Universal stratification) *Let $X \in L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ and let $(A_i)_{i\in I}$ be stratification of $\mathbb{R}^d$. For every $i \in I$, we define the local inertia of the random vector on the stratum A_i by*

$$\sigma_i^2 = \mathbb{E}\left(|X - \mathbb{E}\,(X\,|X \in A_i)|^2\,|X \in A_i\right).$$

) *Then, for every Lipschitz continuous function $F : (\mathbb{R}^d, |\cdot|) \to (\mathbb{R}^d, |\cdot|)$,*

$$\forall\, i \in I, \quad \sup_{[F]_{\mathrm{Lip}}\le 1} \sigma_{F,i} = \sigma_i, \tag{5.33}$$

here $\sigma_{F,i}$ is non-negative and defined by

$$\sigma_{F,i}^2 = \min_{a\in\mathbb{R}^d} \mathbb{E}\left(|F(X) - a|^2\,|X \in A_i\right) = \mathbb{E}\Big(|F(X) - \mathbb{E}\,(\cancel{X}\,|X \in A_i)|^2\,|X \in A_i\Big).$$ F(X)

) Suboptimal choice: $q_i = p_i$.

$$\sup_{[F]_{\mathrm{Lip}}\le 1}\left(\sum_{i\in I} p_i\sigma_{F,i}^2\right) = \sum_{i\in I} p_i\sigma_i^2 = \left\|X - \mathbb{E}\left(X\,|\sigma(\{X \in A_i\},\ i \in I)\right)\right\|_2^2. \tag{5.34}$$

) Optimal choice of the q_i. *(see (3.10) for a closed form of the q_i)*

$$\begin{aligned}\sup_{[F]_{\mathrm{Lip}}\le 1}\left(\sum_{i\in I} p_i\sigma_{F,i}\right)^2 &= \left(\sum_{i\in I} p_i\sigma_i\right)^2 \\ &= \left\|X - \mathbb{E}\left(X\,|\sigma(\{X \in A_i\},\ i \in I)\right)\right\|_1^2.\end{aligned} \tag{5.35}$$

emark. Any real-valued Lipschitz continuous function can be canonically seen as n $\mathbb{R}^d$-valued Lipchitz function, but then the above equalities (5.33)–(5.35) only hold s inequalities.

roof. (a) Note that

$$\begin{aligned}\sigma_{F,i}^2 &= \mathrm{Var}\,(F(X)\,|\,X \in A_i) = \mathbb{E}\left((F(X) - \mathbb{E}\,(F(X)|X \in A_i))^2|\,X \in A_i\right) \\ &\le \mathbb{E}\left((F(X) - F(\mathbb{E}\,(X|X \in A_i)))^2|X \in A_i\right)\end{aligned}$$

At this stage, we assume that $d = 1$. Either $\mathcal{Y}_\infty(\omega) = \{y_*\}$ and the proof is complete, or $\mathcal{Y}_\infty(\omega)$ is a non-trivial compact interval as a compact connected subset of $\mathbb{R}$. The function L is constant on this interval, consequently its derivative L' is zero on $\mathcal{Y}_\infty(\omega)$ so that $\mathcal{Y}_\infty(\omega) \subset \{(\nabla L|h) = 0\} \cap \{L = L(y_*)\}$. Hence the conclusion. When $\{(\nabla L|h) = 0\} \cap \{L = \ell\}$ is locally finite, the conclusion is obvious since its connected components are reduced to single points. $\diamond$

6.3 Applications to Finance

6.3.1 *Application to Recursive Variance Reduction by Importance Sampling*

This section was originally motivated by the seminal paper [12]. Finally, we followed the strategy developed in [199] which provides, in our mind, an easier to implement procedure. Assume we want to compute the expectation

$$\mathbb{E}\,\varphi(Z) = \int_{\mathbb{R}^d} \varphi(z) e^{-\frac{|z|^2}{2}} \frac{dz}{(2\pi)^{\frac{d}{2}}} \tag{6.12}$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is integrable with respect to the normalized Gaussian measure. In order to deal with a consistent problem, we assume throughout this section that

$$\mathbb{P}(\varphi(Z) \neq 0) > 0.$$

$\rhd$ **Examples.** (a) A typical example is provided by an option pricing in a d-dimensional Black–Scholes model where, with the usual notations,

$$\varphi(z) = e^{-r}\phi\left(\left(x_0^i\, e^{(r-\frac{\sigma_i^2}{2})T+\sigma_i\sqrt{T}(Az)_i}\right)_{1\le i\le d}\right),\quad x_0 = (x_0^1, \dots, x_0^d) \in (0, +\infty)^d,$$

with A a lower triangular matrix such that the covariance matrix $R = AA^*$ has diagonal entries equal to 1 and ϕ a non-negative, continuous if necessary, payoff function. The dimension d corresponds to the number of underlying risky assets.

(b) Monte Carlo simulation of functionals of the Euler scheme of a diffusion (or Milstein scheme) appear as integrals with respect to a multivariate Gaussian vector. Then the dimension d can be huge since it corresponds to the product of the number of time steps by the number of independent Brownian motions driving the dynamics of the *SDE*.

Variance reduction by mean translation: first approach (see [12]).

A change of variable $z = \zeta + \theta$, for a fixed $\theta \in \mathbb{R}^d$, leads to

$$\mathbb{E}\,\varphi(Z) = e^{-\frac{|\theta|^2}{2}}\mathbb{E}\left(\varphi(Z+\theta)e^{-(\theta|Z)}\right). \tag{6.13}$$

increasing Lévy process – independent of the standard Brownian motion W). For more insight on such processes, we refer to [40, 261].

6.3.2 Application to Implicit Correlation Search

We consider a 2-dimensional *B-S* toy model as defined by (2.2), *i.e.* $X_t^0 = e^{rt}$ (riskless asset) and

$$X_t^i = x_0^i e^{(r-\frac{\sigma_i^2}{2})t+\sigma_i W_t^i}, \quad x_0^i > 0, \quad i = 1, 2,$$

for the two risky assets, where $\langle W^1, W^2\rangle_t = \rho t$, $\rho \in [-1, 1]$ denotes the correlation between W^1 and W^2, that is, the correlation between the yields of the risky assets X^1 and X^2.

In this market, we consider a *best-of call* option defined by its payoff

$$\big(\max(X_T^1, X_T^2) - K\big)_+.$$

A market of such *best-of calls* is a market of the correlation ρ since the respective volatilities are obtained from the markets of vanilla options on each asset as implicit volatilities. In this 2-dimensional *B-S* setting, there is a closed formula for the premium involving the bi-variate standard normal distribution (see [159]), but what follows can be applied as soon as the asset dynamics – or their time discretization – can be simulated at a reasonable computational cost.

We will use a stochastic recursive procedure to solve the inverse problem in ρ

$$P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T) = P_{market} \;\; \text{[Mark-to-market premium]}, \qquad (6.24)$$

where

/T

/T

$$\begin{aligned} P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T) &:= e^{-r\cancel{t}}\,\mathbb{E}\Big(\big(\max(X_T^1, X_T^2) - K\big)_+\Big) \\ &= e^{-r\cancel{t}}\,\mathbb{E}\bigg(\Big(\max\Big(x_0^1 e^{\mu_1 T+\sigma_1\sqrt{T}Z_1}, x_0^2 e^{\mu_2 T+\sigma_2\sqrt{T}(\rho Z^1+\sqrt{1-\rho^2}Z^2)}\Big) - K\Big)_+\bigg), \end{aligned}$$

where $\mu_i = r - \frac{\sigma_i^2}{2}$, $i = 1, 2$, $Z = (Z^1, Z^2) \stackrel{d}{=} \mathcal{N}(0; I_2)$.

It is intuitive and easy to check (at least empirically by simulation) that the function $\rho \longmapsto P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T)$ is continuous and (strictly) decreasing on $[-1, 1]$. We assume that the market price is at least consistent, *i.e.* that $P_{market} \in [P_{BoC}(1), P_{BoC}(-1)]$ so that Eq. (6.24) in ρ has exactly one solution, say ρ_*. This example is not only a toy model because of its basic *B-S* dynamics, it is also due to the fact that, in such a model, more efficient deterministic procedures can be called upon, based on the closed form for the option premium. Our aim is to propose and illustrate below a general methodology for correlation search.

The most convenient way to prevent edge effects due to the fact that $\rho \in [-1, 1]$ is to use a trigonometric parametrization of the correlation by setting

$$\rho = \cos\theta, \ \theta \in \mathbb{R}.$$

At this stage, note that

$$\sqrt{1-\rho^2}\, Z^2 = |\sin\theta| Z^2 \stackrel{d}{=} (\sin\theta)\, Z^2$$

since $Z^2 \stackrel{d}{=} -Z^2$. Consequently, as soon as $\rho = \cos\theta$,

$$\rho Z^1 + \sqrt{1-\rho^2}\, Z^2 \stackrel{d}{=} (\cos\theta)\, Z^1 + (\sin\theta)\, Z^2$$

owing to the independence of Z^1 and Z^2.

In general, this introduces an over-parametrization, even inside $[0, 2\pi]$, since $\mathrm{Arccos}(\rho^*) \in [0, \pi]$ and $2\pi - \mathrm{Arccos}(\rho^*) \in [\pi, 2\pi]$ are both solutions to our zero search problem, but this is not at all a significant problem for practical implementation: a more careful examination would show that one of these two equilibrium points is "repulsive" and one is "attractive" for the procedure, see Sects. 6.4.1 and 6.4.5 for a brief discussion: this terminology refers to the status of an equilibrium for the *ODE* associated to a stochastic algorithm and the presence (or not) of noise. A noisy repulsive equilibrium cannot, *a.s.*, be the limit of a stochastic algorithm.

From now on, for convenience, we will just mention the dependence of the premium function in the variable θ, namely

$$\theta \longmapsto P_{BoC}(\theta) := e^{-rT}\mathbb{E}\left[\left(\max\left(x_0^1 e^{\mu_1 T+\sigma_1\sqrt{T}Z_1}, x_0^2 e^{\mu_2 T+\sigma_2\sqrt{T}\left((\cos\theta)\, Z^1+(\sin\theta)\, Z^2\right)}\right) - K\right)_+\right].$$

The function P_{BoC} is a 2π-periodic continuous function. Extracting the implicit correlation from the market amounts to solving (with obvious notations) the equation

$$P_{BoC}(\theta) = P_{market} \qquad (\rho = \cos\theta),$$

where P_{market} is the quoted premium of the option (mark-to-market price). We need to slightly strengthen the consistency assumption on the market price, which is in fact necessary with almost any zero search procedure: we assume that P_{market} lies in the open interval

$$P_{market} \in \left(P_{BoC}(1), \max_{\theta} P_{BoC}(-1)\right)$$

i.e. that P_{market} is not an extremal value of P_{BoC}. So we are looking for a zero of the function h defined on $\mathbb{R}$ by

$$h(\theta) = P_{BoC}(\theta) - P_{market}.$$

This function admits a representation as an expectation given by

$$h(\theta) = \mathbb{E}\, H(\theta, Z),$$

where $H : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$ is defined for every $\theta \in \mathbb{R}$ and every $z = (z^1, z^2) \in \mathbb{R}^2$ by

T

$$H(\theta, z) = e^{-r\cancel{t}} \left(\max\left(x_0^1 e^{\mu_1 T + \sigma_1 \sqrt{T} z^1}, x_0^2 e^{\mu_2 T + \sigma_2 \sqrt{T}(z^1 \cos\theta + z^2 \sin\theta)} \right) - K \right)_+ - P_{market}$$

(e^{-rT})

and $Z = (Z^1, Z^2) \stackrel{d}{=} \mathcal{N}(0; I_2)$.

Proposition 6.1 *Assume the above assumptions made on the P_{market} and the function P_{BoC}. If, moreover, the equation $P_{BoC}(\theta) = P_{market}$ has finitely many solutions on $[0, 2\pi]$, then the stochastic zero search recursive procedure defined by*

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, Z_{n+1}), \;\; \theta_0 \in \mathbb{R},$$

where $(Z_n)_{n \geq 1}$ is an i.i.d, $\mathcal{N}(0; I_2)$ distributed sequence and $(\gamma_n)_{n \geq 1}$ is a step sequence satisfying the decreasing step assumption (6.7), a.s. converges toward solution θ_* to $P_{BoC}(\theta) = P_{market}$.

Proof. For every $z \in \mathbb{R}^2$, $\theta \mapsto H(\theta, z)$ is continuous, 2π-periodic and dominated by a function $g(z)$ such that $g(Z) \in L^2(\mathbb{P})$ (g is obtained by replacing $z^1 \cos\theta + z^2 \sin\theta$ by $|z^1| + |z^2|$ in the above formula for H). One deduces that both the mean function h and $\theta \mapsto \mathbb{E}\, H^2(\theta, Z)$ are continuous and 2π-periodic, hence bounded.

The main difficulty in applying the Robbins–Siegmund Lemma is to find the appropriate Lyapunov function.

As the quoted value P_{market} is not an extremum of the function P, $\int_0^{2\pi} h_\pm(\theta)d\theta > 0$ where $h_\pm := \max(\pm h, 0)$. The two functions $h_\pm$ are 2π-periodic so that $\int_t^{t+2\pi} h_\pm(\theta)d\theta = \int_0^{2\pi} h_\pm(\theta)d\theta > 0$ for every $t > 0$. We consider any (fixed) solution θ_0 to the equation $h(\theta) = 0$ and two real numbers $\beta^\pm$ such that

$$0 < \beta^+ < \frac{\int_0^{2\pi} h_+(\theta)d\theta}{\int_0^{2\pi} h_-(\theta)d\theta} < \beta^-$$

and we set, for every $\theta \in \mathbb{R}$,

$$\ell(\theta) := h_+(\theta) - \beta^+ h_-(\theta)\mathbf{1}_{\{\theta \geq \theta_0\}} - \beta^- h_-(\theta)\mathbf{1}_{\{\theta \leq \theta_0\}}.$$

The function ℓ is clearly continuous, 2π-periodic "on the right" on $[\theta_0, +\infty)$ and "on the left" on $(-\infty, \theta_0]$. In particular, it is a bounded function. Furthermore, owing to the definition of $\beta^\pm$, standard Black–Scholes model (starting at $x > 0$ at $t = 0$ with interest rate r and maturity T).

(a) Show that the B-S premium $C_{BS}(\sigma)$ is even, increasing on $[0, +\infty)$ and continuous as a function of the volatility. Show that $\lim_{\sigma\to 0} C_{BS}(\sigma) = (x - e^{-rt}K)_+$ and $\lim_{\sigma\to+\infty} C_{BS}(\sigma) = x$.
(b) Deduce from (a) that for any mark-to-market price $P_{market} \in [(x - e^{-rt}K)_+, x]$, there is a unique (positive) B-S implicit volatility for this price.
(c) Consider, for every $\sigma \in \mathbb{R}$,

$$H(\sigma, z) = \chi(\sigma)\left(xe^{-\frac{\sigma^2}{2}T+\sigma\sqrt{T}z} - Ke^{-rt}\right)_+,$$

where $\chi(\sigma) = (1 + |\sigma|)e^{-\frac{\sigma^2}{2}T}$. Carefully justify this choice of H and implement the algorithm with $x = K = 100$, $r = 0.1$ and a market price equal to 16.73. Choose the step parameter of the form $\gamma_n = \frac{c}{x}\frac{1}{n}$, $n \geq 1$, with $c \in [0.5, 2]$ (this is simply a suggestion).

Warning. The above exercise is definitely a toy exercise! More efficient methods for extracting standard implied volatility are available (see *e.g.* [209], which is based on a Newton–Raphson zero search algorithm; a dichotomy approach is also very efficient).

▶ **Exercise** (*Extension to more general asset dynamics*). We now consider a pair of risky assets following two correlated local volatility models,

$$dX_t^i = X_t^i\big(rdt + \sigma_i(X_t)dW_t^i\big),\ X_0^i = x^i > 0,\ i = 1, 2,$$

where the functions $\sigma_i : \mathbb{R}_+^2 \to \mathbb{R}_+$ are bounded Lipschitz continuous functions and the Brownian motions W^1 and W^2 are correlated with correlation $\rho \in [-1, 1]$ so that $\langle W^1, W^2\rangle_t = \rho t$. (This ensures the existence and uniqueness of strong solutions for this *SDE*, see Chap. 7.)

Assume that we know how to simulate (X_T^1, X_T^2), either exactly, or at least as an approximation by an Euler scheme from a d-dimensional normal vector $Z = (Z^1, \ldots, Z^d) \stackrel{d}{=} \mathcal{N}(0; I_d)$.

Show that the above approach can be extended *mutatis mutandis*.

6.3.3 *The Paradigm of Model Calibration by Simulation*

Let $\Theta \subset \mathbb{R}^d$ be an open convex set of $\mathbb{R}^d$. Let

$$\begin{aligned} Y : (\Theta \times \Omega, \mathcal{B}or(\Theta) \otimes \mathcal{A}) &\longrightarrow (\mathbb{R}^p, \mathcal{B}or(\mathbb{R}^p)) \\ (\theta, \omega) &\longmapsto Y_\theta(\omega) = (Y_\theta^1(\omega), \ldots, Y_\theta^p(\omega)) \end{aligned}$$

lu (b) et du (c) ont été dénumérotées
rapport à ma version.

$$Y_\theta = \left(e^{-rT_i}(X_{T_i}^{x,\sigma,\lambda,a} - K_i)_+ - P_{market}(T_i, K_i)\right)_{i=1,\dots,p}.$$

ill also have to make simulability assumptions on Y_θ and, if necessary, on its es with respect to θ (see below). Otherwise our simulation-based approach e meaningless.

is stage, there are essentially two approaches that can be considered in order this problem by simulation:

bbins–Siegmund zero search approach of ∇L, which needs to have access *epresentation of the gradient* – assumed to exist – as an expectation of the ion L.

re direct treatment based on the so-called *Kiefer–Wolfowitz procedure*, which ariant of the Robbins–Siegmund approach based on a finite difference method decreasing step) which does not require the existence of a representation of s an expectation.

bbins–Siegmund approach

ke the following assumptions: for every $\theta_0 \in \Theta$,

$$(\mathrm{Cal}_{RZ}) \equiv \begin{cases} (i) & \mathbb{P}(d\omega)\text{-}a.s.,\ \theta \longmapsto Y_\theta(\omega) \text{ is differentiable} \\ & \text{at } \theta_0 \text{ with Jacobian } \partial_{\theta_0} Y_\theta(\omega), \\ (ii) & \exists\, U_{\theta_0},\ \text{neighborhood of } \theta_0 \text{ in } \Theta,\ \text{ such that} \\ & \left(\frac{Y_\theta - Y_{\theta_0}}{|\theta-\theta_0|}\right)_{\theta \in U_{\theta_0}\setminus\{\theta_0\}} \text{ is uniformly integrable.} \end{cases}$$

hecks – using the exercise "Extension to uniform integrability" which follows em 2.2 – that $\theta \longmapsto \mathbb{E}\, Y_\theta$ is differentiable and that its Jacobian is given by

$$\partial_\theta \mathbb{E}\, Y_\theta = \mathbb{E}\, \partial_\theta Y_\theta.$$

the function L is differentiable everywhere on Θ and its gradient (with respect canonical Euclidean norm) is given by

$$\forall\, \theta \in \Theta,\quad \nabla L(\theta) = \mathbb{E}\,(\partial_\theta Y_\theta)^t S\, \mathbb{E}\, Y_\theta = \mathbb{E}\left((\partial_\theta Y_\theta)^t\right) S\, \mathbb{E}\, Y_\theta.$$

is stage we need a representation of $\nabla L(\theta)$ as an expectation. To this end, onstruct, for every $\theta \in \Theta$, an independent copy $\widetilde{Y}_\theta$ of Y_θ defined as follows: we ider the product probability space $(\Omega^2, \mathcal{A}^{\otimes 2}, \mathbb{P}^{\otimes 2})$ and set, for every $(\omega, \widetilde{\omega}) \in \Omega^2$, $, \widetilde{\omega}) = Y_\theta(\omega)$ (the extension of Y_θ on Ω^2 still denoted by Y_θ) and $\widetilde{Y}_\theta(\omega, \widetilde{\omega}) =$). It is straightforward by the product measure theorem that the two families $_{\theta\in\Theta}$ and $(\widetilde{Y}_\theta)_{\theta\in\Theta}$ are independent with the same distribution. From now on we will e the usual abuse of notation consisting in assuming that these two independent es live on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Now, one can write

les (b) et les (c) ont été dénumérotées
rapport à ma version.

$$Y_\theta = \left(e^{-rT_i}(X_{T_i}^{x,\sigma,\lambda,a} - K_i)_+ - P_{market}(T_i, K_i)\right)_{i=1,\dots,p}.$$

ill also have to make simulability assumptions on Y_θ and, if necessary, on its 'es with respect to θ (see below). Otherwise our simulation-based approach e meaningless.

is stage, there are essentially two approaches that can be considered in order this problem by simulation:

bbins–Siegmund zero search approach of ∇L, which needs to have access *epresentation of the gradient* – assumed to exist – as an expectation of the ion L.

ıre direct treatment based on the so-called *Kiefer–Wolfowitz procedure*, which ariant of the Robbins–Siegmund approach based on a finite difference method ı decreasing step) which does not require the existence of a representation of ıs an expectation.

ɔbbins–Siegmund approach

ke the following assumptions: for every $\theta_0 \in \Theta$,

$$(\mathrm{Cal}_{RZ}) \equiv \begin{cases} (i) & \mathbb{P}(d\omega)\text{–}a.s.,\ \theta \longmapsto Y_\theta(\omega) \text{ is differentiable} \\ & \text{at } \theta_0 \text{ with Jacobian } \partial_{\theta_0} Y_\theta(\omega), \\ (ii) & \exists\, U_{\theta_0},\ \text{neighborhood of } \theta_0 \text{ in } \Theta,\ \text{such that} \\ & \left(\frac{Y_\theta - Y_{\theta_0}}{|\theta-\theta_0|}\right)_{\theta \in U_{\theta_0}\setminus\{\theta_0\}} \text{ is uniformly integrable.} \end{cases}$$

hecks – using the exercise "Extension to uniform integrability" which follows em 2.2 – that $\theta \longmapsto \mathbb{E}\, Y_\theta$ is differentiable and that its Jacobian is given by

$$\partial_\theta \mathbb{E}\, Y_\theta = \mathbb{E}\, \partial_\theta Y_\theta.$$

, the function L is differentiable everywhere on Θ and its gradient (with respect : canonical Euclidean norm) is given by

$$\forall\, \theta \in \Theta,\quad \nabla L(\theta) = \mathbb{E}\,(\partial_\theta Y_\theta)^t S\, \mathbb{E}\, Y_\theta = \mathbb{E}\left((\partial_\theta Y_\theta)^t\right) S\, \mathbb{E}\, Y_\theta.$$ /* /*

ıis stage we need a representation of $\nabla L(\theta)$ as an expectation. To this end, onstruct, for every $\theta \in \Theta$, an independent copy $\widetilde{Y}_\theta$ of Y_θ defined as follows: we ider the product probability space $(\Omega^2, \mathcal{A}^{\otimes 2}, \mathbb{P}^{\otimes 2})$ and set, for every $(\omega, \widetilde{\omega}) \in \Omega^2$, $, \widetilde{\omega}) = Y_\theta(\omega)$ (the extension of Y_θ on Ω^2 still denoted by Y_θ) and $\widetilde{Y}_\theta(\omega, \widetilde{\omega}) =$ $\widetilde{\omega})$. It is straightforward by the product measure theorem that the two families $_{\theta\in\Theta}$ and $(\widetilde{Y}_\theta)_{\theta\in\Theta}$ are independent with the same distribution. From now on we will e the usual abuse of notation consisting in assuming that these two independent es live on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Jow, one can write

rapport à ma version.

It follows from the Burkhölder–Davis–Gundy Inequality (6.49) that, for every $\in [\Gamma_N, +\infty)$,

$$
\begin{aligned}
\mathbb{E}\left[\sup_{s\le t\le s+\delta} |\widetilde{M}^{(0)}_{(t)} - \widetilde{M}^{(0)}_{(s)}|^{2+\beta}\right] &\le C_\beta\, \mathbb{E}\left(\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k (\Delta \widetilde{M}_k)^2\right)^{1+\frac{\beta}{2}} \\
&\le C_\beta \left(\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k\right)^{1+\frac{\beta}{2}} \mathbb{E}\left(\frac{\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k (\Delta \widetilde{M}_k)^2}{\sum_{k=N(s)+1}^{N(s+\delta)} \cancel{\gamma_{k}}\ \gamma_k}\right)^{1+\frac{\beta}{2}} \\
&\le C_\beta \left(\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k\right)^{1+\frac{\beta}{2}} \mathbb{E}\left(\frac{\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k |\Delta \widetilde{M}_k|^{2+\beta}}{\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k}\right) \\
&\le C_\beta \left(\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k\right)^{\frac{\beta}{2}} \sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k\, \mathbb{E}\,|\Delta \widetilde{M}_k|^{2+\beta},
\end{aligned}
$$

where C_β is a positive real constant. One finally derives that, for every $s \in [\Gamma_N, +\infty)$,

$$
\begin{aligned}
\mathbb{E}\left[\sup_{s\le t\le s+\delta} |M^{(0)}_{(t)} - M^{(0)}_{(s)}|^{2+\delta}\right] &\le C_\delta A(\varepsilon) \left(\sum_{k=N(s)+1}^{N(s+\delta)} \gamma_k\right)^{1+\frac{\beta}{2}} \\
&\le C_\delta A(\varepsilon)\Big(\delta + \sup_{k\ge N(s)+1} \gamma_k\Big)^{1+\frac{\beta}{2}}.
\end{aligned}
$$

Noting that $\widetilde{M}^{(n)}_{(t)} = M^{(0)}_{\Gamma_n+t} - M^{(0)}_{\Gamma_n}$, $t \ge 0$, $n \ge N$, we derive

$$
\forall n \ge N,\ \forall s \ge 0,\quad \mathbb{E} \sup_{s\le t\le s+\delta} |\widetilde{M}^{(n)}_{(t)} - \widetilde{M}^{(n)}_{(s)}|^{2+\beta} \le C'_\delta\Big(\delta + \sup_{k\ge N(\Gamma_n)+1} \gamma_k\Big)^{1+\frac{\beta}{2}}.
$$

Then, by Markov's inequality, we have for every $\varepsilon > 0$ and $T > 0$,

$$
\overline{\lim_n}\, \frac{1}{\delta} \sup_{s\in[0,T]} \mathbb{P}\Big(\sup_{s\le t\le s+\delta} |\widetilde{M}^{(n)}_{(t)} - \widetilde{M}^{(n)}_{(s)}| \ge \varepsilon\Big) \le C'_\delta \frac{\delta^{\frac{\beta}{2}}}{\varepsilon^{2+\beta}}.
$$

The C-tightness of the sequence $(\widetilde{M}^{(n)})_{n\ge N}$ follows again from Theorem 6.7(b). Furthermore, for every $n \ge N$, First note that (6.65) combined with the Koksma–Hlawka Inequality (see Proposition (4.3)) imply for every $n \geq 2$

$$|S_n^*| \leq C_\xi V\big(H(y_*,\ .)\big)(\log n)^q, \tag{6.71}$$

where $V\big(H(y_*,\ .)\big)$ denotes the variation in the measure sense of $H(y_*,\ .)$. An Abel transform yields (with the convention $S_0^* = 0$)

$$\begin{aligned} m_n &= \widetilde{\gamma}_n\big(\nabla\Lambda(y_{n-1})\,|\,S_n^*\big) - \sum_{k=1}^{n-1}\big(\widetilde{\gamma}_{k+1}\nabla\Lambda(y_k) - \widetilde{\gamma}_k\nabla\Lambda(y_{k-1})\,|\,S_k^*\big) \\ &= \underbrace{\widetilde{\gamma}_n\big(\nabla\Lambda(y_{n-1})\,|\,S_n^*\big)}_{(a)} - \underbrace{\sum_{k=1}^{n-1}\widetilde{\gamma}_k\big(\nabla\Lambda(y_k) - \nabla\Lambda(y_{k-1})\,|\,S_k^*\big)}_{(b)} \\ &\quad - \underbrace{\sum_{k=1}^{n-1}\Delta\widetilde{\gamma}_{k+1}\big(\nabla\Lambda(y_k)\,|\,S_k^*\big)}_{(c)}. \end{aligned}$$

We aim at showing that m_n converges in $\mathbb{R}$ toward a finite limit by inspecting the above three terms.

One gets, using that $\gamma_n \leq \widetilde{\gamma}_n$,

$$|(a)| \leq \gamma_n \|\nabla\Lambda\|_{\sup} O\big((\log n)^q\big) = O(\gamma_n(\log n)^q) \to 0 \quad \text{as} \quad n \to +\infty.$$

Owing to (6.68), the partial sum (b) satisfies

$$\begin{aligned} \sum_{k=1}^{n-1}\widetilde{\gamma}_k\big|\big(\nabla\Lambda(y_k) - \nabla\Lambda(y_{k-1})\,|\,S_k^*\big)\big| &\leq C_\Lambda \sum_{k=1}^{n-1}\widetilde{\gamma}_k\gamma_k\frac{|H(y_{k-1},\xi_k)|}{\sqrt{1+L(y_{k-1})}}|S_k^*| \\ &\leq C_\Lambda C_H V\big(H(y_*,\ .)\big)\sum_{k=1}^{n-1}\gamma_k^2(\log k)^q, \end{aligned}$$

where we used Inequality (6.71) in the second inequality.

Consequently the series $\sum_{k\geq 1}\widetilde{\gamma}_k\big(\nabla L(y_k) - \nabla L(y_{k-1})\,|\,S_k^*\big)$ is (absolutely) convergent owing to Assumption (6.65).

Finally, one deals with term (c). First notice that

$$|\widetilde{\gamma}_{n+1} - \widetilde{\gamma}_n| \leq \gamma_{n+1} - \gamma_n + C_\Lambda\gamma_{n+1}^2\gamma_n \leq C'_\Lambda \max(\gamma_n^2, \gamma_{n+1} - \gamma_n)$$

for some real constant C'_Λ. One checks that the series (c) is also (absolutely) convergent owing to the boundedness of ∇L, Assumption (6.65) and the upper-bound (6.71) for S_n^*.

. rappel à ma version.

Then m_n converges toward a finite limit m_∞. This induces that the sequence $(s_n + m_n)_n$ is bounded below since $(s_n)_n$ is non-negative. Now, we know from (6.70) that $(s_n + m_n)$ is also non-increasing, hence convergent in $\mathbb{R}$, which in turn implies that the sequence $(s_n)_{n\geq 0}$ itself is convergent toward a finite limit. The same arguments as in the regular stochastic case yield

$$L(y_n) \longrightarrow L_\infty \text{ as } n \to +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n \Phi^H(y_{n-1}) < +\infty.$$

One concludes, still like in the stochastic case, that (y_n) is bounded and eventually converges toward the unique zero of Φ^H, *i.e.* y_*.

(*b*) is obvious. ◇

ℵ **Practitioner's corner** • The step assumption (6.65) includes all the step sequences of the form $\gamma_n = \frac{c}{n^\alpha}$, $\alpha \in (0, 1]$. Note that as soon as $q \geq 2$, the condition $\gamma_n(\log n)^q \to 0$ is redundant (it follows from the convergence of the series on the right owing to an Abel transform).

• One can replace the (slightly unrealistic) assumption on $H(y_*, .)$ by a more natural Lipschitz continuous assumption, provided one strengthens the step assumption (6.65) into

$$\sum_{n\geq 1} \gamma_n = +\infty, \quad \gamma_n(\log n)n^{1-\frac{1}{q}} \to 0$$

and

$$\sum_{n\geq 1} \max\left(\gamma_n - \gamma_{n+1}, \gamma_n^2\right)(\log n)n^{1-\frac{1}{q}} < +\infty.$$

This is a straightforward consequence of Proïnov's Theorem (Theorem 4.3), which implies that, for every $n \geq 2$,

$$|S_n^*| \leq C(\log n)\, n^{1-\frac{1}{q}}.$$

Note that the above new assumptions are satisfied by the step sequences $\gamma_n = \frac{c}{n^\rho}$, $1 - \frac{1}{q} < \rho \leq 1$.

• It is clear that the mean-reverting assumption on H is much more stringent in the *QMC* setting.

• It remains that theoretical spectrum of application of the above theorem is dramatically more narrow than the original one. However, from a practical viewpoint, one observes on simulations a very satisfactory behavior of such quasi-stochastic procedures, including the improvement of the rate of convergence with respect to the regular *MC* implementation.

▶ **Exercise.** We assume now that the recursive procedure satisfied by the sequence $(y_n)_{n\geq 0}$ is given by

The same reasoning as that carried out above shows that the first order term $\sigma(x)\sigma'(x)\int_0^t W_s dW_s$ in (7.33) becomes

$$\sum_{i,j=1,2} \sigma_i'(x)\sigma_j(x)\int_0^t W_s^i dW_s^j.$$

In particular, when $i \neq j$, this term involves the two Lévy areas $\int_0^t W_s^1 dW_s^2$ and $\int_0^t W_s^2 dW_s^1$, linearly combined with, *a priori*, different coefficients.

If we return to the general setting of a d-dimensional diffusion driven by a q-dimensional standard Brownian motion, with (differentiable) drift $b : \mathbb{R}^d \to \mathbb{R}^d$ and diffusion coefficient $\sigma = [\sigma_{ij}] : \mathbb{R}^d \to \mathcal{M}(d, q, \mathbb{R})$, elementary though tedious computations lead us to define the (discrete time) Milstein scheme with step $\frac{T}{n}$ as follows:

$$\begin{aligned}
\bar{X}_0^{mil} &= X_0, \\
\bar{X}_{t_{k+1}^n}^{mil} &= \bar{X}_{t_k^n}^{mil} + \frac{T}{n} b(\bar{X}_{t_k^n}^{mil}) \\
&\quad + \sigma(\bar{X}_{t_k^n}^{mil})\Delta W_{t_{k+1}^n} + \sum_{1 \le i,j \le q} \partial\sigma_{.i}\sigma_{.j}(\bar{X}_{t_k^n}^{mil})\int_{t_k^n}^{t_{k+1}^n} (W_s^i - W_{t_k^n}^i)dW_s^j, \qquad (7.41) \\
&\qquad k = 0, \ldots, n-1,
\end{aligned}$$

where $\Delta W_{t_{k+1}^n} := W_{t_{k+1}^n} - W_{t_k^n} = \sqrt{\frac{T}{n}} Z_{k+1}^n$, $\sigma_{.i}(x)$ denotes the i-th column of the matrix σ and, for every $i, j \in \{1, \ldots, q\}$,

$$\forall x = (x^1, \ldots, x^d) \in \mathbb{R}^d, \quad \partial\sigma_{.i}\,\sigma_{.j}(x) := \sum_{\ell=1}^d \frac{\partial\sigma_{.i}}{\partial x^\ell}(x)\sigma_{\ell j}(x) \in \mathbb{R}^d. \qquad (7.42)$$

Remark. A more synthetic way to memorize this quantity is to note that it is the Jacobian matrix of the vector $\sigma_{.i}(x)$ applied to the vector $\sigma_{.j}(x)$.

The ability of simulating such a scheme entirely relies on the exact simulations

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

t^n_{k-1}

i.e. of identical copies of the q^2-dimensional random vector

\paragraphs

7.6.1 Main Results for $\mathbb{E}\, f(X_T)$: the Talay–Tubaro and Bally–Talay Theorems

We adopt the notations of the former Sect. 7.1, except that we still consider, for convenience, an autonomous version of the *SDE*, with initial condition $x \in \mathbb{R}^d$,

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dW_t, \quad X_0^x = x.$$

The notations $(X_t^x)_{t\in[0,T]}$ and $(\bar{X}_t^{n,x})_{t\in[0,T]}$ respectively denote the diffusion and the Euler scheme of the diffusion with step $\frac{T}{n}$ of the diffusion starting at x at time 0 (the superscript ^{n} will often be dropped).

The first result is the simplest result on the weak error, obtained under less stringent assumptions on b and σ.

Theorem 7.7 (see [270]) *Assume b and σ are four times continuously differentiable on $\mathbb{R}^d$ with bounded existing partial derivatives (this implies that b and σ are Lipschitz continuous). Assume $f : \mathbb{R}^d \to \mathbb{R}$ is four times differentiable with polynomial growth as well as its existing partial derivatives. Then, for every $x \in \mathbb{R}^d$,*

$$\mathbb{E}\, f(X_T^x) - \mathbb{E}\, f(\bar{X}_T^{n,x}) = O\left(\frac{1}{n}\right) \quad as \quad n \to +\infty. \tag{7.44}$$

Proof (partial). Assume $d = 1$ for notational convenience. We also assume that $b \equiv 0$, σ is bounded and f has bounded first four derivatives, for simplicity. The diffusion $(X_t^x)_{t\geq 0, x\in\mathbb{R}}$ is a homogeneous Markov process with transition semi-group $(P_t)_{t\geq 0}$ (see *e.g.* [162, 251] among other references) reading on Borel test functions g (*i.e.* bounded or non-negative)

$$P_t\, g(x) := \mathbb{E}\, g(X_t^x), \ t \geq 0, \ x \in \mathbb{R}.$$

On the other hand, the Euler scheme with step $\frac{T}{n}$ starting at $x \in \mathbb{R}$, denoted by $(\bar{X}_{t_k^n}^x)_{0\leq k\leq n}$, is a discrete time homogeneous Markov chain with transition reading on Borel test functions g

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

To be more precise, this means for the diffusion process that, for any Borel test function g,

$$\forall s,\ t \geq 0, \quad P_t g(x) = \mathbb{E}\left(g(X_{s+t}) \,|\, X_s = x\right) = \mathbb{E}\, g(X_t^x)$$

7.7.1 Richardson–Romberg Extrapolation with Consistent Brownian Increments

Bias-variance decomposition of the quadratic error in a Monte Carlo simulation

Let V be a vector space of continuous functions with linear growth satisfying $(\mathcal{E}_2)$ (the case of non-continuous functions is investigated in [225]). Let $f \in V$. For notational convenience, in view of what follows, we set $W^{(1)} = W$ and $X^{(1)} = X$ (including $X_0^{(1)} = X_0 \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ throughout this section). 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 $\frac{T}{n}$ induces the following global (squared) quadratic error

$$\begin{aligned}\left\|\mathbb{E}\, f(X_T) - \frac{1}{M}\sum_{m=1}^{M} f\big((\bar{X}_T^{(1)})^m\big)\right\|_2^2 &= \big(\mathbb{E}\, f(X_T) - \mathbb{E}\, f(\bar{X}_T^{(1)})\big)^2 \\ &\quad + \left\|\mathbb{E}\, f(\bar{X}_T^{(1)}) - \frac{1}{M}\sum_{m=1}^{M} f\big((\bar{X}_T^{(1)})^m\big)\right\|_2^2 \\ &= \left(\frac{c_1}{n}\right)^2 + \frac{\mathrm{Var}\big(f(\bar{X}_T^{(1)})\big)}{M} + O(n^{-3}). \quad (7.48)\end{aligned}$$

The above formula is the bias-variance decomposition of the approximation error of the Monte Carlo estimator. The resulting quadratic error bound (7.48) emphasizes that this estimator does not take full advantage of the above expansion $(\mathcal{E}_2)$.

Richardson–Romberg extrapolation

To take advantage of the expansion, we will perform a Richardson–Romberg extrapolation. In this framework (originally introduced in the seminal paper [270]), one considers the strong solution $X^{(2)}$ of a "copy" of Eq. (7.1), driven by a second Brownian motion $W^{(2)}$ and starting from $X_0^{(2)}$ (independent of $W^{(2)}$ with the same distribution as $X_0^{(1)}$) both defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which $W^{(1)}$ and $X_0^{(1)}$ are defined. One may always consider such a Brownian motion by enlarging the probability space Ω if necessary.

Then we consider the Euler scheme with *a twice smaller* step $\frac{T}{2n}$, denoted by $\bar{X}^{(2)}$, associated to $X^{(2)}$, *i.e.* starting from $X_0^{(2)}$ with Brownian increments built from $W^{(2)}$.

We assume from now on that $(\mathcal{E}_3)$ (as defined in (7.47)) holds for f to get more precise estimates but the principle would work with a function simply satisfying $(\mathcal{E}_2)$. Then combining the two time discretization error expansions related to $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$, respectively, we get

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

Then, the new global (squared) quadratic error becomes

... (b) et du (c) ont été dénumérotées
rapport à ma version.

Practitioner's corner

:om a practical viewpoint, one first simulates an Euler scheme with step $\frac{T}{2n}$ using white Gaussian noise $(Z_k^{(2)})_{k\geq 1}$, then one simulates the Gaussian white noise $Z^{(1)}$ ' the Euler scheme with step $\frac{T}{n}$ by setting

$$Z_k^{(1)} = \frac{Z_{2k}^{(2)} + Z_{2k-1}^{(2)}}{\sqrt{2}}, \quad k \geq 1.$$

umerical illustration. We wish to illustrate the efficiency of the Richardson–omberg (*RR*) extrapolation in a somewhat extreme situation where the time dis-etization induces an important bias. To this end, we consider the Euler scheme of e Black–Scholes *SDE*

$$dX_t = X_t\left(r dt + \sigma dW_t\right)$$

ith the following values for the parameters

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

ote that such a volatility $\sigma = 100\%$ per year is equivalent to a 4 year maturity ith volatility 50% (or 16 years with volatility 25%). A high interest rate is chosen :cordingly. We consider the Euler scheme of this *SDE* with step $h = \frac{T}{n}$, namely

$$\bar{X}_{t_{k+1}} = \bar{X}_{t_k}\left(1 + rh + \sigma\sqrt{h}\, Z_{k+1}\right), \qquad \bar{X}_0 = X_0,$$

here $t_k = kh$, $k = 0, \ldots, n$ and $(Z_k)_{1\leq k\leq n}$ is a Gaussian white noise. We purpose-lly choose a coarse discretization step $n = 10$ so that $h = \frac{1}{10}$. One should keep in ind that, in spite of its virtues in terms of closed forms, both coefficients of the lack–Scholes *SDE* have linear growth so that it is quite a demanding benchmark, pecially when the discretization step is coarse. We want to price a vanilla Call ›tion with strike $K = 100$, *i.e.* to compute

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

T
(e^{-rT})

sing a crude Monte Carlo simulation and an *RR* extrapolation with consistent Brow-an increments as described in the above practitioner's corner. The Black–Scholes ference premium is $C_0^{BS} = 42.9571$ (see Sect. 12.2). To equalize the complex-/ of the crude simulation and its *RR* extrapolated counterpart, we use M sample iths, $M = 2^k$, $k = 14, \ldots, 2^{26}$ for the *RR*-extrapolated simulation ($2^{14} \simeq 32\,000$ id $2^{26} \simeq 67\,000\,000$) and $3M$ for the crude Monte Carlo simulation. Figure 7.1 ›picts the obtained results. The simulation is large enough so that, at its end, the ›served error is approximately representative of the residual bias. The blue line rude *MC*) shows the magnitude of the theoretical bias (close to 1.5) for such a coarse ›p whereas the red line highlights the improvement brought by the Richardson–

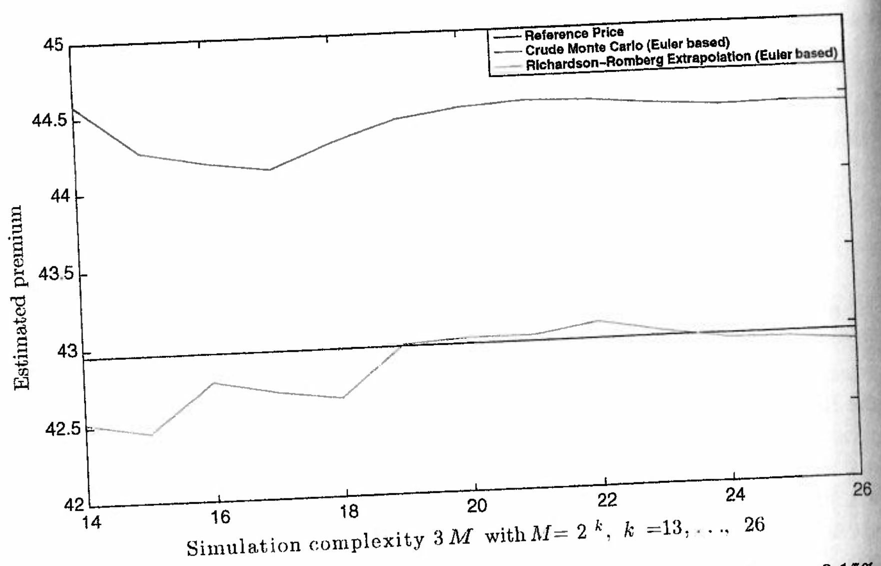


Fig. 7.1 CALL OPTION IN A B-S MODEL PRICED BY AN EULER SCHEME. $\sigma = 1.00, r = 0.15\%$, $T = 1, K = X_0 = 100$. Step $h = 1/10\,(n = 10)$. Black line: reference price; Red line: (Consistent) Richardson–Romberg extrapolation of the Euler scheme of size M; Blue line: Crude Monte Carlo simulation of size $3\,M$ of the Euler scheme (equivalent complexity)

Romberg extrapolation: the residual bias is approximately equal to 0.07, *i.e.* the bias is divided by more than 20.

▶ **Exercises. 1.** Let $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.
(a) Show that

$$\left|\operatorname{cov}(X, Y)\right| \le \sigma(X)\sigma(Y) \quad \text{and} \quad \sigma(X+Y) \le \sigma(X) + \sigma(Y)$$

where $\sigma(X) = \sqrt{\operatorname{Var}(X)} = \|X - \mathbb{E}\,X\|_2$ denotes the *standard-deviation* of X.
(b) Show that $\left|\sigma(X) - \sigma(Y)\right| \le \sigma(X - Y)$ and, for every $\lambda \in \mathbb{R}$, $\sigma(\lambda X) = |\lambda|\sigma(X)$.
2. Let X and $Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ have the same distribution.
(a) Show that $\left|\sigma(X) - \sigma(Y)\right| \le \sigma(X - Y)$. Deduce that, for every $\alpha \in (-\infty, 0] \cup [1, +\infty)$,

$$\operatorname{Var}\big(\alpha X + (1-\alpha)Y\big) \ge \operatorname{Var}(X).$$

(b) Deduce that consistent Brownian increments produce the Richardson-Romberg meta-scheme with the lowest asymptotically variance as n goes to infinity.
3. ($\aleph$ *Practitioner's corner*...). (a) In the above numerical illustration, carry on testing the Richardson–Romberg extrapolation based on Euler schemes *versus* crude Monte Carlo simulation with steps $\frac{T}{n}$ and $\frac{T}{2n}$, $n = 5,\ 10,\ 20,\ 50$, respectively with

A partir de cette race les équations

$$
\begin{aligned}
f_N(t) &\le \|X_0\|_p + C\int_0^t (1+\|X_{s\wedge\tau_N}\|_p)ds \\
&\quad + C_{d,p}^{BDG} C\left(\sqrt{t} + \left[\int_0^t \big\||X_{s\wedge\tau_N}|^2\big\|_{\frac{p}{2}} ds\right]^{\frac12}\right) \\
&= \|X_0\|_p + C\int_0^t (1+\|X_{s\wedge\tau_N}\|_p)ds \\
&\quad + C_{d,p}^{BDG} C\left(\sqrt{t} + \left[\int_0^t \|X_{s\wedge\tau_N}\|_p^2 ds\right]^{\frac12}\right).
\end{aligned}
$$

Consequently, the function f_N satisfies

$$
f_N(t) \le C\left(\int_0^t f_N(s)ds + C_{d,p}^{BDG}\left(\int_0^t f_N^2(s)ds\right)^{\frac12}\right) + \psi(t),
$$

where

$$
\psi(t) = \|X_0\|_p + C\big(t + C_{d,p}^{BDG}\sqrt{t}\big).
$$

bounded

be a bounded Borel function

STEP 2. (*"À la Gronwall" Lemma*).

Lemma 7.3 ("À la Gronwall" Lemma) *Let $f : [0,T] \to \mathbb{R}_+$ and let $\psi : [0,T] \to \mathbb{R}_+$ be two non-negative non-decreasing functions satisfying*

$$
\forall\, t \in [0,T], \qquad f(t) \le A\int_0^t f(s)ds + B\left(\int_0^t f^2(s)ds\right)^{\frac12} + \psi(t),
$$

where A, B are two positive real constants. Then

$$
\forall\, t \in [0,T], \qquad f(t) \le 2\,e^{(2A+B^2)t}\psi(t).
$$

Proof. First, it follows from the elementary inequality $\sqrt{x\,y} \le \frac12\big(\frac{x}{B} + By\big)$, $x, y \ge 0$, $B > 0$, that

$$
\left(\int_0^t f^2(s)ds\right)^{\frac12} \le \left(f(t)\int_0^t f(s)ds\right)^{\frac12} \le \frac{f(t)}{2B} + \frac{B}{2}\int_0^t f(s)ds.
$$

Plugging this into the original inequality yields

$$
f(t) \le (2A + B^2)\int_0^t f(s)ds + 2\,\psi(t).
$$

Gronwall's Lemma 7.2 finally yields the announced result. ◇

$$
\begin{aligned}
f(t) &\le \int_0^t \left\|b(s,X_s)-b(\underline{s},\bar{X}_{\underline{s}})\right\|_p ds + C_{d,p}^{BDG}\left\|\left[\int_0^t \left\|\sigma(s,X_s)-\sigma(\underline{s},\bar{X}_{\underline{s}})\right\|^2 ds\right]^{\frac12}\right\|_p \\
&= \int_0^t \left\|b(s,X_s)-b(\underline{s},\bar{X}_{\underline{s}})\right\|_p ds + C_{d,p}^{BDG}\left\|\int_0^t \left\|\sigma(s,X_s)-\sigma(\underline{s},\bar{X}_{\underline{s}})\right\|^2 ds\right\|_{\frac{p}{2}}^{\frac12} \\
&\le \int_0^t \left\|b(s,X_s)-b(\underline{s},\bar{X}_{\underline{s}})\right\|_p ds + C_{d,p}^{BDG}\left[\int_0^t \left\|\left\|\sigma(s,X_s)-\sigma(\underline{s},\bar{X}_{\underline{s}})\right\|^2\right\|_{\frac{p}{2}} ds\right]^{\frac12} \\
&= \int_0^t \left\|b(s,X_s)-b(\underline{s},\bar{X}_{\underline{s}})\right\|_p ds + C_{d,p}^{BDG}\left[\int_0^t \left\|\left\|\sigma(s,X_s)-\sigma(\underline{s},\bar{X}_{\underline{s}})\right\|\right\|_p^2 ds\right]^{\frac12}.
\end{aligned}
$$

Let us temporarily set $\tau_t^X = \left(1+\left\|\sup_{t\in[0,t]}|X_s|\right\|\right)t$, $t\in[0,T]$. Using Assumption (H_T^β) (see (7.14)) and the Minkowski Inequality on $\left(L^2([0,T],dt), |\,.\,|_{L^2(dt)}\right)$ spaces, we get

$$
\begin{aligned}
f(t) &\le C_{b,\sigma,T}\Bigg(\int_0^t \left((1+\|X_s\|_p)(s-\underline{s})^\beta + \left\|X_s-\bar{X}_{\underline{s}}\right\|_p\right) ds \\
&\qquad + C_{d,p}^{BDG}\left[\int_0^t \left((1+\|X_s\|_p)(s-\underline{s})^\beta + \left\|X_s-\bar{X}_{\underline{s}}\right\|_p\right)^2 ds\right]^{\frac12}\Bigg) \\
&\le C_{b,\sigma,T}\Bigg(\int_0^t \left((1+\|X_s\|_p)(s-\underline{s})^\beta + \left\|X_s-\bar{X}_{\underline{s}}\right\|_p\right) ds \\
&\qquad + C_{d,p}^{BDG}\left[\left(1+\Big\|\sup_{s\in[0,T]}|X_s|\Big\|_p\right)\left[\int_0^t (s-\underline{s})^{2\beta} ds\right]^{\frac12} + \left[\int_0^t \left\|X_s-\bar{X}_{\underline{s}}\right\|_p^2 ds\right]^{\frac12}\right]\Bigg) \\
&\le C_{b,\sigma,T}\Bigg(\tau_t^X\int_0^t (s-\underline{s})^\beta ds + \int_0^t \left\|X_s-\bar{X}_{\underline{s}}\right\|_p ds \\
&\qquad + C_{d,p}^{BDG}\left[\tau_t^X\left[\int_0^t (s-\underline{s})^{2\beta} ds\right]^{\frac12} + \left[\int_0^t \left\|X_s-\bar{X}_{\underline{s}}\right\|_p^2 ds\right]^{\frac12}\right]\Bigg).
\end{aligned}
$$

Now, using that $0\le s-\underline{s}\le \frac{T}{n}$, we obtain

$$
\begin{aligned}
f(t) \le C_{b,\sigma,T}\Bigg(&\left(1+C_{d,p}^{BDG}\right)\left(\frac{T}{n}\right)^\beta \tau_t^X + \int_0^t \|X_s-\bar{X}_{\underline{s}}\|_p\, ds \\
&+ C_{d,p}^{BDG}\left[\int_0^t \left\|X_s-\bar{X}_{\underline{s}}\right\|_p^2 ds\right]^{\frac12}\Bigg),
\end{aligned}
$$

A partir de cette page les équations

STEP 1 (*Representing and estimating* $\mathbb{E}\, f(X_T) - \mathbb{E}\, f(\bar{X}_T^n)$). It follows from the Feynman–Kac formula (7.68) and the terminal condition $u(T,.) = f$ that

$$\mathbb{E}\, f(X_T) = \int_{\mathbb{R}^d} \mathbb{E}\, f(X_T^x)\mathbb{P}_{X_0}(dx) = \int_{\mathbb{R}^d} u(0,x)\mathbb{P}_{X_0}(dx) = \mathbb{E} u(0, X_0)$$

and $\mathbb{E}\, f(\bar{X}_T^n) = \mathbb{E}\, u(T, \bar{X}_T^n)$. It follows that

$$\begin{aligned}\mathbb{E}\left(f(\bar{X}_T^n) - f(X_T)\right) &= \mathbb{E}\left(u(T, \bar{X}_T^n) - u(0, \bar{X}_0^n)\right)\\ &= \sum_{k=1}^{n} \mathbb{E}\left(u(t_k, \bar{X}_{t_k}^n) - u(t_{k-1}, \bar{X}_{t_{k-1}}^n)\right).\end{aligned}$$

In order to evaluate the increment $u(t_k, \bar{X}_{t_k}^n) - u(t_{k-1}, \bar{X}_{t_{k-1}}^n)$, we apply Itô's formula (see Sect. 12.8) between t_{k-1} and t_k to the function u and use that the Euler scheme satisfies the pseudo-*SDE* with "frozen" coefficients

$$d\bar{X}_t^n = b(\underline{t}, \bar{X}_{\underline{t}}^n)dt + \sigma(\underline{t}, \bar{X}_{\underline{t}}^n)dW_t.$$

Doing so, we obtain

$$\begin{aligned}u(t_k, \bar{X}_{t_k}^n) - u(t_{k-1}, \bar{X}_{t_{k-1}}^n) &= \int_{t_{k-1}}^{t_k} \partial_t u(s, \bar{X}_s^n)ds + \int_{t_{k-1}}^{t_k} \partial_x u(s, \bar{X}_s^n)d\bar{X}_s^n\\ &\quad + \frac{1}{2}\int_{t_{k-1}}^{t_k} \partial_{xx} u(s, \bar{X}_s^n)d\langle \bar{X}^n\rangle_s\\ &= \int_{t_{k-1}}^{t_k} \left(\partial_t + \bar{L}\right)u(s, \underline{s}, \bar{X}_s^n, \bar{X}_{\underline{s}}^n)ds\\ &\quad + \int_{t_{k-1}}^{t_k} \sigma(\underline{s}, \bar{X}_{\underline{s}}^n)\partial_x u(s, \bar{X}_s^n)dW_s,\end{aligned}$$

where $\bar{L}$ is the "frozen" infinitesimal generator defined on functions $g \in C^{1,2}([0,T]\times\mathbb{R})$ by

$$\bar{L}g(s, \underline{s}, x, \underline{x}) = b(\underline{s}, \underline{x})\partial_x g(s,x) + \frac{1}{2}\sigma^2(\underline{s}, \underline{x})\partial_{xx} g(s,x)$$

and $\partial_t g(s, \underline{s}, x, \underline{x}) = \partial_t g(s,x)$.

The bracket process of the local martingale $M_t = \displaystyle\int_0^t \partial_x u(s, \bar{X}_s^n)\sigma(\underline{s}, \bar{X}_{\underline{s}}^n)dW_s$ is given for every $t \in [0,T]$ by

$$\langle M\rangle_T = \int_0^T \left(\partial_x\, u(s, \bar{X}_s)\right)^2 \sigma^2(\underline{s}, \bar{X}_{\underline{s}}^n)ds.$$

We briefly reproduce the proof for the reader's convenience. If $z = 0$, the re obvious since $W_T \stackrel{d}{=} -W_T$. If $z > 0$, one introduces the hitting $\tau_z := \inf\{s > 0 : W_s = z\}$ of $[z, +\infty)$ by W (convention $\inf \varnothing = +\infty$). Th $(\mathcal{F}_t^W)$-stopping time since $[z, +\infty)$ is a closed set and W is a continuous p (this uses that $z > 0$). Furthermore, τ_z is $a.s.$ finite since $\overline{\lim}_{t\to+\infty} W_t = +\infty \; a.s.$ C quently, still by continuity of its paths, $W_{\tau_z} = z \; a.s.$ and $W_{\tau_z+t} - W_{\tau_z}$ is indepe of $\mathcal{F}^W_{\tau_z}$. As a consequence, for every $z \geq \max(y, 0)$, using that $W_{\tau_z} = z$ on the $\{\tau_z \leq T\}$,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{t\in[0,T]} W_t \geq z,\; W_T \leq y\Big) &= \mathbb{P}\big(\tau_z \leq T,\, W_T - W_{\tau_z} \leq y - z\big)\\
&= \mathbb{P}\big(\tau_z \leq T,\, -(W_T - W_{\tau_z}) \leq y - z\big)\\
&= \mathbb{P}\big(\tau_z \leq T,\; W_T \geq 2z - y\big)\\
&= \mathbb{P}\big(W_T \geq 2z - y\big)
\end{aligned}$$

since $2z - y \geq z$. Consequently, one may write for every $z \geq \max(y, 0)$,

$$\mathbb{P}\Big(\sup_{t\in[0,T]} W_t \geq z,\; W_T \leq y\Big) = \int_{2z-y}^{+\infty} h_T(\xi)\, d\xi \quad \text{with} \quad h_T(\xi) = \frac{e^{-\frac{\xi^2}{2T}}}{\sqrt{2\pi T}}$$

Hence, since the involved functions are differentiable one has

$$\begin{aligned}
\mathbb{P}\Big(\sup_{t\in[0,T]} W_t \geq z \mid W_T = y\Big) &= \lim_{\eta\to 0} \frac{\big(\mathbb{P}(W_T \geq 2z - (y+\eta)) - \mathbb{P}(W_T \geq 2z - \ldots}{\big(\mathbb{P}(W_T \leq y + \eta) - \mathbb{P}(W_T \leq y)\big)/\eta}\\
&= \frac{h_T(2z-y)}{h_T(y)} = e^{-\frac{(2z-y)^2 - y^2}{2T}} = e^{-\frac{2z(z-y)}{2T}}.
\end{aligned}$$

$e^{-\frac{2z(z-y)}{T}}$

$\to e^{-\frac{2z(z-y)}{T}}$

Corollary 8.1 *Let $\lambda > 0$ and let $x, y \in \mathbb{R}$. If $Y^{W,T}$ denotes the standard Brow bridge of W between 0 and T, then for every $z \in \mathbb{R}$,*

$$\mathbb{P}\left(\sup_{t\in[0,T]} \Big(x + (y-x)\frac{t}{T} + \lambda Y_t^{W,T}\Big) \leq z\right) = \begin{cases} 1 - \exp\big(-\frac{2}{T\lambda^2}(z-x)(z-y)\big) & \text{if } z \geq \max(x,y),\\ 0 & \text{if } z < \max(x,y). \end{cases}$$

A partir de cette page les équations

Proof. First note that

$$x + (y - x)\frac{t}{T} + \lambda Y_t^{W,T} = \lambda x' + \lambda\left(y'\frac{t}{T} + Y_t^{W,T}\right)$$

with $x' = x/\lambda$, $y' = (y - x)/\lambda$.

Then, the result follows by the previous proposition, using that for any real-valued random variable ξ, every $\alpha \in \mathbb{R}$ and every $\beta \in (0, +\infty)$,

$$\mathbb{P}(\alpha + \beta\xi \le z) = \mathbb{P}\Big(\xi \le \frac{z-\alpha}{\beta}\Big) = 1 - \mathbb{P}\Big(\xi > \frac{z-\alpha}{\beta}\Big). \qquad \diamond$$

▶ **Exercise.** Show using $-W \stackrel{d}{=} W$ that

$$\mathbb{P}\left(\inf_{t\in[0,T]}\left(x + (y-x)\frac{t}{T} + \lambda Y_t^{W,T}\right) \le z\right) = \begin{cases} \exp\left(-\frac{2}{T\lambda^2}(z-x)(z-y)\right) & \text{if } z \le \min(x,y), \\ 1 & \text{if } z > \min(x,y). \end{cases}$$

8.2.3 *Application to Lookback Style Path-Dependent Options*

In this section, we focus on Lookback style options (including general barrier options), *i.e.* exotic options related to payoffs of the form $h_T := f\big(X_T, \sup_{t\in[0,T]} X_t\big)$.

We want to compute an approximation of $e^{-rT}\mathbb{E}\, h_T$ using a Monte Carlo simulation based on the continuous time Euler scheme, *i.e.* we want to compute $e^{-rT}\mathbb{E}\, f\big(\bar{X}_T^n, \sup_{t\in[0,T]} \bar{X}_t^n\big)$. We first note, owing to the chaining rule for conditional expectation, that /T /T

e^{rt} e^{rT}

$$\mathbb{E}\, f\big(\bar{X}_T^n, \sup_{t\in[0,T]} \bar{X}_t^n\big) = \mathbb{E}\left[\mathbb{E}\big(f(\bar{X}_T^n, \sup_{t\in[0,T]} \bar{X}_t^n) \,|\, \bar{X}_{t_\ell}^n, \ell = 0, \ldots, n\big)\right].$$

We derive from Proposition 8.2 that

$$\mathbb{E}\left(f(\bar{X}_T^n, \sup_{t\in[0,T]} \bar{X}_t^n) \,|\, \bar{X}_{t_\ell}^n = x_\ell, \ell = 0, \ldots, n\right) = f\big(x_n, \max_{0\le k\le n-1} M_{x_k, x_{k+1}}^{n,k}\big)$$

where, owing to Proposition 8.3,

$$M_{x,y}^{n,k} := \sup_{t\in[0,T/n]}\left(x + \frac{nt}{T}(y - x) + \sigma(t_k^n, x) Y_t^{W^{(t_k^n)}, T/n}\right)$$

are independent. This can also be interpreted as the random variables $M^{n,k}_{\bar{X}^n_{t^n_k},\bar{X}}$ 0, ..., $n-1$, being conditionally independent given $\bar{X}^n_{t^n_k}$, $k = 0, \ldots, n$. Fo Corollary 8.1, the distribution function $G^{n,k}_{x,y}$ of $M^{n,k}_{x,y}$ is given by

$$G^{n,k}_{x,y}(z) = \left[1 - \exp\left(-\frac{2n}{T\sigma^2(t^n_k, x)}(z-x)(z-y)\right)\right]\mathbf{1}_{\{z \geq \max(x,y)\}}, \quad z$$

Then, the inverse distribution simulation rule (see Proposition 1.1) yields th

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

where we used that $U \stackrel{d}{=} 1-U$. To determine $(G^{n,k}_{x,y})^{-1}$ (at $1-u$), it remains the equation $G^{n,k}_{x,y}(z) := 1-u$ under the constraint $z \geq \max(x,y)$, *i.e.*

$$1 - \exp\left(-\frac{2n}{T\sigma^2(t^n_k, x)}(z-x)(z-y)\right) = 1-u, \quad z \geq \max(x,y)$$

or, equivalently,

$$z^2 - (x+y)z + xy + \frac{T}{2n}\sigma^2(t^n_k, x)\log(u) = 0, \quad z \geq \max(x,y).$$

The above equation has two solutions, the solution below satisfying the c Consequently,

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

Finally,

$$\mathcal{L}\Big(\max_{t\in[0,T]}\bar{X}^n_t \mid \{\bar{X}^n_{t_k} = x_k,\ k = 0, \ldots, n\}\Big) = \mathcal{L}\Big(\max_{0\leq k\leq n-1}(G^{n,k}_{x_k,x_{k+1}})^{-1}(1 -$$

where $(U_k)_{1\leq k\leq n}$ are i.i.d. and uniformly distributed random variables ove interval.

Pseudo-code for Lookback style options

We assume for the sake of simplicity that the interest rate r is 0. By l style options we mean the class of options whose payoff involve possibly and the maximum of (X_t) over $[0, T]$, *i.e.*

$$\mathbb{E}\, f\big(\bar{X}^n_T, \sup_{t\in[0,T]} \bar{X}^n_t\big).$$

Regular Call on maximum is obtained by setting $f(x, y) = (y - K)_+$, the (maximum) Lookback option by setting $f(x, y) = y - x$ and the (maximum) partial lookback $f_\lambda(x, y) = (y - \lambda x)_+$ $\lambda > 1$.

We want to compute a Monte Carlo approximation of $\mathbb{E}\, f\big(\bar{X}^n_T, \sup_{t\in[0,T]} \bar{X}^n_t\big)$ using the continuous Euler scheme. We reproduce below a pseudo-script to illustrate how to use the above result on the conditional distribution of the maximum of the Brownian bridge.

• Set $S^f = 0$.
for $m = 1$ to M
  • Simulate a path of the discrete time Euler scheme and set $x_k := \bar{X}^{n,(m)}_{t^n_k}$, $k = 0, \dots, n$.
  • Simulate $\Xi^{(m)} := \max_{0\le k\le n-1} (G^n_{x_k, x_{k+1}})^{-1}(1 - U^{(m)}_k)$, where $(U^{(m)}_k)_{1\le k\le n}$ are i.i.d. with $\mathcal{U}([0, 1])$-distribution.
  • Compute $f(\bar{X}^{n,(m)}_T, \Xi^{(m)})$.
  • Compute $S^f_m := f(\bar{X}^{n,(m)}_T, \Xi^{(m)}) + S^f_{m-1}$.

end. (m)

• Eventually,

$$\mathbb{E}\, f\big(\bar{X}^n_T, \sup_{t\in[0,T]} \bar{X}^n_t\big) \simeq \frac{S^f_M}{M}$$

for large enough M [3].

Once one can simulate $\sup_{t\in[0,T]} \bar{X}^n_t$ (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 a barrier (up or down). The Brownian bridge is also involved in the methods designed for pricing Asian options.

▶ **Exercise.** (a) Show that the distribution of the infimum of the Brownian bridge $(Y^{W,T,y}_t)_{t\in[0,T]}$ starting at 0 and arriving at y at time T is given by

$$\mathbb{P}\Big(\inf_{t\in[0,T]} Y^{W,T,y}_t \le z\Big) = \begin{cases} \exp\big(-\frac{2}{T} z(z-y)\big) & \text{if } z \le \min(y, 0), \\ 1 & \text{if } z \ge \min(y, 0). \end{cases}$$

---

[3]...Of course one needs to compute the empirical variance (approximately) given by

$$\frac{1}{M}\sum_{m=1}^{M} f(\Xi^{(m)})^2 - \Big(\frac{1}{M}\sum_{m=1}^{M} f(\Xi^{(m)})\Big)^2$$

in order to design a confidence interval, without which the method is simply nonsense....

(*b*) Derive a formula similar to (8.7) for the conditional distribution of the mi
of the continuous Euler scheme using now the inverse distribution functions

$$(F^{n,k}_{x,y})^{-1}(u) = \frac{1}{2}\left(x + y - \sqrt{(x-y)^2 - 2\,T\sigma^2(t^n_k, x)\log(u)/n}\right), \quad u \in ((\ldots$$

$\sigma(t_k^n, x)$

$k\, t_k^n,$ of the random variable $\displaystyle\inf_{t\in[0,T/n]}\left(x + \frac{t}{\frac{T}{n}}(y-x) + \sigma(x)Y^{W,T/n}_t\right).$

Warning! The above method is not appropriate for simulating the joint distr
of the $(n+3)$-tuple $\left(\bar{X}^n_{t^n_k},\ k = 0,\ldots,n,\ \inf_{t\in[0,T]}\bar{X}^n_t,\ \sup_{t\in[0,T]}\bar{X}^n_t\right)$.

8.2.4 *Application to Regular Barrier Options: Variance Reduction by Pre-conditioning*

By *regular barrier options* we mean barrier options having a constant lev
barrier. An up-and-out Call is a typical example of such options with a payof
by

$$h_T = (X_T - K)_+ \mathbf{1}_{\{\sup_{t\in[0,T]} X_t \le L\}}$$

where K denotes the strike price of the option and L $(L > K)$ its barrier.

In practice, the "Call" part is activated at T only if the process (X_t) hits the
$L \le K$ between 0 and T. In fact, as far as simulation is concerned, this "Call pa
be replaced by any Borel function f such that both $f(X_T)$ and $f(\bar{X}^n_T)$ are inte
(this is always true if f has polynomial growth owing to Proposition 7.2)
that these so-called barrier options are in fact a sub-class of generalized max
Lookback options having the specificity that the maximum only shows up th
an indicator function.

Then, one may derive a general weighted formula for $\mathbb{E}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]} \bar{X}\ldots}\right.$
which is an approximation of $\mathbb{E}\left(f(X_T)\mathbf{1}_{\{\sup_{t\in[0,T]} X_t \le L\}}\right)$.

Proposition 8.4 (Up-and-Out Call option)

$$\mathbb{E}\Big(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]} \bar{X}^n_t \le L\}}\Big) = \mathbb{E}\left[f(\bar{X}^n_T)\mathbf{1}_{\{\max_{0\le k\le n} \bar{X}^n_{t_k} \le L\}} \prod_{k=0}^{n-1}\left(1 - e^{-\frac{2n}{T}\frac{(\bar{X}^n_{t_k}-L)(\bar{X}^n_{t_k}-L\ldots}{\sigma^2(t^n_k,\bar{X}^n_{t_k})}}\right.\right.$$

Proof of Equation (8.8)**.** This formula is a typical application of pre-conditi
described in Sect. 3.4. We start from the chaining rule for conditional expecta

$$\mathbb{E}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]} \bar{X}^n_t \le L\}}\right) = \mathbb{E}\Big[\mathbb{E}\Big(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]} \bar{X}^n_t \le L\}} \mid \bar{X}^n_{t_k},\ k = 0,\ldots,n$$

Then we use the conditional independence of the bridges of the genuine Euler scheme given the values $\bar{X}^n_{t_k}$, $k = 0, \ldots, n$, established in Proposition 8.2. It follows that

$$\begin{aligned}
\mathbb{E}\Big(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]}\bar{X}^n_t \le L\}}\Big) &= \mathbb{E}\left(f(\bar{X}^n_T)\mathbb{P}\Big(\sup_{t\in[0,T]}\bar{X}^n_t \le L \mid \bar{X}_{t_k},\ k=0,\ldots,n\Big)\right)\\
&= \mathbb{E}\left(f(\bar{X}^n_T)\prod_{k=1}^{n} G^n_{\bar{X}^n_{t_k},\bar{X}^n_{t_{k+1}}}(L)\right)\\
&= \mathbb{E}\left[f(\bar{X}^n_T)\mathbf{1}_{\{\max_{0\le k\le n}\bar{X}^n_{t_k}\le L\}}\prod_{k=0}^{n-1}\left(1-e^{-\frac{2n}{T}\frac{(\bar{X}^n_{t_k}-L)(\bar{X}^n_{t_{k+1}}-L)}{\sigma^2(t^n_k,\bar{X}^n_{t_k})}}\right)\right]. \quad \diamond
\end{aligned}$$

Furthermore, we know that *the random variable in the right-hand side always has a lower variance* since it is a conditional expectation of the random variable in the left-hand side, namely

$$\begin{aligned}
\operatorname{Var}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\max_{0\le k\le n}\bar{X}_{t_k}\le L\}}\prod_{k=0}^{n-1}\left(1-e^{-\frac{2n}{T}\frac{(\bar{X}^n_{t_k}-L)(\bar{X}^n_{t_{k+1}}-L)}{\sigma^2(t^n_k,\bar{X}^n_{t_k})}}\right)\right)&\\
\le \operatorname{Var}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\sup_{t\in[0,T]}\bar{X}^n_t\le L\}}\right)&.
\end{aligned}$$

▶ **Exercises. 1.** *Down-and-Out option.* Show likewise that for every Borel function $f \in L^1(\mathbb{R}, \mathbb{P}_{\bar{X}^n_T})$,

$\prod_{k=1}^{n}\Big(1-$

$$\mathbb{E}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\inf_{t\in[0,T]}\bar{X}^n_t\ge L\}}\right) = \mathbb{E}\left(f(\bar{X}^n_T)\mathbf{1}_{\{\min_{0\le k\le n}\bar{X}^n_{t_k}\ge L\}}e^{-\frac{2n}{T}\sum_{k=0}^{n-1}\frac{(\bar{X}^n_{t_k}-L)(\bar{X}^n_{t_{k+1}}-L)}{\sigma^2(t^n_k,\bar{X}^n_{t_k})}}\right) \tag{8.9}$$

and that the expression in the second expectation has a lower variance.

2. Extend the above results to barriers of the form

$$L(t) := e^{at+b},\ a,\ b\in\mathbb{R}.$$

$\prod_{k=1}^{n}\Big(1-e^{\frac{2n}{T}\sum\cdots}\Big)$

Remark Formulas like (8.8) and (8.9) can be used to produce quantization-based cubature formulas (see [260]).

8.2.5 *Asian Style Options*

The family of Asian options is related to payoffs of the form owing to Kronecker's Lemma (see Lemma 12.1 applied with $a_n = \frac{\|Z_i\|_2}{\sqrt{\pi_i}}$ and b_n $1/\sqrt{\pi_n}$).

(c) First we decompose $I_{\tau\wedge n}$ into

$$I_{\tau\wedge n} = \sum_{i=1}^{n} \frac{\widetilde{Z}_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}} + \sum_{i=1}^{n} \frac{\mathbb{E}\, Z_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}},$$

where the random variables $\widetilde{Z}_i = Z_i - \mathbb{E}\, Z_i$ are independent and centered.

Then, using that τ and $(Z_n)_{n\ge1}$ are independent, we obtain for every $n,\ m \ge 1$,

$$\begin{aligned}\mathbb{E}\left|\sum_{i=n+1}^{n+m} \frac{\widetilde{Z}_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}}\right|^2 &= \sum_{i=n+1}^{n+m} \frac{\mathbb{E}\,\widetilde{Z}_i^2}{\pi_i^2}\pi_i + 2\sum_{n+1\le i<j\le n+m} \frac{\mathbb{E}\,\widetilde{Z}_i\widetilde{Z}_j}{\pi_i\pi_j}\pi_{i\vee j}\\ &= \sum_{i=n+1}^{n+m} \frac{\mathrm{Var}(Z_i)}{\pi_i}\end{aligned}$$

since $\mathbb{E}\,\widetilde{Z}_i\widetilde{Z}_j = \mathbb{E}\,\widetilde{Z}_i\mathbb{E}\,\widetilde{Z}_j = 0$ if $i \ne j$. Hence, for all integers $n,\ m \ge 1$,

$$\left\|\sum_{i=n+1}^{n+m} \frac{\widetilde{Z}_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}}\right\|_2 \le \left(\sum_{i=n+1}^{n+m} \frac{\mathrm{Var}(Z_i)}{\pi_i}\right)^{\frac{1}{2}}$$

which implies, under the assumption $\sum_{n\ge1} \frac{\mathrm{Var}(Z_n)}{\pi_n} < +\infty$, that the series $\sum_{i=1}^{\tau\wedge n} \frac{\widetilde{Z}_i}{\pi_i} =$ $\sum_{i=1}^{n} \frac{\widetilde{Z}_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}}$ is a Cauchy sequence hence convergent in the complete space $\big(L^2(\mathbb{P}), \|\,.\,\|_2\big)$. Its limit is necessarily $I_\tau = \sum_{i=1}^{\tau} \frac{\widetilde{Z}_i}{\pi_i}$ since that is its $a.s.$ limit.

On the other hand, as

$$\sum_{i\ge1}\left\|\frac{|\mathbb{E}\, Z_i|}{\pi_i}\mathbf{1}_{\{i\le\tau\}}\right\|_2 \le \sum_{i\ge1}\frac{|\mathbb{E}\, Z_i|}{\pi_i}\sqrt{\pi_i} = \sum_{i\ge1}\frac{|\mathbb{E}\, Z_i|}{\sqrt{\pi_i}} < +\infty,$$

please skip line

one deduces that $\sum_{i=1}^{n} \frac{\mathbb{E}\, Z_i}{\pi_i}\mathbf{1}_{\{i\le\tau\}}$ is convergent in $L^2(\mathbb{P})$. Its limit is clearly $\sum_{i=1}^{\tau}\frac{|\mathbb{E}\, Z_i|}{\pi_i}$.

Finally, $I_{\tau\wedge n} \to I_\tau$ in $L^2(\mathbb{P})$ (and $\mathbb{P}$-$a.s.$) and the estimate of $\mathrm{Var}(I_\tau)$ is a straightforward consequence of Minkowski's Inequality. ◇

Unbiased Multilevel estimator

The resulting unbiased multilevel estimator reads, for every integer $M \ge 1$,

$$\widehat{I}_M^{RML} = \frac{1}{M}\sum_{m=1}^{M} I^{(m)}_{\tau^{(m)}} = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{\tau^{(m)}} \frac{Z_i^{(m)}}{\pi_i}, \tag{9.109}$$

where $(\tau^{(i)})_{i\geq 1}$ and $(Z_i^{(i)})_{i\geq 1}$, $i \geq 1$, are independent sequences, distributed as τ and $(Z_i)_{i\geq 1}$, respectively.

Definition 9.6 (Randomized Multilevel estimator) *The unbiased multilevel estimator* (9.109) *associated to the random variables* $(Z_i)_{i\in\mathbb{N}^*}$ *and the random time* τ *is called a* Randomized Multilevel estimator *and will be denoted in short by RML from now on.*

However, for a practical implementation, we need to specify the random time τ so that the estimator has both a finite mean complexity and a finite variance or, equivalently, a finite L^2-norm and such a choice is closely connected to the balance between the rate of decay of the variances of the random variables Z_i and the growth of their computational complexity. This is the purpose of the following section, where we will briefly investigate a geometric setting close to that of the multilevel framework with geometric refiners. → Skip here

ℵ Practitioner's corner

In view of practical implementation, we make the following "geometric" assumptions on the complexity and variance (which are consistent with a multilevel approach, see later) and specify a distribution for τ accordingly. See the exercise in the next section for an analysis of an alternative non-geometric framework.

- *Distribution of* τ: we assume *a priori* that $\tau \stackrel{d}{=} G^*(p)$, $p \in (0, 1)$, so that $p_n = p(1-p)^{n-1}$ and $\pi_n = (1-p)^{n-1}$, $n \geq 1$.
  This choice is motivated by the assumptions below.
- *Complexity*. The complexity of simulating Z_n is of the form

  $$\kappa(Z_n) = \kappa N^n, \ n \geq 1.$$

  Then, the mean complexity of the estimator I_τ is clearly

  $$\bar{\kappa}(I_\tau) = \kappa \sum_{n\geq 1} p_n N^n.$$

  As a consequence this mean complexity is finite if and only if

  $$(1-p)N < 1$$

  and $\bar{\kappa}(I_\tau) = \dfrac{\kappa p N}{1-(1-p)N}$.
- *Variance*. We make the assumption that there exists some $\beta > 0$ such that

This approach has already been introduced in Chap. 2 and will be more deeply de oped further on, in Sect. 10.2, mainly devoted to the tangent process method diffusions.

Otherwise, when $\frac{\partial F}{\partial x}(x, z)$ does not exist or cannot be computed easily (whe F can), a natural idea is to introduce a stochastic finite difference approach. O methods based on the introduction of an appropriate weight will be introduced in last two sections of this chapter.

10.1 Finite Difference Method(s)

The finite difference method is in some way the most elementary and natural meth for computing sensitivity parameters, known as Greeks when dealing with financ derivatives, although it is an approximate method in its standard form. This is al known in financial Engineering as the "Bump Method" or "Shock Method". It can described in a very general setting which corresponds to its wide field of applicatio Finite difference methods were been originally investigated in [117, 119, 194].

10.1.1 *The Constant Step Approach*

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

The regular setting

Proposition 10.1 *Let $x \in \mathbb{R}$. Assume that F satisfies the following local mean quadratic Lipschitz continuous assumption at x*

$\forall x', x''$

$$\exists\, \varepsilon_0 > 0,\ \cancel{\forall x'} \in (x - \varepsilon_0, x + \varepsilon_0),\ \big\|F(\underset{x''}{\cancel{x}}, Z) - F(x', Z)\big\|_2 \le C_{F,Z}|\underset{x''}{\cancel{x}} - x'|. \quad (10.1)$$

Assume the function f is twice differentiable with a Lipschitz continuous second derivative on $(x - \varepsilon_0, x + \varepsilon_0)$. Let $(Z_k)_{k \ge 1}$ be a sequence of i.i.d. random vectors with the same distribution as Z. Then for every $\varepsilon \in (0, \varepsilon_0)$, the mean quadratic error or Root Mean Square Error *(RMSE) satisfies* to be sure that the statistical error becomes smaller. However, it is of course useless to carry on the simulation too far since the bias error is not impacted. Note that such specification of the size M of the simulation breaks the recursive feature of the estimator. Another way to use such an error bound is to keep in mind that, in order to reduce the error by a factor of 2, we need to reduce ε and increase M as follows:

$$\varepsilon \rightsquigarrow \varepsilon/\sqrt{2} \quad \text{and} \quad M \rightsquigarrow 4\,M.$$

Warning (what should never be done)! Imagine that we are using two *independent* samples $(Z_k)_{k\geq 1}$ and $(\widetilde{Z}_k)_{k\geq 1}$ to simulate copies $F(x-\varepsilon, Z)$ and $F(x+\varepsilon, Z)$. Then,

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{M}\sum_{k=1}^{M}\frac{F(x+\varepsilon, Z_k)-F(x-\varepsilon, \widetilde{Z}_k)}{2\varepsilon}\right) &\\ &= \frac{1}{4M\varepsilon^2}\Big(\mathrm{Var}\big(F(x+\varepsilon, Z)\big)+\mathrm{Var}\big(F(x-\varepsilon, Z)\big)\Big)\\ &\simeq \frac{\mathrm{Var}\big(F(x,Z)\big)}{2M\varepsilon^2}.\end{aligned}$$

Note that the asymptotic variance of the estimator of $\frac{f(x+\varepsilon)-f(x-\varepsilon)}{2\varepsilon}$ explodes as $\varepsilon \to 0$ and the resulting quadratic error reads approximately

$$[f'']_{\mathrm{Lip}}\frac{\varepsilon^2}{2}+\frac{\sigma\big(F(x,Z)\big)}{\varepsilon\sqrt{2M}},$$

where $\sigma(F(x,Z)) = \sqrt{\mathrm{Var}(F(x,Z))}$ is the standard deviation of $F(x,Z)$. This leads to consider the unrealistic constraint $M(\varepsilon) \propto \varepsilon^{-6}$ to keep the balance between the bias term and the variance term; or equivalently to switch $\varepsilon \rightsquigarrow \varepsilon/\sqrt{2}$ and $M \rightsquigarrow 8\,M$ to reduce the error by a factor of 2.

▷ **Examples (Greeks computation). 1.** *Sensitivity in a Black–Scholes model.* Vanilla payoffs viewed as functions of a normal distribution correspond to functions F of the form

$$F(x,z) = e^{-r\cancel{t}}h\left(x\,e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}\right),\; z\in\mathbb{R},\; x\in(0,+\infty), \qquad T$$

where $h:\mathbb{R}_+\to\mathbb{R}$ is a Borel function with linear growth. If h is Lipschitz continuous, then

$$|F(x,Z)-F(x',Z)| \leq [h]_{\mathrm{Lip}}|x-x'|e^{-\frac{\sigma^2}{2}T+\sigma\sqrt{T}Z}$$

so that elementary computations show, using that $Z \stackrel{d}{=} \mathcal{N}(0;1)$,

$$\|F(x,Z)-F(x',Z)\|_2 \leq [h]_{\mathrm{Lip}}|x-x'|e^{\frac{\sigma^2}{2}T}.$$

The regularity of f follows from the following easy change of variable

/T (2 T (wi...)

$$f(x) = e^{-rT}\int_{\mathbb{R}} h\left(x\, e^{\mu T+\sigma\sqrt{T}z}\right)e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}} = e^{-rT}\int_0^{+\infty} h(y)e^{-\frac{(\log(x/y)+\mu T)^2}{2\sigma^2 T}}\frac{d}{y\sigma\sqrt{}}$$

where $\mu = r - \frac{\sigma^2}{2}$. This change of variable makes the integral appear as a ' convolution on $(0, +\infty)$ [4] with similar regularizing effects as the standar volution on the whole real line. Under the appropriate growth assumption function h (say polynomial growth), one shows from the above identity th function f is in fact infinitely differentiable over $(0, +\infty)$. In particular, it is differentiable with Lipschitz continuous second derivative over any compact ir included in $(0, +\infty)$.

2. *Diffusion model with Lipschitz continuous coefficients.* Let $X^x = (X^x_t$ denote the Brownian diffusion solution of the *SDE*

$$dX_t = b(X_t)dt + \vartheta(X_t)dW_t, \quad X_0 = x,$$

where b and ϑ are locally Lipschitz continuous functions (on the real line) most linear growth (which implies the existence and uniqueness of a strong s $(X^x_t)_{t\in[0,T]}$ starting from $X^x_0 = x$). In such a case, one should instead write

$$F(x, \omega) = h\big(X^x_T(\omega)\big).$$

The Lipschitz continuity of the flow of the above *SDE* (see Theorem 7.10) that

$$\|F(x, .) - F(x', .)\|_2 \le C_{b,\vartheta}[h]_{\mathrm{Lip}}|x - x'|e^{C_{b,\vartheta}T}$$

where $C_{b,\vartheta}$ is a positive constant only depending on the Lipschitz continuou ficients of b and ϑ. In fact, this also holds for multi-dimensional diffusion pr and for path-dependent functionals.

The regularity of the function f is a less straightforward question. But the is positive in two situations: either h, b and σ are regular enough to apply res the flow of the SDE which allows pathwise differentiation of $x \mapsto F(x, \omega)$ (s orem 10.1 further on in Sect. 10.2.2) or ϑ satisfies a uniform ellipticity assu $\vartheta \ge \varepsilon_0 > 0$.

3. *Euler scheme of a Brownian diffusion model with Lipschitz continuous cients.* The same holds for the Euler scheme. Furthermore, Assumption (10.1 uniformly with respect to n if $\frac{T}{n}$ is the step size of the Euler scheme.

4. F can also be a functional of the whole path of a diffusion, provided F is Li continuous with respect to the sup-norm over $[0, T]$.

[4] The convolution on $(0, +\infty)$ is defined between two non-negative functions f and g on by $f \odot g(x) = \displaystyle\int_0^{+\infty} f(x/y)g(y)dy$.

Proposition 10.2 *Let $x \in \mathbb{R}$. Assume that F satisfies in a x open neighb $(x - \varepsilon_0, x + \varepsilon_0)$, $\varepsilon_0 > 0$, of x the following local mean quadratic θ-Hölder assumption, $\theta \in (0, 1]$, at x: there exists a positive real constant $C_{Hol,F,Z}$*

$$\forall x', x'' \in (x - \varepsilon_0, x + \varepsilon_0), \quad \left\|F(x', Z) - F(x'', Z)\right\|_2 \le C_{Hol,F,Z}|x' - x''|^{\theta}.$$

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

$$\left\| f'(x) - \frac{1}{M}\sum_{k=1}^{M} \frac{F(x+\varepsilon, Z_k) - F(x-\varepsilon, Z_k)}{2\varepsilon} \right\|_2 \le \sqrt{\left([f'']_{\text{Lip}}\frac{\varepsilon^2}{2}\right)^2 + \frac{C^2_{Hol,F,Z}}{(2\varepsilon)^{2(1-\theta)}M}}$$

$$\le [f'']_{\text{Lip}}\frac{\varepsilon^2}{2} + \frac{C_{Hol,F,Z}}{(2\varepsilon)^{1-\theta}\sqrt{M}}. \tag{10.6}$$

This variance of the finite difference estimator explodes as $\varepsilon \to 0$ as soon as $\theta < 1$. As a consequence, in such a framework, to divide the quadratic error by a factor of 2, we need to switch

$$\varepsilon \rightsquigarrow \varepsilon/\sqrt{2} \quad \text{and} \quad M \rightsquigarrow 2^{1-\theta} \times 4\,M = 2^{3-\theta}M.$$

A dual point of view in this singular case is to (roughly) optimize the parameter $\varepsilon = \varepsilon(M)$, given a simulation size of M in order to minimize the quadratic error, or at least its natural upper-bounds. Such an optimization performed on (10.6) yields

$$\varepsilon_{opt} = \left(\frac{2^{\theta} C_{Hol,F,Z}}{[f'']_{\text{Lip}}\sqrt{M}}\right)^{\frac{1}{3-\theta}}$$

which of course depends on M so that it breaks the recursiveness of the estimator. Moreover, its sensitivity to $[f'']_{\text{Lip}}$ (and to $C_{Hol,F,Z}$) makes its use rather unrealistic in practice.

The resulting rate of decay of the quadratic error is $O\left(M^{-\frac{2-\theta}{3-\theta}}\right)$. This rate shows that when $\theta \in (0, 1)$, the lack of L^2-regularity of $x \mapsto F(x, Z)$ slows down the convergence of the finite difference method by contrast with the Lipschitz continuous case where the standard rate of convergence of the Monte Carlo method is preserved.

▷ **Example of the digital option.** A typical example of such a situation is the pricing of digital options (or equivalently the computation of the δ-hedge of a Call or Put options).

Let us consider, still in the standard risk neutral Black–Scholes model, a digital Call option with strike price $K > 0$ defined by its payoff

$$h(\xi) = \mathbf{1}_{\{\xi \ge K\}}$$

and set $F(x,z) = e^{-rt} h\left(x\, e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}\right)$, $z \in \mathbb{R}$, $x \in (0,+\infty)$ (r denotes the constant interest rate as usual). We know that the premium of this option is given for every initial price $x > 0$ of the underlying risky asset by /T

$$f(x) = \mathbb{E}\, F(x, Z) \quad \text{with} \quad Z \stackrel{d}{=} \mathcal{N}(0;1).$$

$\cancel{t} \to e^{-rT}$ (4 times)

Set $\mu = r - \frac{\sigma^2}{2}$. It is clear since $Z \stackrel{d}{=} -Z$ that

$$\begin{aligned} f(x) &= e^{-rt}\mathbb{P}\left(x\, e^{\mu T+\sigma\sqrt{T}Z} \geq K\right) \\ &= e^{-rt}\mathbb{P}\left(Z \geq -\frac{\log(x/K)+\mu T}{\sigma\sqrt{T}}\right) \\ &= e^{-rt}\Phi_0\left(\frac{\log(x/K)+\mu T}{\sigma\sqrt{T}}\right), \end{aligned}$$

/T /T /T /T

where Φ_0 denotes the c.d.f. of the $\mathcal{N}(0;1)$ distribution. Hence the function f is infinitely differentiable on $(0,+\infty)$.

On the other hand, still using $Z \stackrel{d}{=} -Z$, for every $x, x' \in \mathbb{R}$,

$$\begin{aligned} &\left\|F(x,Z) - F(x',Z)\right\|_2^2 \\ &= e^{-2rT}\left\| \mathbf{1}_{\left\{Z \geq -\frac{\log(x/K)+\mu T}{\sigma\sqrt{T}}\right\}} - \mathbf{1}_{\left\{Z \geq -\frac{\log(x'/K)+\mu T}{\sigma\sqrt{T}}\right\}} \right\|_2^2 \\ &= e^{-2rT}\mathbb{E}\left| \mathbf{1}_{\left\{Z \leq \frac{\log(x/K)+\mu T}{\sigma\sqrt{T}}\right\}} - \mathbf{1}_{\left\{Z \leq \frac{\log(x'/K)+\mu T}{\sigma\sqrt{T}}\right\}} \right|^2 \\ &= e^{-2rT}\left(\Phi_0\left(\frac{\log(\max(x,x')/K)+\mu T}{\sigma\sqrt{T}}\right) - \Phi_0\left(\frac{\log(\min(x,x')/K)+\mu T}{\sigma\sqrt{T}}\right)\right). \end{aligned}$$

Using that Φ_0' is bounded by $\kappa_0 = \frac{1}{\sqrt{2\pi}}$, we derive that

$$\left\|F(x,Z) - F(x',Z)\right\|_2^2 \leq \frac{\kappa_0 e^{-2rT}}{\sigma\sqrt{T}}\left|\log x - \log x'\right|.$$

Consequently for every interval $I \subset (0,+\infty)$ *bounded away from* 0, there exists a real constant $C_{r,\sigma,T,I} > 0$ such that

$$\forall\, x, x' \in I, \quad \left\|F(x,Z) - F(x',Z)\right\|_2 \leq C_{r,\sigma,T,I}\sqrt{|x-x'|},$$

i.e. the functional F is $\frac{1}{2}$-Hölder in $L^2(\mathbb{P})$ and the above proposition applies.

► **Exercises. 1.** Prove the above Proposition 10.2.

2. *Digital option.* (a) Consider in the risk neutral Black–Scholes model a di option defined by its payoff

$$h(\xi) = \mathbf{1}_{\{\xi \geq K\}}$$

and set $F(x, z) = e^{-rt} h\left(x\, e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}\right)$, $z \in \mathbb{R}$, $x \in (0, +\infty)$ (r is a con interest rate as usual). We still consider the computation of $f(x) = \mathbb{E}\, F(x, Z)$ w $Z \stackrel{d}{=} \mathcal{N}(0; 1)$.

Verify on a numerical simulation that the variance of the finite difference est tor introduced in Proposition 10.1 explodes as $\varepsilon \to 0$ at the rate expected fron preceding computations.

(b) Derive from the preceding a way to "synchronize" the step ε and the size l the simulation.

10.1.2 A Recursive Approach: Finite Difference with Decreasing Step

In the former finite difference method with constant step, the bias never fades. Co quently, increasing the accuracy of the sensitivity computation, it has to be resu from the beginning with a new ε. In fact, it is easy to propose a recursive versio the above finite difference procedure by considering some variable steps ε whicl to 0. This can be seen as an application of the Kiefer–Wolfowitz principle origin developed for Stochastic Approximation purposes.

We will focus on the "regular setting" (F Lipschitz continuous in L^2) in section, the singular setting is proposed as an exercise. Let $(\varepsilon_k)_{k\geq 1}$ be a sequenc positive real numbers decreasing to 0. With the notations and the assumptions of former section, consider the estimator

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

It can be computed in a recursive way since

$$\widehat{f'(x)}_{M+1} = \widehat{f'(x)}_M + \frac{1}{M+1}\left(\frac{F(x+\varepsilon_{M+1}, Z_{M+1}) - F(x-\varepsilon_{M+1}, Z_{M+1})}{2\varepsilon_{M+1}} - \widehat{f'(x)}_M\right)$$

Elementary computations show that the mean squared error satisfies

If one notes that, for any $s \geq t$, $X_s^x = X_s^{X_t^x, t}$, the first term "$\frac{\partial F(X^x)}{\partial X_t^x}$" in the above product is clearly equal to $DF(X) \cdot (1_{[t,T]} Y_{\cdot}^{(t)})$ whereas the second term is the result of a formal differentiation of the *SDE* at time t with respect to W, namely $\vartheta(X_t^x)$.

An interesting feature of this derivative in practice is that it satisfies the usual chaining rules like

$$D_t F^2(X^x) = 2F(X^x) D_t F(X^x)$$

and more generally

$$D_t \Phi(F(X^x)) = D\Phi(F(X^x)) D_t F(X^x),$$

etc.

What is called *Malliavin calculus* is a way to extend this notion of differentiation to more general functionals using some functional analysis arguments (closure of operators, etc) using, for example, the domain of the operator $D_t F$ (see *e.g.* [16, 208]).

Using the Haussman–Clark–Occone formula to get Bismut's formula

As a first conclusion we will show that the Haussman–Clark–Occone formula contains the Bismut formula. Let X^x, H, f and T be as in Sect. 10.4.1. We consider the two true martingales

\vartheta(X^x_s) ← LaTeX

$$M_t = \int_0^t H_s dW_s \quad \text{and} \quad N_t = \mathbb{E} f(X_T^x) + \int_0^t \mathbb{E}\left(f'(X_T) Y_T^{(s)} \mid \mathcal{F}_s\right) dW_s, \quad t \in [0, T]$$

$\vartheta(X^x_s)$

and perform a (stochastic) integration by parts. Owing to (10.17), we get, under appropriate integrality conditions,

$$\begin{aligned}
\mathbb{E}\left(f(X_T^x) \int_0^T H_s dW_s\right) &= 0 + \mathbb{E}\int_0^T [\ldots] dM_t + \mathbb{E}\int_0^T [\ldots] dN_t \\
&\quad + \mathbb{E}\left(\int_0^T \mathbb{E}\left(f'(X_T) Y_T^{(s)} \mid \mathcal{F}_s\right) \vartheta(X_s^x) H_s \, ds\right) \\
&= \int_0^T \mathbb{E}\left(\mathbb{E}\left(f'(X_T) Y_T^{(s)} \mid \mathcal{F}_s\right) \vartheta(X_s^x) H_s\right) ds
\end{aligned}$$

owing to Fubini's Theorem. Finally, using the characterization of conditional expectation to get rid of the conditioning, we obtain

$$\mathbb{E}\left(f(X_T^x) \int_0^T H_s dW_s\right) = \int_0^T \mathbb{E}\left(f'(X_T) Y_T^{(s)} \vartheta(X_s^x) H_s\right) ds.$$

Finally, a reverse application of Fubini's Theorem and the identity $Y_T^{(s)} = \frac{Y_T}{Y_s}$ leads to

$$\mathbb{E}\left(f(X_T^x) \int_0^T H_s dW_s\right) = \mathbb{E}\left(f'(X_T) Y_T \int_0^T \frac{\vartheta(X_s^x)}{Y_s} H_s \, ds\right),$$

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$\Phi_0(t)$	.99865	.99904	.99931	.99952	.99966	.99976	.999841	.999928	.999968	.999997

12.2 Black–Scholes Formula(s) (To Compute Reference Prices)

In a risk-neutral Black–Scholes model, the quoted price of a risky asset is a solutic to the *SDE* $dX_t = X_t(rdt + \sigma dW_t)$, $X_0 = x_0 > 0$, where r is the interest rate ar $\sigma > 0$ is the volatility and W is a standard Brownian motion. Itô's formula (se Sect. 12.8) yields that

$$X_t^{x_0} = x_0\, e^{(r-\frac{\sigma^2}{2})t+\sigma W_t}, \quad W_t \stackrel{d}{=} \mathcal{N}(0;1).$$

A vanilla (European) payoff of maturity $T > 0$ is of the form $h_T = \varphi(X_T)$. European option contract written on the payoff h_T is the right to receive h_T the maturity T. Its price – or premium – at time $t = 0$ is given by $e^{-rT}\mathbb{E}\,\varphi(X_T^{x_0}$ and, more generally at time $t \in [0, T]$, it is given by $e^{-r(T-t)}\mathbb{E}\left(\varphi(X_T^{x_0}) \mid X_t^{x_0}\right) = e^{-r(T-t)}\mathbb{E}\,\varphi(X_{T-t}^{x_0})$. In the case where $\varphi(x) = (x - K)_+$ (*call* with *strike price K* this premium at time t has a closed form given by

$$\mathrm{Call}_t(x_0, K, R, \sigma, T) = \mathrm{Call}_0(x_0, K, R, \sigma, T - t),$$

where

$$\mathrm{Call}_0(x_0, K, r, \sigma, \tau) = x_0\Phi_0(d_1) - e^{-r\tau}K\Phi_0(d_2), \quad \tau > 0, \qquad (12.1$$

with

$$d_1 = \frac{\log\left(\frac{x_0}{K}\right) + (r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}. \qquad (12.2)$$

As for the *put* option written on the payoff $h_T = (K - X_T^{x_0})_+$, the premium is

$$\mathrm{Put}_t(x_0, K, r, \sigma, T) = \mathrm{Put}_0(x_0, K, r, \sigma, T - t),$$

where

$$\mathrm{Put}_0(x_0, K, r, \sigma, \tau) = e^{-r\tau}K\Phi_0(-d_2) - x_0\Phi_0(-d_1). \qquad (12.3)$$

The avatars of the regular Black–Scholes formulas can be obtained as follows:

- *Stock without dividend (Black–Scholes):* the risky asset is X.
- *Stock with continuous yield $\lambda > 0$ of dividends:* the risky asset is $e^{\lambda t}X_t$ and one has to replace x_0 by $e^{-\lambda\tau}x_0$ in the right-hand sides of (12.1), (12.2) and (12.3).

Gilles Pagès 8.02.2023

A partir de la page 413 du livre ligne 7, les numérotations diffèrent d'avec ma dernière version "perso" car les équations des item (b) et (c) ne sont pas numérotées dans le livre mais le sont dans ma version perso.

Mais bizarrement cela ~~crée~~ semble-t-il des références fausses (et juste décalées de 2 unités) relatives à l'oscillateur de Cameron-Martin : - p.160 dans l'exercice 5 il faudrait lire Using the identity (9.103) et non (9.105).

- Idem p.560 ligne -4 où il faudrait lire (9.102) et non (9.104)
- Idem p.562 ligne 8 ________ (9.103) et non (9.105

Je n'ai pas trouvé d'autres occurrences et il me semble que la labellisation et le référencement sont corrects.

$$\Phi_n(x, y) = 1 + \sum_{k=1}^{n} \mathbf{1}_{\{\xi_k^1 < x\}} \mathbf{1}_{\{\xi_k^2 < y\}}$$

$$= 1 - 1 + \sum_{r=0}^{r_1} \sum_{s=0}^{r_2} \sum_{u_r=0}^{x_{r+1}-1} \sum_{v_s=0}^{y_{s+1}-1} \left\lfloor \frac{n}{p_1^{r+1} p_2^{s+1}} \right\rfloor + \eta_{r,s,u_r,v_s}$$

$$= \sum_{r=0}^{r_1} \sum_{s=0}^{r_2} x_{r+1} y_{s+1} \left(\left\lfloor \frac{n}{p_1^{r+1} p_2^{s+1}} \right\rfloor + \eta_{r,s} \right),$$

where $\eta_{r,s} \in [0, 1]$. Owing to (12.8), (12.9) and the obvious fact that nx $\sum_{0 \le r \le r_1} \sum_{0 \le s \le r_2} \frac{n\, x_{r+1} y_{s+1}}{p^{r+1} q^{s+1}}$, we derive that for every $(x, y) \in \mathcal{X}^n_{1,2}$,

$$\left|\widetilde{\Phi}_n(x, y)\right| \le \sum_{r=0}^{r-1} \sum_{s=0}^{r_2} \sum_{u_r=0}^{x_{r+1}-1} \sum_{v_s=0}^{y_{s+1}-1} x_{r+1} y_{s+1} \left| \underbrace{\left\lfloor \frac{n}{p_1^{r+1} p_2^{s+1}} \right\rfloor - \frac{n}{p_1^{r+1} p_2^{s+1}}}_{\in[-1,0]} + \underbrace{\eta_{r,}}_{\in[0,} \right.$$

$$\le (r_1 + 1)(r_2 + 1)(p_1 - 1)(p_2 - 1).$$

We conclude by noting that, on the one hand, $r_i + 1 = \left\lfloor \frac{\log p_i n}{\log p_i} \right\rfloor$, $i = 1, 2$ ar the other hand,

$$n\, D_n^*(\xi) \le \max\left(\sup_{(x,y)\in[0,1]^2} \left|\widetilde{\Phi}_n(x, y)\right|, n\, D_n^*(\xi^1), n\, D_n^*(\xi^2) \right).$$

Finally, following the above lines – in a simpler way – one shows that

$$n\, D_n^*(\xi^i) \le (r_i + 1)(p_i - 1), i = 1, 2.$$

This completes the proof since $(r_i + 1)(p_i - 1) \ge 1$, $i = 1, 2$, and $\max(a, b)$ for every $a, b \in \mathbb{N}^*$.

12.11 A Pitman–Yor Identity as a Benchmark

We aim at computing $\mathbb{E} \cos(X_t^2)$, where X^2 denotes the second compone
12 Clark–Cameron oscillator defined by (9.104), namely

$$X_t^2 = \sigma \int_0^t (W_s^1 + \mu s) dW_s^2, \quad t \in [0, T].$$

Conditional on the process $(W_s^1)_{0 \le s \le t}$, X_t^2 has a centered Gaussian distributi stochastic variance

$$\mathbb{E}^*\left[e^{\mu B_t+\frac{B_t^2}{2t}(1-\sigma t\coth\sigma t)}\right]=\int_{-\infty}^{+\infty}e^{\mu x+\frac{x^2}{2t}(1-\sigma t\coth\sigma t)}e^{-\frac{x^2}{2t}}\frac{dx}{\sqrt{2\pi}}$$

$$=\frac{1}{\sqrt{t}}\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}\left(x^2\sigma\coth\sigma t-2\mu x\right)\Big)\frac{dx}{\sqrt{2\pi}}.$$

We set $a=\sqrt{\sigma\coth\sigma t}$ and $b=\mu/a$ and we get

$$\mathbb{E}^*\left[e^{\mu B_t+\frac{B_t^2}{2t}(1-\sigma t\coth(\sigma t))}\right]=\frac{1}{\sqrt{t}}e^{\frac{b^2}{2}}\int_{-\infty}^{+\infty}\exp\Big(-\frac{(ax-b)^2}{2}\Big)\frac{dx}{\sqrt{2\pi}}=\frac{e^{\frac{b^2}{2}}}{a\sqrt{t}}.$$

Hence,

$$\mathbb{E}\,e^{\bar{i}X_t^2}=\sqrt{\frac{\sigma t}{\sinh\sigma t}}\frac{1}{\sqrt{t}}\frac{1}{\sqrt{\sigma\coth\sigma t}}e^{-\frac{1}{2}\mu^2t+\frac{b^2}{2}}=\frac{1}{\sqrt{\cosh\sigma t}}e^{-\frac{\mu^2t}{2}\left(1-\frac{\tanh\sigma t}{\sigma t}\right)}.$$

3 Since the right term of the above equality has no imaginary part, we obtained the announced formula (9.10~~5~~1)

3

$$\mathbb{E}\,\cos(\sigma X_t^2)=\frac{e^{-\frac{\mu^2t}{2}\left(1-\frac{\tanh\sigma t}{\sigma t}\right)}}{\sqrt{\cosh\sigma t}}.$$

Remark. Note that these computations are shortened when there is no drift term, *i.e.*

$$X_t^1=W_t^1,\quad X_t^2=\sigma\int_0^t X_s^1\,dW_s^1.$$

Indeed, owing to the Cameron–Martin formula (see [251] p. 445),

$$\mathbb{E}\,e^{-\sigma\int_0^1(W_s^1)^2ds}=\left(\cosh\sqrt{2\sigma}\right)^{-\frac{1}{2}}$$

and the scaling property of the Brownian motion yields

$$\mathbb{E}\,e^{\bar{i}X_t^2}=\mathbb{E}\,e^{-\frac{\sigma^2}{2}\int_0^t(W_s^2)^2ds}=\mathbb{E}\,e^{-\frac{\sigma^2t^2}{2}\int_0^1(W_s^2)^2ds}=(\cosh\sigma t)^{-\frac{1}{2}}.$$

► **Exercise.** Prove in detail the identity (12.12).

denote the Lévy area associated to (W^1, W^2) at time 1. We admit that the characteristic function of X reads

$$\chi(u) = \mathbb{E}\, e^{\tilde{\imath} u X} = \frac{1}{\sqrt{\cosh u}}, \quad u \in \mathbb{R},$$

(see Formula (9.105) in Chap. 9 further on, applied here with $\mu = 0$, and Sect. 12.11 of the Miscellany Chapter for a proof).

(a) Show that

$$C := \mathbb{E}\, X_+ = \frac{1}{\sqrt{2\pi}}\, \mathbb{E}\, \|B\|_{L^2([0,1],dt)},$$

where $B = (B_t)_{t\in[0,1]}$ denotes a standard Brownian motion.

(b) Establish the elementary identity

$$\mathbb{E}\, \|B\|_{L^2([0,1],dt)} = \frac{1}{4} + \mathbb{E}\left(\|B\|_{L^2([0,1],dt)} - \frac{1}{2}\, \|B\|^2_{L^2([0,1],dt)} \right)$$

and justify why $\mathbb{E}\, \|B\|_{L^2([0,1],dt)}$ should be computed by a Monte Carlo simulation using this identity. [Hint: An appropriate Monte Carlo simulation should yield a result close to 0.2485, but this approximation is not accurate enough to compute optimal quantizers ([4].]

(c) Describe in detail a method (or possibly two methods) for computing a small database of N-quantizers of the Lévy area for levels running from $N = 1$ to $N_{\max} = 50$, including, for every level $N \geq 1$, both their weights and their induced quadratic mean quantization error. [Hint: Use (5.15) to compute $\|X - \widehat{X}^{\Gamma}\|_2^2$ when Γ is a stationary quantizer.]

3

5. *Clark–Cameron oscillator*. Using the identity (9.105) in its full generality, extend the quantization procedure of Exercise **4**, to the case where

$$X = \int_0^1 (W_s^1 + \mu s)\, dW_s^2,$$

with μ a fixed real constant.

6. *Supremum of the Brownian bridge*. Let

$$X = \sup_{t\in[0,1]} |W_t - t\, W_1|$$

denote the supremum of the standard Brownian bridge (see Chap. 8 for more details, see also Sect. 4.3 for the connection with uniformly distributed sequences and

---

[4]A more precise approximation is $C = 0.24852267852801818 \pm 2.033\ 10^{-7}$ obtained by implementing an *ML2R* estimator with a target *RMSE* $\varepsilon = 3.0\ 10^{-7}$, see Chap. 9.