Méthodes de Monte Carlo et Algorithmes Stochastiques

BERNARD LAPEYRE

Septembre 2017

Contents

Chapter 1

Monte-Carlo Methods for Options

Monte-Carlo methods are extensively used in financial institutions to compute European options prices, to evaluate sensitivities of portfolios to various parameters and to compute risk measurements.

Let us describe the principle of the Monte-Carlo methods on an elementary example. Let

$$\int_{[0,1]^d} f(x)dx,$$

where $f(\cdot)$ is a bounded real valued function. Represent I as $\mathbb{E}(f(U))$, where U is a uniformly distributed random variable on $[0,1]^d$. By the Strong Law of Large Numbers, if $(U_i, i \geq 1)$ is a family of uniformly distributed independent random variables on $[0,1]^d$, then the average

$$S_N = \frac{1}{N}\sum_{i=1}^{N} f(U_i) \tag{1.1}$$

converges to $\mathbb{E}(f(U))$ almost surely when N tends to infinity. This suggests a very simple algorithm to approximate I: call a random number generator N times and compute the average (1.1). Observe that the method converges for *any* integrable function on $[0,1]^d$: f is not necessarily a smooth function.

In order to efficiently use the above Monte-Carlo method, we need to know its rate of convergence and to determine when it is more efficient than deterministic algorithms. The Central Limit Theorem provides the asymptotic distribution of $\sqrt{N}(S_N - I)$ when N tends to $+\infty$. Various refinements of the Central Limit Theorem, such as Berry-Essen and Bikelis theorems, provide non asymptotic estimates.

The preceding consideration shows that the convergence rate of a Monte Carlo method is rather slow ($1/\sqrt{N}$). Moreover, the approximation error is random and may take large values even if N is large (however, the probability of such an event tends to 0 when N tends to infinity). Nevertheless, the Monte-Carlo methods are useful in practice. For instance, consider an integral in a hypercube $[0,1]^d$, with d large ($d = 40$, e.g.). It is clear that the quadrature methods require too many points (the number of points increases exponentially with the dimension of the space). Low discrepancy sequences are efficient for moderate value of d but this efficiency decreases drastically when d becomes large (the discrepancy behaves like $C(d)\frac{\log^d(N)}{N}$ where the constant

$C(d)$ may be extremely large.). A Monte-Carlo method does not have such disadvantages : it requires the simulation of independent random vectors $(X_1, \ldots, X_d)$, whose coordinates are independent. Thus, compared to the computation of the one-dimensional situation, the number of trials is multiplied by d only and therefore the method remains tractable even when d is large. In addition, another advantage of the Monte-Carlo methods is their parallel nature: each processor of a parallel computer can be assigned the task of making a random trial.

To summarize the preceding discussion : probabilistic algorithms are used in situations where the deterministic methods are unefficient, especially when the dimension of the state space is very large. Obviously, the approximation error is random and the rate of convergence is slow, but in these cases it is still the best method known.

1.1 On the convergence rate of Monte-Carlo methods

In this section we present results which justify the use of Monte-Carlo methods and help to choose the appropriate number of simulations N of a Monte-Carlo method in terms of the desired accuracy and the confidence interval on the accuracy.

Theorem 1.1.1 (Strong Law of Large Numbers)**.** *Let $(X_i, i \geq 1)$ be a sequence of independent identically distributed random variables such that $\mathbb{E}(|X_1|) < +\infty$. The one has :*

$$\lim_{n\to+\infty} \frac{1}{n}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) \text{ a.s.}$$

Remark 1.1.1. The random variable X_1 needs to be integrable. Therefore the Strong Law of Large Numbers does not apply when X_1 is Cauchy distributed, that is when its density is $\frac{1}{\pi(1+x^2)}$.

Convergence rate We now seek estimates on the error

$$\epsilon_n = \mathbb{E}(X) - \frac{1}{n}(X_1 + \cdots + X_n).$$

The Central Limit Theorem precises the asymptotic distribution of $\sqrt{N}\epsilon_N$.

Theorem 1.1.2 (Central Limit Theorem)**.** *Let $(X_i, i \geq 1)$ be a sequence of independent identically distributed random variables such that $\mathbb{E}(X_1^2) < +\infty$.Let σ^2 denote the variance of X_1, that is*

$$\sigma^2 = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \mathbb{E}\left((X_1 - \mathbb{E}(X_1))^2\right).$$

Then :

$$\left(\frac{\sqrt{n}}{\sigma}\epsilon_n\right) \text{ converges in distribution to } G,$$

where G is a Gaussian random variable with mean 0 and variance 1.

Remark 1.1.2. From this theorem it follows that for all $c_1 < c_2$

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\sigma}{\sqrt{n}}c_1 \leq \epsilon_n \leq \frac{\sigma}{\sqrt{n}}c_2\right) = \int_{c_1}^{c_2} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

In practice, one applies the following approximate rule, for n large enough, the law of ϵ_n is a Gaussian random variable with mean 0 and variance σ^2/n.

Note that it is impossible to bound the error, since the support of any (non degenerate) Gaussian random variable is $\mathbb{R}$. Nevertheless the preceding rule allow one to define a confidence interval : for instance, observe that

$$\mathbb{P}\left(|G| \leq 1.96\right) \approx 0.95.$$

Therefore, with a probability closed to 0.95, for n is large enough, one has :

$$|\epsilon_n| \leq 1.96\frac{\sigma}{\sqrt{n}}.$$

How to estimate the variance The previous result shows that it is crucial to estimate the standard deviation σ of the random variable. Its easy to do this by using the same samples as for the expectation. Let X be a random variable, and $(X_1, \ldots, X_N)$ a sample drawn along the law of X. We will denote by $\bar{X}_N$ the Monte-Carlo estimator of $\mathbb{E}(X)$ given by

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

A standard estimator for the variance is given by

$$\bar{\sigma}_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X}_N\right)^2,$$

$\bar{\sigma}_N^2$ is often called the empirical variance of the sample. Note that $\bar{\sigma}_N^2$ can be rewritten as

$$\bar{\sigma}_N^2 = \frac{N}{N-1} \left(\frac{1}{N} \sum_{i=1}^{N} X_i^2 - \bar{X}_N^2 \right).$$

On this last formula, it is obvious that $\bar{X}_N$ and $\bar{\sigma}_N^2$ can be computed using only $\sum_{i=1}^{N} X_i$ and $\sum_{i=1}^{N} X_i^2$.

Moreover, one can prove, when $\mathbb{E}(X^2) < +\infty$, that $\lim_{N \to +\infty} \bar{\sigma}_N^2 = \sigma^2$, almost surely, and that $\mathbb{E}\left(\bar{\sigma}_N^2\right) = \sigma^2$ (the estimator is unbiased). This leads to an (approximate) confidence interval by replacing σ par $\bar{\sigma}_n$ in the standard confidence interval. With a probability near of 0.95, $\mathbb{E}\left(X\right)$ belongs to the (random) interval given by

$$\left[\bar{X}_N - \frac{1.96 \bar{\sigma}_N}{\sqrt{N}}, \bar{X}_N + \frac{1.96 \bar{\sigma}_N}{\sqrt{N}} \right].$$

So, with very little additional computations, (we only have to compute $\bar{\sigma}_N$ on a sample already drawn) we can give an reasonable estimate of the error done by approximating $\mathbb{E}(X)$ with $\bar{X}_N$. The possibility to give an error estimate with a small numerical cost, is a very useful feature of Monte-Carlo methods.

1.2 Simulation methods of classical laws

The aim of this section is to give a short introduction to sampling methods used in finance. Our aim is *not* to be exhaustive on this broad subject (for this we refer to, e.g., [Devroye(1986)]) but to describe methods needed for the simulation of random variables widely used in finance. Thus we concentrate on Gaussian random variables and Gaussian vectors.

1.2.1 Simulation of the uniform law

In this subsection we present basic algorithms producing sequences of "pseudo random numbers", whose statistical properties mimic those of sequences of independent and identically

uniformly distributed random variables. For a recent survey on random generators see, for instance, [L'Ecuyer(1990)] and for mathematical treatment of these problems, see Niederreiter [Niederreiter(1995)] and the references therein. To generate a deterministic sequence which "looks like" independent random variables uniformly distributed on $[0,1]$, the simplest (and the most widely used) methods are congruential methods. They are defined through four integers a, b, m and U_0. The integer U_0 is the seed of the generator, m is the order of the congruence, a is the multiplicative term. A pseudo random sequence is obtained from the following inductive formula:

$$U_n = (aU_{n-1} + b) \text{ (mod. } m)$$

In practice, the seed is set to U_0 at the beginning of a program and must never be changed inside the program.

Observe that a pseudo random number generator consists of a completely deterministic algorithm. Such an algorithm produces sequences which statistically behaves (almost) like sequences of independent and identically uniformly distributed random variables. There is no theoretical criterion which ensures that a pseudo random number generator is statistically acceptable. Such a property is established on the basis of empirical tests. For example, one builds a sample from successive calls to the generator, and one then applies the Chi–square test or the Kolmogorov–Smirnov test in order to test whether one can reasonably accept the hypothesis that the sample results from independent and uniformly distributed random variables. A generator is good when no severe test has rejected that hypothesis. Good choice for a, b, m are given in [L'Ecuyer(1990)] and [Knuth(1998)]. The reader is also refered to the following web site entirely devoted to Monte-Carlo simulation : `http://random.mat.sbg.ac.at/links/`.

1.2.2 Simulation of some common laws of finance

We now explain the basic methods used to simulate laws in financial models.

Using the distribution function in simulation The simplest method of simulation relies on the use of the distribution function.

Proposition 1.2.1. *Let X be a real random variable with strictly positive and continuous density $p_X(x)$. Let F be its distribution function defined by*

$$F(x) = P(X \leq x).$$

Let U be a uniformly distributed in $[0,1]$ random variable. Then X and $F^{-1}(U)$ have the same distribution function, that is to say X and $F^{-1}(U)$ have the same law.

Démonstration : Clearly, as F^{-1} is strictly increasing, we have

$$\mathbb{P}\left(F^{-1}(U) \leq x\right) = \mathbb{P}\left(U \leq F(x)\right).$$

Now, as $F(x) \leq 1$, we have $\mathbb{P}\left(F^{-1}(U) \leq x\right) = F(x)$. So $F^{-1}(U)$ and X have the same distribution function and, hence, the same law. □

Remark 1.2.2. Simulation of an exponential law
This result can be extended to a general case (that is to say a law which does not admit a density, or with density not necessarily strictly positive). In this case we have to define the inverse F^{-1} of the increasing function F by

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

If we note that $F^{-1}(u) \leq x$ if and only if $u \leq F(x)$ the end of the previous proof remains the same.

Remark 1.2.3. Simulation of asset models with jumps uses random variables with exponential laws. The preceding proposition applies to the simulation of an exponential law of parameter $\lambda > 0$, whose density is given by

$$\lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}_+}(x).$$

In this case, a simple computation leads to $F(x) = 1 - e^{-\lambda x}$, so the equation $F(x) = u$ can be solved as $x - \frac{\log(1-u)}{\lambda}$. If U follows a uniform distribution on $[0, 1]$, $-\frac{\log(1-U)}{\lambda}$ (or $-\frac{\log(U)}{\lambda}$ follows an exponential law with parameter λ).

Remark 1.2.4. This method can also be used to sample Gaussian random variables. Of course neither the distribution function nor its inverse are exactly known but some rather good polynomial approximations can be found, e.g. , in [Abramovitz and Stegun(1970)]. This method is numerically more complex than Box-Muller method (see below) but can be used when using low discrepancy sequences to sample Gaussian random variables.

Conditional simulation using the distribution function In stratification methods, described later in this chapter, it is necessary to sample real random variable X, given that this random variable belongs to a given interval $]a, b]$. This can be easily done by using the distribution function. Let U be a random variable uniform on $[0, 1]$, F be the distribution function of X, $F(x) = \mathbb{P}(X \leq x)$ and F^{-1} be its inverse. The law of Y defined by

$$Y = F^{-1}\left(F(a) + (F(b) - F(a))U\right),$$

is equal to the conditional law of X given that $X \in]a, b]$. This can be easily proved by checking that the distribution function of Y is equal to the one of X knowing that $X \in]a, b]$.

Gaussian Law The Gaussian law with mean 0 and variance 1 on $\mathbb{R}$ is the law with the density given by

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Therefore, this distribution function of the Gaussian random variable X is given by

$$\mathbb{P}(X \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{x^2}{2}\right) dx, \ \forall z \in \mathbb{R}.$$

The most widely used simulation method of a Gaussian law is the Box-Muller method. This method is based upon the following result (See exercise 2 for a proof.).

Proposition 1.2.5. *Let U and V be two independent random variables which are uniformly distributed on $[0,1]$. Let X and Y be defined by*

$$\begin{aligned} X &= \sqrt{-2\log U}\sin(2\pi V), \\ Y &= \sqrt{-2\log U}\cos(2\pi V). \end{aligned}$$

Then X and Y are two independent Gaussian random variables with mean 0 and variance 1.

Of course, the method can be used to simulate N independent realizations of the same real Gaussian law. The simulation of the two first realizations is performed by calling a random number generator twice and by computing X and Y as above. Then the generator is called two other times to compute the corresponding two new values of X and Y, which provides two new realizations which are independent and mutually independent of the two first realizations, and so on.

Simulation of a Gaussian vector To simulate a Gaussian vector

$$X = (X^1, \ldots, X^d)$$

with zero mean and with a $d \times d$ covariance matrix $C = (c_{ij}, 1 \leq i, j \leq n)$ with $c_{ij} = \mathbb{E}(X^i X^j)$ one can proceed as follows.

C is a covariance matrix, so it is positive (since, for each $v \in \mathbb{R}^d$, $v.Cv = \mathbb{E}\left((v.X)^2\right) \geq 0$). Standard results of linear algebra prove that there exists a $d \times d$ matrix A, called a square root of C such that

$$AA^* = C,$$

where A^* is the transposed matrix of $A = (a_{ij}, 1 \leq i, j \leq n)$.

Moreover one can compute a square root of a given positive symmetric matrix by specifying that $a_{ij} = 0$ for $i < j$ (i.e. A is a lower triangular matrix). Under this hypothesis, its easy to see that A is uniquely determined by the following algorithm

$$\begin{aligned} a_{11} &:= \sqrt{c_{11}} \\ \text{For } 2 < i \leq d & \\ a_{i1} &:= \frac{c_{i1}}{a_{11}}, \end{aligned}$$

then i increasing from 2 to d,

$$\begin{aligned} a_{ii} &:= \sqrt{c_{ii} - \sum_{j=1}^{i-1} |a_{ij}|^2}, \\ \text{For } j < i \leq d & \\ a_{ij} &:= \frac{c_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}}{a_{jj}}, \\ \text{For } 1 < i < j & \\ a_{ij} &:= 0. \end{aligned}$$

This way of computing a square root of a positive symmetric matrix is known as the Cholevsky algorithm.

Now, if we assume that $G = (G^1, \dots, G^d)$ is a vector of independent Gaussian random variables with mean 0 and variance 1 (which are easy to sample as we have already seen), one can check that $Y = AG$ is a Gaussian vector with mean 0 et with covariance matrix given by $AA^* = C$. As X et Y are two Gaussian vectors with the same mean and covariance matrix, the law of X and Y are the same. This leads to the following simulation algorithm.

> Simulate the vector $(G^1, \dots, G^d)$ of *independent* Gaussian variables as explained above. Then return the vector $X = AG$.

Discrete law Consider a random variable X taking values in a finite set $\{x_k, k = 1, \dots, N\}$. The value x_k is taken with probability p_k. To simulate the law of X, one simulates a random variable U uniform on $[0, 1]$. If the value u of the trial satisfies

$$\sum_{j=0}^{k-1} p_j < u \leq \sum_{j=0}^{k} p_j,$$

one decides to return the value x_k. Clearly the random variable obtained by using this procedure follows the same law as X.

Bibliographic remark A very complete discussion on the simulation of non uniform random variables can be found in [Devroye(1986)], results and discussion on the construction of pseudo-random sequences in Knuth [Knuth(1998)].

[Ripley(2006)],[Rubinstein(1981)] and [Hammersley and Handscomb(1979)] are reference books on simulation methods. See also the survey paper by Niederreiter [Niederreiter(1995)] and the references therein, in particular these which concern nonlinear random number generators.

1.3 Variance Reduction

All the results of the preceding section show that the ratio $\sigma/\sqrt{N}$ governs the accuracy of a Monte-Carlo method with N simulations. An obvious consequence of this fact is that one always has interest to rewrite the quantity to compute as the expectation of a random variable which has a smaller variance : this is the basic idea of variance reduction techniques. For complements, we refer the reader to [Kalos and Whitlock(2008)],[Hammersley and Handscomb(1979)],[Rubinstein(1981)] or [Ripley(2006)].

Suppose that we want to evaluate $\mathbb{E}(X)$. We try to find an alternative representation for this expectation as

$$\mathbb{E}(X) = \mathbb{E}(Y) + C,$$

using a random variable Y with lower variance and C a known constant. A lot of techniques are known in order to implement this idea. This paragraph gives an introduction to some standard methods.

1.3.1 Control variates

The basic idea of control variate is to write $\mathbb{E}(f(X))$ as

$$\mathbb{E}(f(X)) = \mathbb{E}(f(X) - h(X)) + \mathbb{E}(h(X)),$$

where $\mathbb{E}(h(X))$ can be explicitly computed and $\mathrm{Var}\ (f(X)-h(X))$ is smaller than $\mathrm{Var}\ (f(X))$. In these circumstances, we use a Monte-Carlo method to estimate $\mathbb{E}(f(X) - h(X))$, and we add the value of $\mathbb{E}(h(X))$. Let us illustrate this principle by several financial examples.

Using call-put arbitrage formula for variance reduction Let S_t be the price at time t of a given asset and denote by C the price of the European call option

$$C = \mathbb{E}\left(e^{-rT}\left(S_T - K\right)_+\right),$$

and by P the price of the European put option

$$P = \mathbb{E}\left(e^{-rT}\left(K - S_T\right)_+\right).$$

There exists a relation between the price of the put and the call which does not depend on the models for the price of the asset,namely, the "call-put arbitrage formula" :

$$C - P = \mathbb{E}\left(e^{-rT}\left(S_T - K\right)\right) = S_0 - Ke^{-rT}.$$

This formula (easily proved using linearity of the expectation) can be used to reduce the variance of a call option since

$$C = \mathbb{E}\left(e^{-rT}\left(K - S_T\right)_+\right) + S_0 - Ke^{-rT}.$$

The Monte-Carlo computation of the call is then reduced to the computation of the put option.

Remark 1.3.1. For the Black-Scholes model explicit formulas for the variance of the put and the call options can be obtained. In most cases, the variance of the put option is smaller than the variance of the call since the payoff of the put is bounded whereas the payoff of the call is not. Thus, one should compute put option prices even when one needs a call prices.

Remark 1.3.2. Observe that call-put relations can also be obtained for Asian options or basket options.

For example, for Asian options, set $\bar{S}_T = \frac{1}{T}\int_0 S_s ds$. We have :

$$\mathbb{E}\left(\left(\bar{S}_T - K\right)_+\right) - \mathbb{E}\left(\left(K - \bar{S}_T\right)_+\right) = \mathbb{E}\left(\bar{S}_T\right) - K,$$

and, in the Black-Scholes model,

$$\mathbb{E}\left(\bar{S}_T\right) = \frac{1}{T}\int_0^T \mathbb{E}(S_s)ds = \frac{1}{T}\int_0^T S_0 e^{rs} ds = S_0\frac{e^{rT}-1}{rT}.$$

Basket options. A very similar idea can be used for pricing basket options. Assume that, for $i = 1, \dots, d$

$$S_T^i = x_i e^{\left(r - \frac{1}{2}\sum_{j=1}^{p} \sigma_{ij}^2\right) T + \sum_{j=1}^{p} \sigma_{ij} W_T^j},$$

where $W^1, \dots, W^p$ are independent Brownian motions. Let a_i, $1 \leq i \leq p$, be positive real numbers such that $a_1 + \cdots + a_d = 1$. We want to compute a put option on a basket

$$\mathbb{E}\left((K - X)_+\right),$$

where $X = a_1 S_T^1 + \cdots + a_d S_T^d$. The idea is to approximate

$$\frac{X}{m} = \frac{a_1 x_1}{m} e^{rT + \sum_{j=1}^{p} \sigma_{1j} W_T^j} + \cdots + \frac{a_d x_d}{m} e^{rT + \sum_{j=1}^{p} \sigma_{dj} W_T^j}$$

where $m = a_1 x_1 + \cdots + a_d x_d$, by $\frac{Y}{m}$ where Y is the log-normal random variable

$$Y = m e^{\sum_{i=1}^{d} \frac{a_i x_i}{m} \left(rT + \sum_{j=1}^{p} \sigma_{ij} W_T^j\right)}.$$

As we can compute an explicit formula for

$$\mathbb{E}\left[(K - Y)_+\right],$$

one can to use the control variate $Z = (K - Y)_+$ and sample $(K - X)_+ - (K - Y)_+$.

1.3.2 Importance sampling

Importance sampling is another variance reduction procedure. It is obtained by changing the sampling law.

We start by introducing this method in a very simple context. Suppose we want to compute

$$\mathbb{E}(g(X)),$$

X being a random variable following the density $f(x)$ on $\mathbb{R}$, then

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f(x) dx.$$

Let $\tilde{f}$ be another density such that $\tilde{f}(x) > 0$ and $\int_{\mathbb{R}} \tilde{f}(x) dx = 1$. Clearly one can write $\mathbb{E}(g(X))$ as

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} \frac{g(x) f(x)}{\tilde{f}(x)} \tilde{f}(x) dx = \mathbb{E}\left(\frac{g(Y) f(Y)}{\tilde{f}(Y)}\right),$$

where Y has density $\tilde{f}(x)$ under $\mathbb{P}$. We thus can approximate $\mathbb{E}(g(X))$ by

$$\frac{1}{n}\left(\frac{g(Y_1) f(Y_1)}{\tilde{f}(Y_1)} + \cdots + \frac{g(Y_n) f(Y_n)}{\tilde{f}(Y_n)}\right),$$

where $(Y_1, \ldots, Y_n)$ are independant copies of Y. Set $Z = g(Y)f(Y)/\tilde{f}(Y)$. We gave decreased the variance of the simulation if $\text{Var }(Z) < \text{Var }(g(X))$. It is easy to compute the variance of Z as

$$\text{Var }(Z) = \int_{\mathbb{R}} \frac{g^2(x)f^2(x)}{\tilde{f}(x)}dx - \mathbb{E}(g(X))^2.$$

From this and an easy computation it follows that if $g(x) > 0$ and $\tilde{f}(x) = g(x)f(x)/\mathbb{E}(g(X))$ then $\text{Var }(Z) = 0$! Of course this result cannot be used in practice as it relies on the exact knowledge of $\mathbb{E}(g(X))$, which is the exactly what we want to compute. Nevertheless, it leads to a heuristic approach : choose $\tilde{f}(x)$ as a good approximation of $|g(x)f(x)|$ such that $\tilde{f}(x)/\int_{\mathbb{R}} \tilde{f}(x)dx$ can be sampled easily.

An elementary financial example Suppose that G is a Gaussian random variable with mean zero and unit variance, and that we want to compute

$$\mathbb{E}\left(\phi(G)\right),$$

for some function ϕ. We choose to sample the law of $\tilde{G} = G + m$, m being a real constant to be determined carefully. We have :

$$\mathbb{E}\left(\phi(G)\right) = \mathbb{E}\left(\phi(\tilde{G})\frac{f(\tilde{G})}{\tilde{f}(\tilde{G})}\right) = \mathbb{E}\left(\phi(\tilde{G})e^{-m\tilde{G}+\frac{m^2}{2}}\right).$$

This equality can be rewritten as

$$\mathbb{E}\left(\phi(G)\right) = \mathbb{E}\left(\phi(G+m)e^{-mG-\frac{m^2}{2}}\right).$$

Suppose we want to compute a European call option in the Black and Scholes model, we have

$$\phi(G) = \left(\lambda e^{\sigma G} - K\right)_+,$$

and assume that $\lambda << K$. In this case, $\mathbb{P}(\lambda e^{\sigma G} > K)$ is very small and unlikely the option will be exercised. This fact can lead to a very large error in a standard Monte-Carlo method. In order to increase to exercise probability, we can use the previous equality

$$\mathbb{E}\left(\left(\lambda e^{\sigma G} - K\right)_+\right) = \mathbb{E}\left(\left(\lambda e^{\sigma(G+m)} - K\right)_+ e^{-mG-\frac{m^2}{2}}\right),$$

and choose $m = m_0$ with $\lambda e^{\sigma m_0} = K$, since

$$\mathbb{P}\left(\lambda e^{\sigma(G+m_0)} > K\right) = \frac{1}{2}.$$

This choice of m is certainly not optimal; however it drastically improves the efficiency of the Monte-Carlo method when $\lambda << K$ (see exercise 4 for a mathematical hint of this fact).

The multidimensional case Monte-Carlo simulations are really useful for problems with large dimension, and thus we have to extend the previous method to multidimensional setting. The ideas of this section come from [Glasserman et al.(1999)].

Let us start by considering the pricing of index options. Let σ be a $n \times d$ matrix and $(W_t, t \geq 0)$ a d-dimensional Brownian motion. Denote by $(S_t, t \geq 0)$ the solution of

$$\left\{ \begin{array}{rcl} dS_t^1 &=& S_t^1 \left(rdt + [\sigma dW_t]_1 \right) \\ & \cdots & \\ dS_t^n &=& S_t^n \left(rdt + [\sigma dW_t]_n \right) \end{array} \right.$$

where $[\sigma dW_t]_i = \sum_{j=1}^d \sigma_{ij} dW_t^j$.

Moreover, denote by I_t the value of the index

$$I_t = \sum_{i=1}^n a_i S_t^i,$$

where $a_1, \ldots, a_n$ is a given set of positive numbers such that $\sum_{i=1}^n a_i = 1$. Suppose that we want to compute the price of a European call option with payoff at time T given by

$$h = (I_T - K)_+ .$$

As

$$S_T^i = S_0^i \exp\left(\left(r - \frac{1}{2} \sum_{j=1}^d \sigma_{ij}^2 \right) T + \sum_{j=1}^d \sigma_{ij} W_T^j \right),$$

there exists a function ϕ such that

$$h = \phi\left(G_1, \ldots, G_d\right),$$

where $G_j = W_T^j / \sqrt{T}$. The price of this option can be rewritten as

$$\mathbb{E}\left(\phi(G)\right)$$

where $G = (G_1, \ldots, G_d)$ is a d-dimensional Gaussian vector with unit covariance matrix.

As in the one dimensional case, it is easy (by a change of variable) to prove that, if $m = (m_1, \ldots, m_d)$,

$$\mathbb{E}\left(\phi(G)\right) = \mathbb{E}\left(\phi(G+m) e^{-m.G - \frac{|m|^2}{2}} \right), \tag{1.2}$$

where $m.G = \sum_{i=1}^d m_i G_i$ and $|m|^2 = \sum_{i=1}^d m_i^2$. In view of 1.2, the variance $V(m)$ of the random variable

$$X_m = \phi(G+m) e^{-m.G - \frac{|m|^2}{2}}$$

is

$$\begin{array}{rcl} V(m) &=& \mathbb{E}\left(\phi^2(G+m) e^{-2m.G - |m|^2} \right) - \mathbb{E}\left(\phi(G)\right)^2, \\ &=& \mathbb{E}\left(\phi^2(G+m) e^{-m.(G+m) + \frac{|m|^2}{2}} e^{-m.G - \frac{|m|^2}{2}} \right) - \mathbb{E}\left(\phi(G)\right)^2, \\ &=& \mathbb{E}\left(\phi^2(G) e^{-m.G + \frac{|m|^2}{2}} \right) - \mathbb{E}\left(\phi(G)\right)^2. \end{array}$$

Exercise 5 provides an example of the use of this formula to reduce variance. The reader is refered to [Glasserman et al.(1999)] for an almost optimal way to choose the parameter m based on this representation.

1.3.3 Antithetic variables

The use of antithetic variables is widespread in Monte-Carlo simulation. This technique is often efficient but its gains are less dramatic than other variance reduction techniques.

We begin by considering a simple and instructive example. Let

$$I = \int_0^1 g(x)dx.$$

If U follows a uniform law on the interval $[0,1]$, then $1-U$ has the same law as U, and thus

$$I = \frac{1}{2}\int_0^1 (g(x)+g(1-x))dx = \mathbb{E}\left(\frac{1}{2}(g(U)+g(1-U))\right).$$

Therefore one can draw n independent random variables $U_1,\dots,U_n$ following a uniform law on $[0,1]$, and approximate I by

$$\begin{aligned} I_{2n} &= \tfrac{1}{n}\left(\tfrac{1}{2}(g(U_1)+g(1-U_1))+\cdots+\tfrac{1}{2}(g(U_n)+g(1-U_n))\right) \\ &= \tfrac{1}{2n}\left(g(U_1)+g(1-U_1)+\cdots+g(U_n)+g(1-U_n)\right). \end{aligned}$$

We need to compare the efficiency of this Monte-Carlo method with the standard one with $2n$ drawings

$$\begin{aligned} I^0_{2n} &= \tfrac{1}{2n}\left(g(U_1)+g(U_2)+\cdots+g(U_{2n-1})+g(U_{2n})\right) \\ &= \tfrac{1}{n}\left(\tfrac{1}{2}(g(U_1)+g(U_2))+\cdots+\tfrac{1}{2}(g(U_{2n-1})+g(U_{2n}))\right). \end{aligned}$$

We will now compare the variances of I_{2n} and I^0_{2n}. Observe that in doing this we assume that most of numerical work relies in the evaluation of f and the time devoted to the simulation of the random variables is negligible. This is often a realistic assumption.

An easy computation shows that the variance of the standard estimator is

$$\text{Var}\,(I^0_{2n}) = \frac{1}{2n}\text{Var}\,\left(g(U_1)\right),$$

whereas

$$\begin{aligned} \text{Var}\,(I_{2n}) &= \frac{1}{n}\text{Var}\,\left(\frac{1}{2}(g(U_1)+g(1-U_1))\right) \\ &= \frac{1}{4n}\left(\text{Var}\,(g(U_1))+\text{Var}\,(g(1-U_1))+2\text{Cov}\,(g(U_1),g(1-U_1))\right) \\ &= \frac{1}{2n}\left(\text{Var}\,(g(U_1))+\text{Cov}\,(g(U_1),g(1-U_1))\right). \end{aligned}$$

Obviously, $\text{Var}\,(I_{2n}) \leq \text{Var}\,(I^0_{2n})$ if and only if $\text{Cov}\,(g(U_1),g(1-U_1)) \leq 0$. One can prove that if f is a monotonic function this is always true (see 6 for a proof) and thus the Monte-Carlo method using antithetic variables is better than the standard one.

This ideas can be generalized in dimension greater than 1, in which case we use the transformation

$$(U_1,\dots,U_d) \to (1-U_1,\dots,1-U_d).$$

More generaly, if X is a random variable taking its values in $\mathbb{R}^d$ and T is a transformation of $\mathbb{R}^d$ such that the law of $T(X)$ is the same as the law of X, we can construct an antithetic method using the equality

$$\mathbb{E}(g(X)) = \frac{1}{2}\mathbb{E}\left(g(X) + g(T(X))\right).$$

Namely, if $(X_1, \ldots, X_n)$ are independent and sampled along the law of X, we can consider the estimator

$$I_{2n} = \frac{1}{2n}\left(g(X_1) + g(T(X_1)) + \cdots + g(X_n) + g(T(X_n))\right)$$

and compare it to

$$I^0_{2n} = \frac{1}{2n}\left(g(X_1) + g(X_2)) + \cdots + g(X_{2n-1}) + g(X_{2n})\right).$$

The same computations as before prove that the estimator I_{2n} is better than the crude one if and only if $\mathrm{Cov}\ (g(X), g(T(X))) \leq 0$. We now show a few elementary examples in finance.

A toy financial example. Let G be a standard Gaussian random variable and consider the call option

$$\mathbb{E}\left(\left(\lambda e^{\sigma G} - K\right)_+\right).$$

Clearly the law of $-G$ is the same as the law of G, and thus the function T to be considered is $T(x) = -x$. As the payoff is increasing as a function of G, the following antithetic estimator certainly reduces the variance :

$$I_{2n} = \frac{1}{2n}\left(g(G_1) + g(-G_1) + \cdots + g(G_n) + g(-G_n)\right),$$

where $g(x) = \left(\lambda e^{\sigma x} - K\right)_+$.

Antithetic variables for path-dependent options. Consider the path dependent option with payoff at time T

$$\psi\left(S_s, s \leq T\right),$$

where $(S_t, t \geq 0)$ is the lognormal diffusion

$$S_t = x \exp\left((r - \frac{1}{2}\sigma^2)t + \sigma W_t\right).$$

As the law of $(-W_t, t \geq 0)$ is the same as the law of $(W_t, t \geq 0)$ one has

$$\mathbb{E}\left(\psi\left(x \exp\left((r - \frac{1}{2}\sigma^2)s + \sigma W_s\right), s \leq T\right)\right)$$
$$= \mathbb{E}\left(\psi\left(x \exp\left((r - \frac{1}{2}\sigma^2)s - \sigma W_s\right), s \leq T\right)\right),$$

and, for appropriate functionals ψ, the antithetic variable method may be efficient.

1.3.4 Stratification methods

These methods are widely used in statistics (see [Cochran(1953)]). Assume that we want to compute the expectation

$$I = \mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x)f(x)dx,$$

where X is a $\mathbb{R}^d$ valued random variable with density $f(x)$.

Let $(D_i, 1 \leq i \leq m)$ be a partition of $\mathbb{R}^d$. I can be expressed as

$$I = \sum_{i=1}^{m} \mathbb{E}(\mathbf{1}_{X\in D_i} g(X)) = \sum_{i=1}^{m} \mathbb{E}(g(X)|X \in D_i)\mathbb{P}(X \in D_i),$$

where

$$\mathbb{E}(g(X)|X \in D_i) = \frac{\mathbb{E}(\mathbf{1}_{X\in D_i} g(X))}{\mathbb{P}(X \in D_i)}.$$

Note that $\mathbb{E}(g(X)|X \in D_i)$ can be interpreted as $\mathbb{E}(g(X^i))$ where X^i is a random variable whose law is the law of X conditioned by X belongs to D_i, whose density is

$$\frac{1}{\int_{D_i} f(y)dy} \mathbf{1}_{x\in D_i} f(x)dx.$$

Remark 1.3.3. The random variable X^i is easily simulated using an acceptance rejection procedure. But this method is clearly unefficient when $\mathbb{P}(X \in D_i)$ is small.

When the numbers $p_i = \mathbb{P}(X \in D_i)$ can be explicitly computed, one can use a Monte-Carlo method to approximate each conditional expectation $I_i = \mathbb{E}(g(X)|X \in D_i)$ by

$$\tilde{I}_i = \frac{1}{n_i} \left(g(X_1^i) + \cdots + g(X_{n_i}^i) \right),$$

where $(X_1^i, \ldots, X_{n_i}^i)$ are independent copies of X^i. An estimator $\tilde{I}$ of I is then

$$\tilde{I} = \sum_{i=1}^{m} p_i \tilde{I}_i.$$

Of course the samples used to compute $\tilde{I}_i$ are supposed to be independent and so the variance of $\tilde{I}$ is

$$\sum_{i=1}^{m} p_i^2 \frac{\sigma_i^2}{n_i},$$

where σ_i^2 be the variance of $g(X^i)$.

Fix the total number of simulations $\sum_{i=1}^{m} n_i = n$. This minimization the variance above, one must choose

$$n_i = n \frac{p_i \sigma_i}{\sum_{i=1}^{m} p_i \sigma_i}.$$

For this values of n_i, the variance of $\tilde{I}$ is given in this case by

$$\frac{1}{n} \left(\sum_{i=1}^{m} p_i \sigma_i \right)^2.$$

Note that this variance is smaller than the one obtained without stratification. Indeed,

$$\begin{aligned}\text{Var } (g(X)) &= \mathbb{E}\left(g(X)^2\right) - \mathbb{E}\left(g(X)\right)^2 \\ &= \sum_{i=1}^{m} p_i \mathbb{E}\left(g^2(X)|X \in D_i\right) - \left(\sum_{i=1}^{m} p_i \mathbb{E}\left(g(X)|X \in D_i\right)\right)^2 \\ &= \sum_{i=1}^{m} p_i \text{Var } (g(X)|X \in D_i) + \sum_{i=1}^{m} p_i \mathbb{E}\left(g(X)|X \in D_i\right)^2 \\ &\qquad - \left(\sum_{i=1}^{m} p_i \mathbb{E}\left(g(X)|X \in D_i\right)\right)^2.\end{aligned}$$

Using the convexity inequality for x^2 we obtain $\left(\sum_{i=1}^m p_i a_i\right)^2 \leq \sum_{i=1}^m p_i a_i^2$ if $\sum_{i=1}^m p_i = 1$, and the inequality

$$\text{Var } (g(X)) \geq \sum_{i=1}^{m} p_i \text{Var } (g(X)|X \in D_i) \geq \left(\sum_{i=1}^{m} p_i \sigma_i\right)^2,$$

follows.

Remark 1.3.4. The optimal stratification involves the σ_i's which are seldom explicitly known. So one needs to estimate these σ_i's by Monte-Carlo simulations.

Moreover note that arbitrary choices of n_i may *increase* the variance. Common way to circumvent this difficulty is to choose

$$n_i = np_i.$$

The corresponding variance

$$\frac{1}{n}\sum_{i=1}^{m} p_i \sigma_i^2,$$

is always smaller than the original one as $\sum_{i=1}^m p_i \sigma_i^2 \leq \text{Var } (g(X))$. This choice is often made when the probabilities p_i can be computed. For more considerations on the choice of the n_i and also, for hints on suitable choices of the sets D_i, see [Cochran(1953)].

A toy example in finance In the standard Black and Scholes model the price of a call option is

$$\mathbb{E}\left(\left(\lambda e^{\sigma G} - K\right)_+\right).$$

It is natural to use the following strata for G : either $G \leq d = \frac{\log(K/\lambda)}{\sigma}$ or $G > d$. Of course the variance of the stratum $G \leq d$ is equal to zero, so if you follow the optimal choice of number, you do not have to simulate points in this stratum : all points have to be sampled in the stratum $G \geq d$! This can be easily done by using the (numerical) inverse of the distribution function of a Gaussian random variable.

Of course, one does not need Monte-Carlo methods to compute call options for the Black and Scholes models; we now consider a more convincing example.

Basket options Most of what follows comes from [Glasserman et al.(1999)]. The computation of an European basket option in a multidimensional Black-Scholes model can be expressed as

$$\mathbb{E}(h(G)),$$

for some function h and for $G = (G_1, \ldots, G_n)$ a vector of independent standard Gaussian random variables. Choose a vector $u \in \mathbb{R}^n$ such that $|u| = 1$ (note that $< u, G >= u_1 G_1 + \cdots + u_n G_n$ is also a standard Gaussian random variable.). Then choose a partition $(B_i, 1 \leq i \leq n)$ of $\mathbb{R}$ such that

$$\mathbb{P}(< u, G > \in B_i) = \mathbb{P}(G_1 \in B_i) = 1/n.$$

This can be done by setting

$$B_i =]N^{-1}((i-1)/n), N^{-1}(i/n)],$$

where N is the distribution function of a standard Gaussian random variable and N^{-1} is its inverse. We then define the strata by setting

$$D_i = \{< u, x > \in B_i\}.$$

In order to implement our stratification method we need to solve two simulation problems

- sample a Gaussian random variable $< u, G >$ given that $< u, G >$ belongs to B_i,
- sample a new vector G knowing the value $< u, G >$.

The first problem is easily solved since the law of

$$N^{-1}\left(\frac{i-1}{N} + \frac{U}{N}\right), \tag{1.3}$$

is precisely the law a standard Gaussian random variable conditioned to be in B_i (see page 6).

To solve the second point, observe that

$$G - < u, G > u$$

is a Gaussian vector independent of $< u, G >$ with covariance matrix $I - u \otimes u'$ (where $u \otimes u'$ denotes the matrix defined by $(u \otimes u')_{ij} = u_i u_j$). Let Y be a copy of the vector G. Obviously $Y - < u, Y > u$ is independent of G and has the same law as $G - < u, G > u$. So

$$G =< u, G > u + G - < u, G > u \text{ and } < u, G > u + Y - < u, Y > u,$$

have the same probability law. This leads to the following simulation method of G given $< u, G >= \lambda$:

- sample n independent standard Gaussian random variables Y^i,
- set $G = \lambda u + Y - < u, Y > u$.

To make this method efficient, the choice of the vector u is crucial : an almost optimal way to choose the vector u can be found in [Glasserman et al.(1999)].

1.3.5 Mean value or conditioning

This method uses the well known fact that conditioning reduces the variance. Indeed, for any square integrable random variable Z, we have

$$\mathbb{E}(Z) = \mathbb{E}(\mathbb{E}(Z|Y)),$$

where Y is any random variable defined on the same probability space as Z. It is well known that $\mathbb{E}(Z|Y)$ can be written as

$$\mathbb{E}(Z|Y) = \phi(Y),$$

for some measurable function ϕ. Suppose in addition that Z is square integrable. As the conditional expectation is a L^2 projection

$$\mathbb{E}\left(\phi(Y)^2\right) \leq \mathbb{E}(Z^2),$$

and thus $\text{Var}\,(\phi(Y)) \leq \text{Var}\,(Z)$.

Of course the practical efficiency of simulating $\phi(Y)$ instead of Z heavily relies on an explicit formula for the function ϕ. This can be achieved when $Z = f(X, Y)$, where X and Y are independent random variables. In this case, we have

$$\mathbb{E}(f(X, Y)|Y) = \phi(Y),$$

where $\phi(y) = \mathbb{E}(f(X, y))$.

A basic example. Suppose that we want to compute $\mathbb{P}\left(X \leq Y\right)$ where X and Y are independent random variables. This situation occurs in finance, when one computes the hedge of an exchange option (or the price of a digital exchange option).

Using the preceding, we have

$$\mathbb{P}\left(X \leq Y\right) = \mathbb{E}\left(F(Y)\right),$$

where F is the distribution function of X. The variance reduction can be significant, especially when the probability $\mathbb{P}\left(X \leq Y\right)$ is small.

1.4 Low discrepancy sequences

Using sequences of points "more regular" than random points may sometimes improve Monte-Carlo methods. We look for deterministic sequences $(x_i, i \geq 1)$ such that

$$\int_{[0,1]^d} f(x)dx \approx \frac{1}{n}(f(x_1) + \cdots + f(x_n)),$$

for all function f in a large enough set.

When the considered sequence is deterministic, the method is called a *quasi Monte-Carlo* method. One can find sequences such that the speed of convergence of the previous approximation is of the order $K\frac{\log(n)^d}{n}$ (when the function f is regular enough). Such a sequence is called a "low discrepancy sequence".

We give now a mathematical definition of a uniformly distributed sequence. In this definition the notion of discrepancy is involved. By definition, if x and y are two points in $[0,1]^d$, $x \leq y$ if and only if $x_i \leq y_i$, for all $1 \leq i \leq d$.

Definition 1.4.1. *A sequence $(x_n)_{n\geq 1}$ is said to be uniformly distributed on $[0,1]^d$ if one of the following equivalent properties is fulfilled :*

1. *For all $y = (y_1, \cdots, y_d) \in [0,1]^d$:*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{x_k \in [0,y]} \prod_{i=1}^{d} y^i = \text{Volume}([0,y]),$$

   where $[0,y] = \{z \in [0,1]^d, 0 \leq z \leq y\}$.

2. *Let $D_n^*(x) = \sup_{y\in[0,1]^d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{x_k\in[0,y]} - \text{Volume}([0,y]) \right|$ be the discrepancy of the sequence, then*

$$\lim_{n\to+\infty} D_n^*(x) = 0,$$

3. *For every bounded continuous function f on $[0,1]^d$*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k) = \int_{[0,1]^d} f(x)dx,$$

Remark 1.4.1.

- If $(U_n)_{n\geq 1}$ is a sequence of independent random variables with uniform law on $[0,1]$, the random sequence

$$(U_n(\omega), n \geq 1),$$

  is almost surely uniformly distributed. Moreover, we have an iterated logarithm law for the discrepancy, namely,

$$\limsup_n \sqrt{\frac{2n}{\log(\log n)}} D_n^*(U) = 1 \text{ a.s.}$$

- The discrepancy of any infinite sequence satisfies the following property

$$D_n^* > C_d \frac{(\log n)^{\max(\frac{d-1}{2},1)}}{n} \text{ for an infinite number of values of } n,$$

  where C_d is a constant which depends on d only. This result is known as the Roth theorem (see [Roth(1954)]).

- It is possible to construct d-dimensional sequences with discrepancies bounded by $(\log n)^d/n$. We will see later in this section some examples of such sequences. Note that, using the Roth theorem, these sequences are almost optimal. These sequences are, in principle, asymptotically better than random numbers.

  In practice we use a number of drawing between 10^3 and 10^8 and, in this case, the best known sequences are not clearly better than random numbers in term of discrepancy. This is especially true in large dimension (greater than 100).

The discrepancy allows one to give an estimation of the approximation error

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) - \int_{[0,1]^d} f(x)dx,$$

when f has a finite variation in the sense of Hardy and Krause. This estimate is known as the Koksma-Hlawka inequality.

Proposition 1.4.2 (Koksma-Hlawka inequality)**.** *Let g be a finite variation function in the sense of Hardy and Krause and denote by $V(g)$ its variation. Then for $n \geq 1$*

$$\left| \frac{1}{N}\sum_{k=1}^{N} g(x_k) - \int_{[0,1]^d} g(u)du \right| \leq V(g) D_N^*(x).$$

Remark 1.4.3. This result is very different from the central limit theorem used for random sequences, which leads to a confidence interval for a given probability. Here, this estimation is deterministic. This can be seen as a useful property of low discrepancy sequences, but this estimation involves $V(g)$ and $D_N^*(x)$ and both of these quantities are extremely hard to estimate in practice. So, the theorem gives in most cases a large overestimation of the real error (very often, too large to be useful) .

For a general definition of finite variation function in the sense of Hardy and Krause see [Niederreiter(1992)]. In dimension 1, this notion coincides with the notion of a function with finite variation in the classical sense. In dimension d, when g is d times continuously differentiable, the variation of $V(g)$ is given by

$$\sum_{k=1}^{d} \sum_{1\leq i_1<\cdots<i_k\leq d} \int_{\left\{ \begin{array}{l} x\in[0,1]^d \\ x_j = 1, \text{for } j\neq i_1,\ldots,i_k \end{array} \right\}} \left| \frac{\partial^k g(x)}{\partial x_{i_1}\cdots\partial x_{i_k}} \right| dx_{i_1}\ldots dx_{i_k}.$$

When the dimension d increases, a function with finite variation has to be smoother. For instance, the set function $\mathbf{1}_{f(x_1,\ldots,x_d)>\lambda}$ has an infinite variation when $d \geq 2$. Moreover, most of the standard option payoffs for basket options such as

$$(a_1x_1 + \cdots + a_dx_d - K)_+ \text{ or } (K - (a_1x_1 + \cdots + a_dx_d))_+$$

do not have finite variation when $d \geq 3$ (see [Ksas(2000)] for a proof).

Note that the efficiency of a law discrepancy method depends not only on the representation of the expectation, but also on the way the random variable is simulated. Moreover, the method chosen can lead to functions with infinite variation, even when the variance is bounded.

For instance, assume that we want to compute $\mathbb{E}(f(G))$, where G is a real random variable and f is a function such that $\text{Var}\ (f(G)) < +\infty$, f is increasing, $f(-\infty) = 0$ and $f(+\infty) = +\infty$. Assume that we simulate along the law of G using the inverse of the distribution function denoted by $N(x)$. For the sake of simplicity, we will assume that N is differentiable and strictly increasing. If U is a random variable drawn uniformly on $[0,1]$, we have

$$\mathbb{E}(f(G)) = \mathbb{E}\left(f(N^{-1}(U)\right) = \mathbb{E}(g(U)).$$

In order to use the Koksma-Hlawka inequality we need to compute the variation of g. But

$$\begin{aligned} V(g) &= \int_0^1 |g'|(u)du \\ &= \int_0^1 f'(N^{-1}(u))dN^{-1}(u) \\ &= \int_{\mathbb{R}} f'(x)dx = f(+\infty) - f(-\infty) = +\infty. \end{aligned}$$

An example in finance is given by the call option where

$$f(G) = \left(\lambda e^{\sigma G} - K\right)_+,$$

and G is a standard Gaussian random variable. Of course, it is easy in this case to solve this problem by first computing the price of the put option and then by using the call-put arbitrage relation to retrieve the call price.

We will now give examples of some of the most widely used low discrepancy sequences in finance. For other examples and an exhaustive and rigorous presentation of this subject see [Niederreiter(1992)].

The Van Der Corput sequence Let p be an integer, $p \geq 2$ and n a positive integer. We denote by $a_0, a_1, \ldots, a_r$ the p-adic decomposition of n, that is to say the unique set of integers a_i such that $0 \leq a_i < p$ for $0 \leq i \leq r$ and $a_r > 0$ with

$$n = a_0 + a_1 p + \cdots + a_r p^r.$$

Using standard notations, n can be writen as

$$n = a_r a_{r-1} \ldots a_1 a_0 \text{ in base } p.$$

The Van Der Corput sequence in base p is given by

$$\phi_p(n) = \frac{a_0}{p} + \cdots + \frac{a_r}{p^r}.$$

The definition of $\phi_p(n)$ can be rewriten as follows

$$\text{if } n = a_r a_{r-1} \ldots a_0 \text{ then } \phi_p(n) = 0, a_0 a_2 \ldots a_r,$$

where $0, a_0 a_2 \ldots a_r$ denotes the $p-$adic decomposition of a number.

Halton sequences. Halton sequences are multidimensional generalizations of Van Der Corput sequence. Let $p_1, \cdots, p_d$ be the first d prime numbers. The Halton sequence is defined by

$$x_n^d = (\phi_{p_1}(n), \cdots, \phi_{p_d}(n)) \tag{1.4}$$

for an integer n and where $\phi_{p_i}(n)$ is the Van Der Corput sequence in base p_i.

One can prove that the discrepancy of a $d-$dimensional Halton sequence can be estimated by

$$D_n^* \leq \frac{1}{n} \prod_{i=1}^{d} \frac{p_i \log(p_i n)}{\log(p_i)}.$$

Faure sequence. These sequences are defined in [Faure(1981)] and [Faure(1982)]. The Faure sequence in dimension d is defined as follows. Let r be an odd integer greater than d. Now define a function T on the set of numbers x such that

$$x = \sum_{k \geq 0} \frac{a_k}{r^{k+1}},$$

where the sum is finite and each a_k belongs to $\{0, \ldots, r-1\}$, by

$$T(x) = \sum_{k \geq 0} \frac{b_k}{r^{k+1}},$$

where

$$b_k = \sum_{i \geq k} \left(\begin{array}{c} i \\ k \end{array} \right) a_i \bmod r,$$

and $\left(\begin{array}{c} i \\ k \end{array} \right)$ denote the binomial coefficients. The Faure sequence is then defined as follows

$$x_n = \left(\phi_r(n-1), T(\phi_r(n-1)), \cdots, T^{d-1}(\phi_r(n-1)) \right), \tag{1.5}$$

where $\phi_r(n)$ is the Van Der Corput sequence of basis r. The discrepancy of this sequence is bounded by $C \frac{\log(n)^d}{n}$.

Irrational translation of the torus These sequences are defined by

$$x_n = (\{n\alpha_1\}, \ldots, \{n\alpha_d\}), \tag{1.6}$$

where $\{x\}$ is the fractional part of the number x and $\alpha = (\alpha_1, \cdots, \alpha_d)$ is a vector of real numbers such that $(1, \alpha_1, \cdots, \alpha_d)$ is a free family on $\mathbb{Q}$. This is equivalent to say that there is no linear relation with integer coefficients $(\lambda_i, i = 0, \ldots, d)$ such that

$$\lambda_0 + \lambda_1 \alpha_1 + \cdots + \lambda_d \alpha_d = 0.$$

Note that this condition implies that the α_i are irrational numbers.

One convenient way to choose such a family is to define α by

$$\left(\sqrt{p_1}, \cdots, \sqrt{p_d} \right),$$

where $p_1, \ldots, p_d$ are the d first prime numbers. See [Pagès and Xiao(1997)] for numerical experiments on this sequence.

Sobol sequence ([Sobol'(1967)]) One of the most used low discrepancy sequences is the Sobol sequence. This sequence uses the binary decomposition of a number n

$$n = \sum_{k \geq 1} a_k(n) 2^{k-1},$$

where the $a_k(n) \in \{0, 1\}$. Note that $a_k(n) = 0$, for k large enough.

First choose a polynomial of degree q with coefficient in $\mathbb{Z}/2\mathbb{Z}$

$$P = \alpha_0 + \alpha_1 X + \cdots + \alpha_q X^q,$$

such that $\alpha_0 = \alpha_q = 1$. The polynomial P is supposed to be irreducible and primitive in $\mathbb{Z}/2\mathbb{Z}$. See [Roman(1992)] for definitions and appendix A.4 of this book for an algorithm for computing such polynomials (a table of (some) irreducible polynomials is also available in this book and algorithm for testing the primitivity of a polynomial is available in `Maple`).

Choose an arbitrary vector of $(M_1, \ldots, M_q) \in \mathbb{N}^q$, such that M_k is odd and less than 2^k. Define M_n, for $n > q$ by

$$M_n = \oplus_{i=1}^{q} 2^i \alpha_i M_{n-i} \oplus M_{k-q},$$

where $\oplus$ is defined by

$$m \oplus n = \sum_{k \geq 0} (a_k(m) \text{ XOR } a_k(n)) 2^k,$$

and XOR is the bitwise operator defined by

$$a\text{XOR}b = (a + b) \text{ mod. } 2.$$

A direction sequence $(V_k, k \geq 0)$ of real numbers is then defined by

$$V_k = \frac{M_k}{2^k},$$

and a one dimensional Sobol sequence x_n, by

$$x_n = \oplus_{k \geq 0} a_k(n) V_k,$$

if $n = \sum_{k \geq 1} a_k(n) 2^{k-1}$, A multidimensional sequence can be constructed by using different polynomials for each dimension.

A variant of the Sobol sequence can be defined using a "Gray code". For a given integer n, we can define a Gray code of n, $(b_k(n), k \geq 0)$, by the binary decomposition of $G(n) = n \oplus [n/2]$

$$n \oplus [n/2] = \sum_{k \geq 0} b_k(n) 2^k.$$

Note that the function G is bijective from $\{0, \ldots, 2^N - 1\}$ to itself. The main interest of Gray codes is that the binary representation of $G(n)$ and $G(n+1)$ differ in exactly one bit. The variant proposed by Antonov et Salev (see [Antonov and Saleev(1980)]) is defined by

$$x_n = b_1(n) V_1 \oplus \cdots \oplus b_r(n) V_r.$$

For an exhaustive study of the Sobol sequence, see [Sobol'(1967)] and [Sobol'(1976)]. A program allowing to generate some Sobol sequences for small dimensions can be found in [Press et al.(1992)], see also [Fox(1988)]. Empirical studies indicate that Sobol sequences are among the most efficient low discrepancy sequences (see [Fox et al.(1992)] and [Radovic et al.(1996)] for numerical comparisons of sequences).

1.5 Monte–Carlo methods for pricing American options.

We present some recent Monte–Carlo algorithms devoted to the pricing of American options. This part follows the work done by Pierre Cohort (see [Cohort(2001)]). We give the description of two class of algorithm. The first class use an approximation of conditional expectation (Longstaff and Schwartz [Longstaff and Schwartz(1998)] and Tsitsiklis and VanRoy [Tsitsiklis and Roy(2000)]). The second one is based on an approximation a the underlying Markov chain (quantization algorithms [Pages and Bally(2000), Pages et al.(2001)], Broadie and Glassermann [Broadie and Glassermann(1997a), Broadie and Glassermann(1997b)] and Barraquand and Martineau [Barraquand and Martineau(1995)]).

Introduction

We assume that an American option can only be exercised at a given set of time $(t_0, \ldots, t_N = T)$. In this case we say that the option is a Bermudean option.

The asset price at time t_n is supposed to be given by a Markov chain $(X_n, 0 \leq n \leq N)$ with a transition matrix between n and $n+1$ given by $P_n(x, dy)$. This means that for every bounded f

$$\mathbb{E}\left(f(X_{n+1})|\mathcal{F}_n\right) = P_n f(X_n) := \int f(y) P_n(X_n, dy),$$

where $\mathcal{F}_n = \sigma\left(X_k, k \leq n\right)$.

We assume that the payoff at time t_n is given by $\varphi\left(X_n\right)$ where φ is a bounded function and that the actualization between time t_j and t_{j+1} can be expressed as $1/(1+R_j)$, R_j being a constant positive interest rate between time t_j and t_{j+1}. This allows us to define an actualization coefficient B between arbitrary time j and k, when $j < k$, by setting

$$B(j,k) = \prod_{l=j}^{k-1} \frac{1}{1+R_l}.$$

By convention we set $B(j,j) = 1$. Note that $B(j, j+1) = 1/(1+R_j)$. With these notations, the price at time 0 of this option is

$$Q_0 = \sup_{\tau \in \mathcal{T}_{0,N}} \mathbb{E}\left(B\left(0,\tau\right)\varphi\left(X_\tau\right)\right) \tag{1.7}$$

where $\mathcal{T}_{j,N}$ is the set of $\mathcal{F}_n$-stopping times taking values in $\{j, \ldots, N\}$.

Remark 1.5.1. For the d-dimensional Black and Scholes model the assets prices $(S_t^{x,i}, 1 \leq i \leq d)$ can be written as

$$S_t^{x,i} = x_i e^{\left(r - \frac{1}{2}\sum_{1\leq j\leq p}\sigma_{ij}^2\right)t + \sum_{1\leq j\leq p}\sigma_{ij}W_t^j}$$

where $(W_t, t \geq 0)$ is a p–dimensional Brownian motion, σ is a $d \times p$ matrix, x belongs to $\mathbb{R}^d$, $\sigma_i^2 = \sum_{1\leq j\leq p}\sigma_{ij}^2$ and $r > 0$ is the risk-less interest rate. So P_n is defined by

$$P_n f(x) = \mathbb{E}\left(f(S^x_{t_{n+1}-t_n})\right).$$

Moreover, we have

$$1 + R_n = \exp(r(t_{n+1} - t_n)).$$

Standard theory of optimal stopping proves that $Q_0 = u(0, X_0)$, where u is the solution of the dynamic programming algorithm

$$\begin{cases} u(N, x) = \varphi(x) \\ u(j, x) = \max\left(\varphi(x), \tilde{P}_j u(j+1, x)\right), \quad 0 \leq j \leq N-1. \end{cases} \tag{1.8}$$

and $\tilde{P}_j(x, dy)$ is the discounted transition kernel of the Markov chain $(X_j, j = 0, \ldots, N)$ between j and $j+1$ given by

$$\tilde{P}_j f(x) = B(j, j+1) P_j f(x).$$

The stopping time

$$\tau^* = \inf\left\{j \geq 0, u(j, X_j) = \varphi\left(X_j\right)\right\}. \tag{1.9}$$

is optimal, that is to say

$$Q_0 = u(0, X_0) = \mathbb{E}\left(B\left(0, \tau^*\right) \varphi\left(X_{\tau^*}\right)\right).$$

Moreover, the price Q_j of this American option at time j given by

$$Q_j = \sup_{\tau \in \mathcal{T}_{j,N}} \mathbb{E}\left(B\left(j, \tau\right) \varphi\left(X_\tau\right) | \mathcal{F}_j\right),$$

can be computed as $u(j, X_j)$ and and optimal stopping time at time j is given by τ_j^* where

$$\tau_j^* = \inf\left\{i \geq j; u(i, X_i) = \varphi(X_i)\right\}.$$

The Longstaff and Schwartz algorithm.

This algorithm approximate the optimal stopping time τ^* on M given paths by random variables $\tau^{(m)}$ depending on the path m and then estimate the price Q_0 according to a Monte-Carlo formula

$$Q_0^M = \frac{1}{M} \sum_{1 \leq i \leq M} B(0, \tau^{(m)}) \varphi\left(X_{\tau^{(m)}}^{(m)}\right). \tag{1.10}$$

The construction of $(\tau^m, 1 \leq m \leq M)$ is presented according to the following steps. First we write the dynamic programming algorithm directly on the optimal stopping time τ^*. Then we present the Longstaff and Schwartz approximation of this recursion. Finally we give a Monte-Carlo version, leading to $\tau^{(m)}$.

A dynamical programming algorithm for optimal stopping time. The main feature of the Longstaff and Schwartz algorithm is to use a dynamic principle for optimal stopping times rather than for the value function. Note that that τ^* can be computed using the sequence $(\tau_j^*, 0 \leq j \leq N)$ defined by the backward induction

$$\begin{cases} \tau_N^* = N, \\ \tau_j^* = j \mathbf{1}_{\varphi(X_j) \geq u(j, X_j)} + \tau_{j+1}^* \mathbf{1}_{\varphi(X_j) < u(j, X_j)}. \end{cases} \tag{1.11}$$

Note that

$$\begin{aligned}\{\varphi(X_j) < u(j, X_j)\} &= \left\{\varphi(X_j) < \tilde{P}_j u(j+1, X_j))\right\} \\ &= \left\{\varphi(X_j) < \mathbb{E}\left(B(j, \tau^*_{j+1})\varphi(X_{\tau^*_{j+1}})\Big|X_j\right)\right\}.\end{aligned} \tag{1.12}$$

With this definition for τ^*_j, it is easy to check recursively that

$$\tau^*_j = \min\left\{i \geq j, \varphi\left(X_i\right) = u(i, X_i)\right\}. \tag{1.13}$$

Thus τ^*_j (and thus $\tau^* = \tau^*_0$) are optimal stopping times at time j. In order to estimate τ^* we have to find a way to approximate the conditional expectation

$$\mathbb{E}\left(B\left(j, \tau^*_{j+1}\right)\varphi\left(X_{\tau^*_{j+1}}\right)|X_j\right). \tag{1.14}$$

The basic idea of Longstaff and Schwartz is to introduce a least square regression method to perform this approximation.

The regression method We denote by Z_{j+1} the random variable

$$Z_{j+1} = B\left(j, \tau^*_{j+1}\right)\varphi\left(X_{\tau^*_{j+1}}\right).$$

Obviously, as B and φ are bounded, Z_{j+1} is also a bounded random variable. Let us recall that the conditional expectation

$$\mathbb{E}\left(Z_{j+1}|X_j\right) \tag{1.15}$$

can be expressed as $\psi_j(X_j)$, where ψ_j minimizes

$$\mathbb{E}\left([X_{j+1} - f(X_j)]^2\right)$$

among all functions f such that $\mathbb{E}\left(f(X_j)^2\right) < +\infty$. This come from the definition of the conditional expectation as a L^2-projection on the set of the $\sigma(X_j)$-measurable random variables and the well known property that all these random variables can be writen as $f(X_j)$.

Assume now that we have a sequence of functions $(g_l, l \geq 1)$ which is a total basis of $L^2_j = L^2\left(\mathbb{R}^d, \text{law of } X_j\right)$ for every time $j, 1 \leq j \leq N$. For all time index j we are thus able to express ψ_j as

$$\psi_j = \sum_{l \geq 1} \alpha_l g_l,$$

where the convergence of the series has to be understood in L^2_j. So we have a way to compute recursively the optimal exercise time

1. initialize $\tau_N = N$. Then inductively
2. define $\alpha^j = (\alpha^j_l, l \geq 0)$ as the sequence which minimizes

$$\mathbb{E}\left(\left[B\left(j, \tau^*_{j+1}\right)\varphi\left(X_{\tau^*_{j+1}}\right) - (\alpha \cdot g)(X_j)\right]^2\right)$$

where $\alpha \cdot g = \sum_{l \geq 1} \alpha_l g_l$.

3. define $\tau_j^* = j\mathbf{1}_{\{\varphi(X_j)\geq(\alpha^j\cdot g)(X_j)\}} + \tau_{j+1}^*\mathbf{1}_{\{\varphi(X_j)<(\alpha^j\cdot g)(X_j)\}}$.

This program is not really implementable since we have to perform a minimization in an infinite dimensional space and we are not able to compute the expectation involved in the minimization problem in step 2.

In order to implement the algorithm, the basic idea is to truncate the series at index k. This leads to the following modified program.

1. initialize $\hat{\tau}_N = N$. Then inductively
2. define $\widehat{\alpha}^{j,k} = (\widehat{\alpha}_l^{j,k}, 0 \leq l \leq k)$ as the vector which minimizes

$$\mathbb{E}\left(\left[B\left(j,\hat{\tau}_{j+1}\right)\varphi\left(X_{\hat{\tau}_{j+1}}\right) - (\hat{\alpha}^{j,k}\cdot g)(X_j)\right]^2\right) \tag{1.16}$$

   where $(\widehat{\alpha}^{j,k}\cdot g) = \sum_{l=1}^k \widehat{\alpha}_l^{j,k} g_l$.

3. define

$$\widehat{\tau}_j = j\mathbf{1}_{\varphi(X_j)\geq(\widehat{\alpha}^{j,k}\cdot g)(X_j)} + \widehat{\tau}_{j+1}\mathbf{1}_{\varphi(X_j)<(\widehat{\alpha}^{j,k}\cdot g)(X_j)}.$$

To make the previous program implementable, we introduce a Monte-Carlo version of (1.16).

A Monte-Carlo approach to the regression problem Let $(X_n^{(m)}, 0 \leq n \leq N)$, for $1 \leq m \leq M$ be M paths sampled along the law of the process $(X_n, 0 \leq n \leq N)$. We replace the minimization problem 1.16 by

1. initialize $\tau_N^M = N$. Then inductively
2. define $\alpha_M^{j,k} = (\alpha_M^{j,k}, 0 \leq j \leq k)$ as the vector which minimizes

$$\frac{1}{M}\sum_{1\leq m\leq M}\left((\alpha\cdot g)\left(X_j^{(m)}\right) - B\left(j,\tau_{j+1}\right)\varphi\left(X_{\tau_{j+1}}^{(m)}\right)\right)^2 \tag{1.17}$$

3. define for each trajectory m

$$\tau_j^{(m)} = j\mathbf{1}_{\left\{\varphi(X_j)\geq(\alpha_M^{j,k}\cdot g)(X_j)\right\}} + \tau_{j+1}^{(m)}\mathbf{1}_{\left\{\varphi(X_j)<(\alpha_M^{j,k}\cdot g)(X_j)\right\}}.$$

This algorithm is now fully implementable and that the estimator of the price is given by

$$\frac{1}{M}B(0,\tau^{(m)})\varphi(X_{\tau^{(m)}}^{(m)}).$$

Remark 1.5.2.

- The minimization problem (1.17) is a standard least square approximation problem. An algorithm for solving it can be found for instance in [Press et al.(1992)].
- The basis $(g_k, k \geq 1)$ is supposed to be independent of the time index j. This is just for sake of simplicity and one can change the basis at each time index if needed for numerical efficiency. For instance, it is quite usual to add the payoff function in the basis, at least when j approaches N.

- Very often for financial products such as call or put options,

$$\mathbb{P}\left(\varphi(X_j)=0\right)>0,$$

for every time j. Since $\{\varphi\left(X_j\right)=0\}\subset A_j$, it is useless to compute

$$\mathbb{E}\left(B\left(j,\tau_{j+1}^*\right)\varphi\left(X_{\tau_{j+1}^*}\right)\Big|X_j\right)$$

on the set $\{\varphi\left(X_j\right)=0\}$ and we can thus restrict ourself in the regression to trajectories such that $\{\varphi\left(X_j\right)>0\}$.

- A rigorous proof of the convergence of the Longstaff Schwartz algorithm is given in [Clément et al.(2002)].

The Tsitsiklis–VanRoy algorithm.

This algorithm use a revised dynamical programming algorithm similar to (1.8) but involving

$$\widetilde{u}(j,x)=\tilde{P}_j u(j+1,x)$$

instead of $u(j,x)$. An approximation of $\widetilde{u}$ is then performed by using a regression method.

Another form of the dynamic programming algorithm. Using $\widetilde{u}$, the algorithm (1.8) can be rewriten as

$$\left\{\begin{array}{l}\widetilde{u}(N,x)=\tilde{P}_{n-1}\varphi(x)\\ \text{For } 0\leq j\leq N-2,\\ \qquad \widetilde{u}(j,x)=\tilde{P}_j\max\left(\varphi(.),\tilde{u}(j+1,.)\right)(x).\end{array}\right. \tag{1.18}$$

So that $Q_0=\max(\varphi(x_0),\widetilde{u}(0,x_0)$.

For every $1\leq j\leq N-1$, let $(g_l,l\geq 1)$ be a basis of L_j^2. At time $N-1$, the authors approximate $\tilde{u}(j,x)$ by its projection on the subspace spanned by $\{g_1,\ldots,g_k\}$, according to the L_j^2 norm. They iterate this procedure backward in time, projecting at time j the function

$$B\left(j,j+1\right)\max\left(\varphi(x),(\alpha^{j+1}\cdot g)(x)\right)$$

on the vector space generated by $g_1,\ldots,g_k$ according to the L_j^2 norm. The resulting approximated dynamical programming algorithm is then

$$\left\{\begin{array}{l}\alpha^{N-1}=D_{N-1}^{-1}\mathbb{E}\left(B\left(N-1,N\right)\varphi\left(X_N\right){}^t g^{N-1}\left(X_{N-1}\right)\right)\\ \text{For } 1\leq j\leq N-2\\ \quad \alpha^j=D_j^{-1}\mathbb{E}\left(B\left(j,j+1\right)\max\left(\varphi\left(X_{j+1}\right),\left(\alpha^{j+1}\cdot g\right)\left(X_{j+1}\right)\right){}^t g\left(X_j\right)\right)\\ \alpha=\mathbb{E}\left(B\left(0,1\right)\max\left(\varphi\left(X_1\right),\left(\alpha^1\cdot g\right)\left(X_1\right)\right)\right)\end{array}\right. \tag{1.19}$$

where D_j^M is the covariance matrix of the vector $g(X_j)$ defined as

$$(D_j)_{k,l}=\mathbb{E}\left(g_k(X_j)g_l(X_j)\right)-\mathbb{E}\left(g_k(X_j)\right)\mathbb{E}\left(g_l(X_j)\right).$$

At time 0, the proposed approximation of the price is given by

$$\max\left(\varphi\left(x_0\right),\alpha\right).$$

The Monte-Carlo algorithm. In order to obtain an implementable algorithm we use a set of trajectories denoted by $(X^{(m)}, 1 \leq m \leq M)$ sampled along the law of X. The Monte-Carlo version of (1.19) is

$$\left\{ \begin{array}{l} \alpha_M^{N-1} = (D_{N-1}^M)^{-1} \frac{1}{M} \sum_{1 \leq l \leq M} B\left(N-1, N\right) \varphi\left(X_N^{(m)}\right) {}^t g\left(X_{N-1}^{(m)}\right) \\ \text{For } 1 \leq j \leq N-2, \\ \quad \alpha_M^j = (D_j^M)^{-1} B\left(j, j+1\right) \times \\ \qquad \times \frac{1}{M} \sum_{1 \leq l \leq M} \max\left(\varphi\left(X_{j+1}^{(m)}\right), (\alpha_M^{j+1} \cdot g)\left(X_{j+1}^{(m)}\right)\right) {}^t g\left(X_j^{(m)}\right) \\ \alpha_M^0 = B\left(0,1\right) \frac{1}{M} \sum_{1 \leq l \leq M} \max\left(\varphi\left(X_1^{(m)}\right), (\alpha_M^1 \cdot g)\left(X_1^{(m)}\right)\right), \end{array} \right. \tag{1.20}$$

where D_j denotes the empirical dispersion matrix of the set of points

$$(g(X_j^{(m)}), 1 \leq m \leq M)$$

of $\mathbb{R}^k$.The price at time 0 is approximated by $\max\left(\varphi\left(X_0\right), \alpha_M^0\right)$.

The quantization algorithms.

Algorithms based on quantization methods have been introduced by Bally and Pagès in [Pages and Bally(2000)]. The principle of these algorithms is to discretize, at time j, the underlying Markov process X_j by a set of "quantization levels" $y^j = \left(y_i^j\right)_{1 \leq i \leq n_j}$ and then to approximate the transition kernel of X at time j by a discrete kernel $\widehat{P}_j = (\widehat{P}_j(y_k^j, y_l^{j+1}), 1 \leq k \leq n_j, l \leq n_{j+1})$.

More precisely, let $P_j\left(x, dy\right)$ be the transition kernel between times j and $j+1$ of a Markov chain $(X_n, 0 \leq n \leq N)$ taking its values in $\mathbb{R}^n$ and starting from x_0.

For $1 \leq j \leq N$, let y^j be a set of vectors of $\mathbb{R}^d$

$$y^j = \left\{y_1^j, \ldots, y_{n_j}^j\right\}.$$

Note that $X_0 = x_0$ is supposed to be constant and need not to be discretized. We recall that

$$\sup_{\tau \in \mathcal{T}_{0,N}} \mathbb{E}\left(B\left(0, \tau\right) \varphi(X_\tau)\right) = u(0, x_0),$$

where u is given by the dynamic programming algorithm (1.8)

$$\left\{ \begin{array}{l} u(N, x) = \varphi(x) \\ u(j, x) = \max\left(\varphi(x), \tilde{P}_j u(j+1, x)\right) \quad 0 \leq j \leq N-1. \end{array} \right. \tag{1.21}$$

We will now approxime the actualized transition kernel $\tilde{P}_j B(j, j+1) P_j$ by a discrete transition kernel $\widehat{P}_j$ defined on $y^j \times y^{j+1}$. This kernel must satisfy the following requirement : $\widehat{P}_j\left(y_k^j, y_l^{j+1}\right)$ approximate the actualized probability that the process X moves from a neighborhood of y_k^j at time j to a neighborhood of y_l^{j+1} at time $j+1$. This can be done using $\widehat{P}_j$ defined by

$$\widehat{P}_j\left(y_k^j, y_l^{j+1}\right) = B(j, j+1) \frac{\mathbb{P}(X_j \in C_k(y^j), X_{j+1} \in C_l(y^{j+1}))}{\mathbb{P}(X_j \in C_k(y^j))} \tag{1.22}$$

where the set $C_i\left(y^j\right)$ is the i^{th} Voronoï cell of the set of points y^j, defined by

$$C_i\left(y^j\right) = \left\{ z \in \mathbb{R}^d, \left\| z - y_i^j \right\| = \min_{1 \le l \le n_j} \left\| z - y_l^j \right\| \right\}. \tag{1.23}$$

If we assume that, for every $x \in \mathbb{R}^d$ and l, $P_j\left(x, \partial C_l\left(y^{j+1}\right)\right) = 0$, equation (1.22) defines a probability transition kernel.

The approximation of the dynamic programming algorithm (1.21) using the matrix $\widehat{P}_j$ is given by

$$\begin{cases} \widehat{u}(N, y_k^N) = f\left(y_k^N\right) \quad 1 \le k \le n_N \\ \text{For } 0 \le j \le N-1,\ 1 \le k \le n_j, \\ \qquad \widehat{u}(j, y_k^j) \max\left(f\left(y_k^j\right), \displaystyle\sum_{1 \le l \le n_{j+1}} \widehat{P}_j\left(y_k^j, y_l^{j+1}\right) \widehat{u}(j+1, y_l^{j+1}) \right) \end{cases} \tag{1.24}$$

and the proposed price approximation is $\widehat{u}(0, x_0)$.

In order to have an algorithm we need to construct a Monte-Carlo procedure to approximate $\widehat{P}_j$. Let $(X^{(m)}, 1 \le m \le M)$ be M sample paths drawned along the law of the markov chain X and independent. The formula (1.22) suggests to consider the MonteCarlo estimator $\bar{P}_j$ defined by

$$\bar{P}_j\left(y_k^j, y_l^{j+1}\right) = B(j, j+1) \frac{\sum_{1 \le m \le M} \mathbf{1}_{\left\{X_j^{(m)} \in C_k(y^j), X_{j+1}^{(m)} C_l(y^{j+1})\right\}}}{\sum_{1 \le m \le M} \mathbf{1}_{\left\{X_j^{(m)} \in C_k(y^j)\right\}}} \tag{1.25}$$

An implementable version of (1.24) obtained by replacing $\widehat{P}_j$ by $\bar{P}_j$ in (1.24).

Remark 1.5.3. When $\widehat{u}$ has been computed at the point $(y_i^j, 1 \le i \le n_j)$, we can extend it to $\mathbb{R}^n$ by setting $\widehat{u}(j, x) = \widehat{u}(j, y_i^j)$ for $x \in C_i(y^j)$. This allows us to approximate the stopping time τ^* by

$$\widehat{\tau} = \min\left\{ j; \varphi(X_j) = \widehat{u}(j, X_j) \right\}, \tag{1.26}$$

and leads to another approximation for the price at time zero by

$$\frac{1}{M} \sum_{1 \le m \le M} B\left(0, \widehat{\tau}^{(m)}\right) f\left(\mathcal{Q}_{\widehat{\tau}^{(m)}}, X_{\widehat{\tau}^{(m)}}^{(m)}\right),$$

where $\widehat{\tau}^{(m)}$ is the stopping time computed using 1.26 on the trajectory m.

Choice of the quantization sets There are many ways to implement the quantization method, depending on the choice of the quantization sets y^j. We present here two of them.

The first one is the Random Quantization algorithm. We assume that $n_j = n$ for all j. We use a self–quantization procedure to generate the quantization sets, that is to say we set

$$y_m^j = X_j^{(m)} \quad 1 \le j \le n \quad 1 \le m \le n.$$

where $X^{(m)}, 1 \le m \le M$ are independently drawn along the law of X. This choice is acceptable but is not the best among all the random choices (see [Cohort(2004)]).

A second way to choose the quantization sets is to minimize a quantity reflecting the quality of the quantization set. The distortion $\mathcal{D}$ defined for a set $y = (y_k, 1_l eqk \leq n)$ and a random variable X by

$$\mathcal{D}(y, X)\, \mathbb{E}\left(\min_{1\leq k\leq n} \|X - y_k\|^2\right)$$

is a classical choice for this. Moreover, an error bound for the previously defined approximation of the price can be derived in terms of the distortions $(\mathcal{D}(y^j, X_j), 1 \leq j \leq N)$ (see [Pages and Bally(2000)]). It can be shown that there exists an optimal set y^* satisfying

$$\mathcal{D}(y^*, X) = \min_{y\in\left(\mathbb{R}^d\right)^n} \mathcal{D}(y, X)$$

This optimal set can be computed numerically by using the Lloyd algorithm : let $y^{(0)} \in \left(\mathbb{R}^d\right)^n$ be such that all points are different and define a sequence $y^{(n)}$ setting

$$y_k^{(n+1)} = \mathbb{E}\left(Y | Y \in C_k\left(y^{(n)}\right)\right).$$

Note that a Monte-Carlo method must be used to approximate the previous expectation. Then $\mathcal{D}\left(y^{(n)}, X\right)$ is decreasing and the sequence $y^{(n)}$ converges to a point y^* which realizes a local minimum of $\mathcal{D}(., X)$. For more details on the Lloyd algorithm and its convergence see [Sabin and Gray(1986)]. Another classical optimization procedure for $\mathcal{D}$ uses the Kohonen algorithm, see [Pages(1995)].

The Broadie–Glassermann algorithm.

For this algorithm we use a random mesh depending on time and write a dynamical programming algorithm at points of this mesh.

Mesh generation. We assume that the transition matrices on $\mathbb{R}^n$ of the process $(X_n, n_g eq0)$ $P_j(x, dy)$ have a density

$$f_j(x, y).$$

This is obviously the case in the Black and Scholes model. We assume moreover that $X_0 = x_0 \in \mathbb{R}^n$.

Let $(Y_1^{(m)}, 1 \leq m \leq n)$ be a sample following the density $f_0(x_0, y)$. The set $(Y_1^{(m)}, 1 \leq m \leq n)$ define the mesh at time 1.

Then the mesh at time $j+1$ $(Y_{j+1}^{(m)}, 1 \leq m \leq n)$ is defined assuming that $Y_j = (Y_j^{(m)}, 1 \leq m \leq n)$ is already drawn and that the conditional law of Y^{j+1} knowing Y^j is the law of n independant random variables with (conditional) density

$$\frac{1}{n}\sum_{1\leq l\leq n} f_j\left(Y_j^{(l)}, y\right) dy. \tag{1.27}$$

The next step consists in writing an approximation of the dynamical programming algorithm

$$\begin{cases} u(N, x) = \varphi(x), \\ u(j, x) \max\left(\varphi(x), \tilde{P}_j u(j+1, x)\right) \quad 0 \leq j \leq N-1. \end{cases} \tag{1.28}$$

using only the points $(Y_j^{(m)}, 1 \leq m \leq n)$. For this Broadie and Glassermann propose to approximate

$$P_j u(j+1, Y_j^{(k)}),$$

by a sum

$$\frac{1}{n} \sum_{1 \leq l \leq n} \hat{\beta}_j^{k,l}(Y_j^{(k)}, Y_{j+1}^{(m)}) u(j+1, Y_{j+1}^{(m)}). \tag{1.29}$$

They choose the weights $\hat{\beta}_{kl}^j$ by considering (1.29) as a Monte-Carlo sum. Indeed, one has

$$\begin{aligned}
&\mathbb{E}\left(u(j+1, Y_{j+1}^{(m)}) \frac{f_j\left(Y_j^{(k)}, Y_{j+1}^{(m)}\right)}{\frac{1}{n}\sum_{1 \leq l \leq n} f_j\left(Y_j^{(l)}, Y_{j+1}^{(m)}\right)} \middle| Y_j^{(k)}, k = 1, \ldots, n \right) = \\
&\qquad = \int_{\mathbb{R}^d} u(j+1, u) \frac{f_j\left(Y_j^{(k)}, u\right)}{\frac{1}{n}\sum_{1 \leq l \leq n} f_j(Y_j^{(l)}, y)} \frac{1}{n} \sum_{1 \leq l \leq n} f_j\left(Y_j^{(l)}, y\right) dy \qquad (1.30) \\
&\qquad = \int_{\mathbb{R}^d} u(j+1, u) f_j\left(Y_j^{(k)}, y\right) \, dy \\
&\qquad = P_j u(j+1, Y_j^{(k)})
\end{aligned}$$

But $(Y_{j+1}^{(m)}, 1 \leq m \leq n)$ are, conditionally to $(Y_j^{(m)}, 1 \leq m \leq n)$, independant and identically distributed. So a natural estimator for (1.30) is given by

$$\frac{1}{n} \sum_{1 \leq m \leq n} \hat{\beta}_{k,m}^j u(j+1, Y_{j+1}^{(m)}) \tag{1.31}$$

where

$$\hat{\beta}_{k,m}^j = \frac{f_j\left(Y_j^{(k)}, Y_{j+1}^{(m)}\right)}{\frac{1}{n}\sum_{1 \leq l \leq n} f_j\left(Y_j^{(k)}, Y_{j+1}^{(l)}\right)}. \tag{1.32}$$

The approximation of the dynamical programming algorithm is then given by

$$\begin{cases}
\widehat{u}(N, Y_{(k)}^N) = \varphi(Y_{(k)}^N), \\
\text{for } 0 \leq k \leq n,\ 1 \leq j \leq N, \\
\quad \widehat{u}(j, Y_j^{(k)}) \max\left(\varphi\left(Y_j^{(k)}\right), \sum_{1 \leq l \leq n} \hat{\beta}_{k,m}^j, \widehat{u}(j+1, Y_{j+1}^{(m)})\right),
\end{cases}$$

and the price approximation by $\widehat{u}(0, x_0)$.

The Barraquand–Martineau algorithm

To avoid dimensionality issue, the authors propose to approximate the optimal stopping strategy a sub–optimal one : assume that the option holder knows at time j the payoff values $\{\varphi(X_i), i \leq n\}$ but not the stock values $\{X_i, i \leq n\}$. Then the option holder can only exercise according to a stopping time with respect to $(\mathcal{G}_n, n \geq 0)$ where $\mathcal{G}_n = \sigma(X_i, i \leq n)$. Doing this, the price is approximated (by below) by

$$\sup_{\tau \in \mathcal{G}_{0,N}} \mathbb{E}\left(B(0, \tau) \varphi(X_\tau)\right) \tag{1.33}$$

where $\mathcal{G}_{0,N}$ is the set of the $\mathcal{G}_n$–stopping times taking values in $\{0,\ldots,N\}$.

To compute (1.33), the authors propose a quantization method for the one-dimensional dynamic programming algorithm. Obviously, in general, the process $(\varphi(X_n), n \geq 0)$ is not Markovian with respect to $\mathcal{G}$ and a second approximation has to be done in order to write the following dynamical programming algorithm assuming that $(\varphi(X_n), n \geq 0)$ is Markovian

$$\begin{cases} u(N,x) = \varphi(x) \\ u(j,x) = \max\left(\varphi(x), \hat{P}_j u(j+1,x)\right), \quad 1 \leq j \leq N, \end{cases} \tag{1.34}$$

where $\hat{P}_j(x,dy)$ is given by

$$\mathbb{E}\left(B(n,n+1) f\left(\varphi(X_{n+1})\right) | \varphi(X_n)\right) = \hat{P}_n f\left(\varphi(X_n)\right).$$

Assume that the two previous approximations are acceptable. The algorithm (1.34) can now be handled by a one–dimensional quantization technique. For $1 \leq j \leq N$, let

$$\left\{z_2^j < \ldots < z_{n_j}^j\right\} \subset \mathbb{R}$$

and let $z_1^j = -\infty$, $z_{n_j+1}^j = +\infty$. Then, define the sets $y^j = \left\{y_1^j, \ldots, y_{n_j}^j\right\}$ by

$$y_k^j = \mathbb{E}\left(\varphi\left(X_j\right) | \varphi\left(X_j\right) \in \left[z_k^j, z_{k+1}^j\right]\right). \tag{1.35}$$

As in the quantization algorithm, the kernel $\hat{P}_j$ is discretized on $y^j \times y^{j+1}$ by $\hat{\hat{P}}_j$ where

$$\hat{\hat{P}}_j\left(y_k^j, y_l^{j+1}\right) = B(j,j+1) \frac{\mathbb{P}\left(\varphi(X_j) \in [z_k^j, z_{k+1}^j], \varphi(X_{j+1}) \in [z_l^{j+1}, z_{l+1}^{j+1}]\right)}{\mathbb{P}\left(\varphi(X_j) \in [z_k^j, z_{k+1}^j]\right)} \tag{1.36}$$

The next step is to approximate (1.34) using $\widehat{P}_j\left(y_k^j, y_l^{j+1}\right)$,

$$\begin{cases} \hat{\hat{u}}(N, y_k^N) &= y_k^N, \quad 1 \leq k \leq n_N \\ \hat{\hat{u}}(j, y_k^j) &= \max\left(y_k^j, \sum_{1 \leq l \leq n_{j+1}} \hat{\hat{P}}_j\left(y_k^j, y_l^{j+1}\right) \hat{\hat{u}}(j+1, y_l^{j+1})\right), \\ & \qquad \text{For } 1 \leq k \leq n_j,\ 0 \leq j \leq N. \end{cases} \tag{1.37}$$

The approximation of the price at time 0 is given by $\hat{\hat{u}}(0,x_0)$.

A Monte-Carlo version of (1.37) is obtain using m samples of the process X, $(X_j^{(m)}, 1 \leq j \leq N)$ and approximating

$$y_k^j \text{ with } \frac{1}{\text{Card}\left(A_k^j\right)} \sum_{m \in A_k^j} \varphi(X_j^{(m)}),$$

and

$$\hat{\hat{P}}_j \text{ with } \frac{\text{Card}\left(A_k^j \cap A_l^{j+1}\right)}{\text{Card}\left(A_k^j\right)},$$

where

$$A_k^j = \left\{1 \leq m \leq n, \varphi(X_j^{(m)}) \in \left[z_k^j, z_{k+1}^j\right]\right\}.$$

Remark 1.5.4. Comparison of most of the algorithms presented here can be found in [Fu et al.(2001)].

1.6 Exercises and problems

Exercise 1. Let Z be a Gaussian random variable and K a positive real number.

1. Let $d = \frac{\mathbb{E}(Z)-\log(K)}{\sqrt{\mathrm{Var}\ (Z)}}$, prove that

$$\mathbb{E}\left(\mathbf{1}_{\{Z\geq\log(K)\}}(e^Z\right) = e^{\mathbb{E}(Z)+\frac{1}{2}\mathrm{Var}\ (Z)}N\left(d+\sqrt{\mathrm{Var}\ (Z)}\right).$$

2. Prove the formula (Black and Scholes formula)

$$\mathbb{E}\left(\left(e^Z-K\right)_+\right) = e^{\mathbb{E}(Z)+\frac{1}{2}\mathrm{Var}\ (Z)}N\left(d+\sqrt{\mathrm{Var}\ (Z)}\right) - KN(d),$$

Exercise 2. Let X and Y be two independent Gaussian random variables with mean 0 and variance 1.

1. Prove that, if $R = \sqrt{X^2+Y^2}$ and θ is the polar angle, then $\theta/2\pi$ and $\exp(-R^2/2)$ are two independent random variables following a uniform law on $[0,1]$.

2. Using the previous result, deduce proposition 1.2.5.

Exercise 3. Consider the case of a European call in the Black and Scholes model with a stochastic interest rate. Suppose that the price of the stock is 1, and the option price at time 0 is given $\mathbb{E}(Z)$ with Z defined by

$$Ze^{-\int_0^T r_\theta d\theta}\left[e^{\int_0^T r_\theta d\theta-\frac{\sigma^2}{2}T+\sigma W_T}-K\right]_+.$$

1. Prove that the variance of Z is bounded by $\mathbb{E}e^{-\sigma^2T+2\sigma W_T}$.

2. Prove that $\mathbb{E}e^{-\frac{1}{2}\gamma^2T+\gamma W_T} = 1$, and deduce an estimate for the variance of Z

Exercise 4. Let λ and K be two real positive numbers such that $\lambda < K$ and X_m be the random variable

$$X_m = \left(\lambda e^{\sigma(G+m)}-K\right)_+ e^{-mG-\frac{m^2}{2}}.$$

We denote its variance by σ_m^2. Give an expression for the derivative of σ_m^2 with respect to m as an expectation, then deduce that σ_m^2 is a decreasing function of m when $m \leq m_0 = \log(K/\lambda)/\sigma$.

Exercise 5. Let $G = (G_1,\ldots,G_d)$ be a d-dimensional Gaussian vector with covariance equal to the identity matrix. For each $m \in \mathbb{R}^d$, let $V(m)$ denote the variance of the random variable $X_m = \phi(G+m)e^{-m.G-\frac{|m|^2}{2}}$.

1. Prove that

$$\frac{\partial V}{\partial m_i}(m) = \mathbb{E}\left(\phi^2(G)e^{-m.G+\frac{|m|^2}{2}}(m_i-G_i)\right).$$

2. Assume that $\phi(G) = \left(\sum_{i=1}^d \lambda_i e^{\sigma_i G_i} - K\right)_+$, λ_i, σ_i, K being real positive constant. Let $m_i(\alpha) = \alpha\lambda_i\sigma_i$. Prove that, if $\sum_{i=1}^d \lambda_i < K$ then

$$\frac{dV(m(\alpha))}{d\alpha} \leq 0,$$

for $0 \leq \alpha \leq \frac{K-\sum_{i=1}^d \lambda_i}{\sum_{i=1}^d (\lambda_i\sigma_i)^2}$ and deduce that if

$$m_0^i = \sigma_i\lambda_i \frac{K - \sum_{i=1}^d \lambda_i}{\sum_{i=1}^d (\lambda_i\sigma_i)^2},$$

then $V(m_0) \leq V(0)$.

Problem 6. The aim of this problem is to prove that the antithetic variable method decreases the variance for a function which is monotonous with respect to each of its arguments.

1. Let f and g be two increasing functions from $\mathbb{R}$ to $\mathbb{R}$. Prove that, if X and Y are two real random variables then we have

$$\mathbb{E}\left(f(X)g(X)\right) + \mathbb{E}\left(f(Y)g(Y)\right) \geq \mathbb{E}\left(f(X)g(Y)\right) + \mathbb{E}\left(f(Y)g(X)\right).$$

2. Deduce that, if X is a real random variable, then

$$\mathbb{E}\left(f(X)g(X)\right) \geq \mathbb{E}\left(f(X)\right)\mathbb{E}\left(g(X)\right).$$

3. Prove that if $X_1, \ldots, X_n$ are n independent random variables then

$$\mathbb{E}\left(f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)|X_n\right) = \phi(X_n),$$

where ϕ is a function which can be computed as an expectation.

4. Deduce from this property that if f and g are two increasing (with respect each of its argument) functions then

$$\mathbb{E}\left(f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)\right) \geq \mathbb{E}\left(f(X_1, \ldots, X_n)\right)\mathbb{E}\left(g(X_1, \ldots, X_n)\right).$$

5. Let h be a function from $[0,1]^n$ in $\mathbb{R}$ which is monotonous with respect to each of its arguments. Let $U_1, \ldots, U_n$ be n independent random variables following the uniform law on $[0,1]$. Prove that

$$\mathrm{Cov}\left(h(U_1, \ldots, U_n)h(1-U_1, \ldots, 1-U_n)\right) \leq 0,$$

and deduce that in this case the antithetic variable method decreases the variance.

Problem 7. Let X and Y be independent real random variables. Let F and G be the distribution functions of X and G respectively We want to compute by a Monte-Carlo method the probability

$$\theta = \mathbb{P}\left(X + Y \leq t\right).$$

1. Propose a variance reduction procedure using a conditioning method.

2. We assume that F and G are (at least numerically) easily invertible. Explain how to implement the antithetic variates methods. Why does this method decrease the variance in this case ?

3. Assume that h is a function such that $\int_0^1 |h(s)|^2 ds < +\infty$. Let $(U_i, i \geq 1)$ be a sequence of independent random variates with a uniform distribution on $[0,1]$. Prove that $\frac{1}{N}\sum_{i=1}^N h\left((i-1+U_i)/n\right)$ has a lower variance than $\frac{1}{N}\sum_{i=1}^N h\left(U_i\right)$.

Problem 8. Let Z be a random variable given by

$$Z = \lambda_1 e^{\beta_1 X_1} + \lambda_2 e^{\beta_2 X_2},$$

where (X_1, X_2) is a couple of real random variables and λ_1, λ_2, β_1 and β_2 are real positive numbers. This problem studies various methods to compute the price of an index option given by $p = \mathbb{P}\left(Z > t\right)$.

1. In this question, we assume that (X_1, X_2) is a Gaussian vector with mean 0 such that $\mathrm{Var}\ (X_1) = \mathrm{Var}\ (X_2) = 1$ and $\mathrm{Cov}\ (X_1, X_2) = \rho$, with $|\rho| \leq 1$. Explain how to simulate random samples along the law of Z. Describe a Monte-Carlo method allowing to estimate p and explain how to estimate the error of the method.

2. Explain how to use low discrepancy sequences to compute p.

3. We assume that X_1 and X_2 are two independent Gaussian random variables with mean 0 and variance 1. Let m be a real number. Prove that p can be written as

$$p = \mathbb{E}\left[\phi(X_1, X_2)\mathbf{1}_{\lambda_1 e^{\beta_1(X_1+m)} + \lambda_2 e^{\beta_2(X_2+m)} \geq t}\right],$$

for some function ϕ. How can we choose m such that

$$\mathbb{P}(\lambda_1 e^{\beta_1(X_1+m)} + \lambda_2 e^{\beta_2(X_2+m)} \geq t) \geq \frac{1}{4}?$$

Propose a new Monte-Carlo method which allows to compute p. Explain how to check on the drawings that the method does reduce the variance.

4. Assuming now that X_1 and X_2 are two independent random variables with distribution functions $F_1(x)$ and $F_2(x)$ respectively. Prove that

$$p = \mathbb{E}\left[1 - G_2\left(t - \lambda_1 e^{\beta_1 X_1}\right)\right],$$

where $G_2(x)$ is a function such that the variance of

$$1 - G_2\left(t - \lambda_1 e^{\lambda_1 X_1}\right),$$

is always less than the variance of $\mathbf{1}_{\lambda_1 e^{\beta_1 X_1} + \lambda_2 e^{\lambda_2 X_2} > t}$. Propose a new Monte-Carlo method to compute p.

5. We assume again that (X_1, X_2) is a Gaussian vector with mean 0 and such that $\text{Var}\ (X_1) = \text{Var}\ (X_2) = 1$ and $\text{Cov}\ (X_1, X_2) = \rho$, with $|\rho| \leq 1$. Prove that $p = \mathbb{E}\left[1 - F_2\left(\phi(X_1)\right)\right]$ where F_2 is the repartition function of X_2 and ϕ a function to be computed.

   Deduce a variance reduction method computing p.

Problem 9. Calcul par une méthode de Monte-Carlo du prix d'un CDS.
On considère une suite de temps déterministes $0 < T_1 < T_2 < \cdots < T_n$ et un taux d'intérêt $r \geq 0$. On note par τ un temps (de défaut) aléatoire et l'on cherche à évaluer par une méthode de Monte-Carlo le prix d'un CDS donné par :

$$P = \mathbb{E}\left(H(\tau)\right)$$

où :

$$H(\tau) = R\sum_{i=1}^{n-1} e^{-rT_i}\mathbf{1}_{\{\tau > T_i\}} - e^{-r\tau}\mathbf{1}_{\{\tau \leq T_n\}}$$

Intensité déterministe On suppose que $(\lambda_s, s \geq 0)$ est une fonction continue non aléatoire de s, telle que, pour tout $s \geq 0$, $\lambda_s > 0$. On suppose que τ suit la loi

$$\mathbf{1}_{\{t \geq 0\}} e^{-\int_0^t \lambda_s ds}\lambda_t dt.$$

1. Reconnaître la loi de τ lorsque λ_s ne dépend pas de s. Comment peut on, alors, simuler un variable aléatoire suivant la loi de τ ? Montrer que l'on peut calculer P à l'aide d'une formule simple.

2. Soit ξ une variable aléatoire de loi exponentielle de paramètre 1. On note Λ la fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$ définie par :
   $$\Lambda(t) = \int_0^t \lambda_s ds,$$
   et Λ^{-1} son inverse. Montrer que $\Lambda^{-1}(\xi)$ suit la même loi que τ et, en déduire une méthode de simulation selon la loi de τ.

3. En supposant Λ et Λ^{-1} calculable explicitement, proposer une méthode de Monte-Carlo permettant de calculer P. Expliquer comment l'on peut estimer, alors, l'erreur commise.

4. On suppose que $\Lambda(T_n) \leq K$. Montrer que :
   $$P = \left(R\sum_{i=1}^{n-1} e^{-rT_i}\right)\mathbb{P}\left(\xi > K\right) + \mathbb{E}\left(H(\tau)|\xi \leq K\right)\mathbb{P}\left(\xi \leq K\right).$$

   Expliquer comment simuler efficacement une variable aléatoire selon la loi de ξ conditionnellement à l'événement $\{\xi \leq K\}$.

   En déduire une méthode de réduction de variance pour le calcul de P.

Intensité aléatoire On suppose dans ce paragraphe on suppose que λ_t est un processus aléatoire donné par $\lambda_t = \exp(X_t)$ où $(X_t, t \geq 0)$ est solution de :

$$dX_t = -c\,(X_t - \alpha)\,dt + \sigma dW_t, X_0 = x.$$

avec c, σ, α des réels donnés et $(W_t, t \geq 0)$ un mouvement brownien.

On pose $\tau = \Lambda^{-1}(\xi)$, ξ étant un variable aléatoire de loi exponentielle de paramètre 1 indépendante de W.

1. Calculer $\mathbb{E}(\lambda_t)$. Proposer une variable de contrôle pour le calcul de P. Comment peut on vérifier numériquement que cette variable de contrôle réduit effectivement la variance ?

2. Calculer :
$$\mathbb{E}\left(R\sum_{i=1}^{n-1} e^{-rT_i}\mathbf{1}_{\{\tau>T_i\}} - e^{-r\tau}\mathbf{1}_{\{\tau<T_n\}}\Big|\lambda_t, t \geq 0\right).$$
En déduire un méthode de simulation évitant de simuler la variable aléatoire ξ dont on montrera qu'elle réduit la variance.

Problem 10. Une méthode de fonction d'importance

1. Soit $\beta > 0$, on pose $\xi_\beta = \xi/\beta$. Montrer que, pour toute fonction f mesurable positive et pour tout $\beta > 0$ on a :
$$\mathbb{E}\left(f(\xi)\right) = \mathbb{E}\left(f(\xi_\beta)g_\beta(\xi_\beta)\right).$$
g_β étant une fonction que l'on explicitera en fonction de β.

2. En déduire que pour tout $\beta > 0$:
$$P = \mathbb{E}\left(g_\beta(\xi/\beta)H(\Lambda^{-1}(\xi/\beta))\right)$$
avec $H(t) = R\sum_{i=1}^{n} e^{-rT_i}\mathbf{1}_{\{t>T_i\}} - e^{-rt}\mathbf{1}_{\{t<T_n\}}$.

3. On pose $X_\beta = g_\beta(\xi/\beta)H(\Lambda^{-1}(\xi/\beta))$, montrer que pour tout $\beta > 0$:
$$\mathrm{Var}\,(X_\beta) = \mathbb{E}\left(g_\beta(\xi)H^2(\Lambda^{-1}(\xi))\right) - P^2.$$

4. Montrer que $\mathrm{Var}\,(X_\beta) < +\infty$ si $\beta < 2$ et que :
$$\mathrm{Var}\,(X_2) = \frac{1}{2}\mathbb{E}\left(\int_0^{+\infty} H^2(\Lambda^{-1}(t))dt\right) - P^2.$$

5. Montrer que $\beta \to \mathrm{Var}\,(X_\beta)$ est une fonction strictement convexe sur l'intervalle $]0, 2[$ et que
$$\lim_{\beta\to 0+} \mathrm{Var}\,(X_\beta) = +\infty.$$
En admettant que $\lim_{t\to+\infty}\Lambda^{-1}(t) = +\infty$ presque sûrement, montrer que $\lim_{\beta\to 2^-}\mathrm{Var}\,(X_\beta) = +\infty$.

6. Quel β est il souhaitable de choisir dans une méthode de Monte-Carlo ? Proposer un méthode permettant d'approcher cette valeur.

Exercise 11. On considère X une variable aléatoire gaussienne centrée réduite.

1. Pour f une fonction bornée, on note $I = \mathbb{E}(f(X))$ et I_n^1 et I_n^2 les estimateurs :

$$I_n^1 = \frac{1}{2n}\left(f(X_1) + f(X_2) + \ldots + f(X_{2n-1}) + f(X_{2n})\right).$$

$$I_n^2 = \frac{1}{2n}\left(f(X_1) + f(-X_1) + \ldots + f(X_n) + f(-X_n)\right).$$

   où $(X_n, n \geq 1)$ est une suite de variables aléatoires indépendantes tirées selon la loi de X.

   Identifier la limite en loi des variables aléatoires $\sqrt{n}(I_n^1 - I)$. Même question pour la famille de variables aléatoires $\sqrt{n}(I_n^2 - I)$.

   On calculera la variance des lois limites.

2. Comment peut on estimer la variance de lois limites précédentes à l'aide de l'échantillon $(X_n, 1 \leq i \leq 2n)$ pour I_n^1 et $(X_n, 1 \leq i \leq n)$ pour I_n^2?

   Comment évaluer l'erreur dans une méthode de Monte-Carlo utilisant I_n^1 ou I_n^2 ?

3. Montrer que si f est une fonction croissante $\mathrm{Cov}\ (f(X), f(-X)) \leq 0$. Quel est dans ce cas, l'estimateur qui vous parait préférable I_n^1 ou I_n^2 ? Même question si f est décroissante.

Exercise 12. Soit G une variable aléatoire gaussienne centrée réduite.

1. On pose $L^m = \exp\left(-mG - \frac{m^2}{2}\right)$, montrer que $\mathbb{E}(L^m f(G+m)) = \mathbb{E}(f(G))$ pour toute fonction f bornée.

   Soit X^m une autre variable aléatoire, intégrable telle que $\mathbb{E}(X^m f(G+m)) = \mathbb{E}(f(G))$ pour toute fonction f bornée. Montrer que $\mathbb{E}\left(X^m | G\right) = L^m$.

   Dans une méthode de simulation quelle représentation de $\mathbb{E}(f(G))$ vaut il mieux utiliser $\mathbb{E}(X^m f(G+m))$ ou $\mathbb{E}(L^m f(G+m))$?

2. Montrer que la variance de $L^m f(G+m)$ se met sous la forme

$$\mathbb{E}\left(e^{-mG+\frac{m^2}{2}} f^2(G)\right) - \mathbb{E}(f(G))^2,$$

   et que la valeur de m qui minimise cette variance est donnée par $m = \frac{\mathbb{E}(Gf^2(G))}{\mathbb{E}(f^2(G))}$. Que vaut ce m optimum lorsque $f(x) = x$? Commentaire.

3. Soit p_1 et p_2 deux nombres positifs de somme 1 et m_1 et m_2 2 réels, on pose :

$$l(g) = p_1 e^{m_1 g - \frac{m_1^2}{2}} + p_2 e^{m_2 g - \frac{m_2^2}{2}}.$$

   On pose pour f mesurable bornée $\mu(f) = \mathbb{E}(l(G)f(G))$. Montrer que

$$\mu(f) = \int_{\mathbb{R}} f(x)p(x)dx,$$

   p étant une densité que l'on précisera.

4. Proposer une technique de simulation selon la loi de densité p.

5. On suppose que $\tilde{G}$ est une variable aléatoire suivant la loi précédente. Montrer que :

$$\mathbb{E}(l^{-1}(\tilde{G})f(\tilde{G})) = \mathbb{E}(f(G)),$$

$$\text{Var }(l^{-1}(\tilde{G})f(\tilde{G})) = \mathbb{E}\left(l^{-1}(G)f^2(G)\right) - \mathbb{E}(f(G))^2.$$

6. On s'intéresse au cas $p_1 = p_2 = 1/2$, $m_1 = -m_2 = m$ et $f(x) = x$. Montrer que :

$$\text{Var }(l^{-1}(\tilde{G})\tilde{G}) = \mathbb{E}\left(\frac{e^{m^2/2}G^2}{\cosh(mG)}\right).$$

   On note $v(m)$ cette variance comme fonction de m. Vérifier que $v'(0) = 0$ et $v''(0) < 0$.

   Comment choisir m pour réduire la variance lors d'un calcul de $\mathbb{E}(G)$?

Exercise 13. Soit X une variable aléatoire réelle et Y une variable de contrôle réelle. On supposera que $\mathbb{E}(X^2) < +\infty$ et que $\mathbb{E}(Y^2) < +\infty$, $\mathbb{E}(Y) = 0$ pour $i = 1, \ldots, n$.

1. Soit λ un vecteur de $\mathbb{R}^n$, calculer $\text{Var }(X - \lambda Y)$ et la valeur λ^* qui minimise cette variance. As t'on intérêt à supposer X et Y indépendantes ?

2. On suppose que $((X_n, Y_n), n \geq 0)$ est une suite de variables aléatoires indépendantes tirées selon la loi du couple (X, Y). On définit λ_n^* par

$$\lambda_n^* = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n}\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum_{i=1}^n X_i)^n}.$$

   Montrer que λ_n^* tends presque surement vers λ^* lorsque n tend vers $+\infty$.

3. Montrez que $\sqrt{n}\left(\lambda_n^* - \bar{\lambda}_n\right)\bar{Y}_n$, où $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$ tend vers 0 presque sûrement.

4. En utilisant le théorème de Slutsky (voir question suivante) montrer que :

$$\sqrt{n}\left(\frac{1}{n}(X_1 - \lambda_n^* Y_1 + \cdots + X_n - \lambda_n^* Y_n) - \mathbb{E}(X)\right)$$

   tends vers un loi gaussienne de variance $\text{Var }(X - \lambda_n^* Y)$.

   Comment interpreter le résultat pour une méthode de Monte-Carlo utilisant λX comme variable de contrôle ?

5. Montrez le théorème de Slutsky, c'est à dire que si X_n converge en loi vers X et Y_n converge en loi vers une constante a alors le couple (X_n, Y_n) converge en loi vers (X, a) (on pourra considérer la fonction caractéristique du couple (X_n, Y_n).

Problem 14. Méthode de biaisage d'un tirage uniforme par une loi β
Le cas de la dimension 1. On note U une variable aléatoire de loi uniforme sur l'intervalle $[0, 1]$. On considère la famille de loi $\beta(a, 1)$ dont la densité est donnée, pour $a > 0$, par :

$$au^{a-1}\mathbf{1}_{\{u\in[0,1]\}}.$$

On note V_a une variable aléatoire de de loi $\beta(a, 1)$.

1. Proposer une méthode de simulation selon la loi $\beta(a,1)$. Pour g une fonction bornée, comment peut-on estimer $\mathbb{E}\left(g(V_a)\right)$ à l'aide d'une méthode de Monte-Carlo ? Comment obtenir un ordre de grandeur de l'erreur dans cette méthode ?

2. Vérifier que, si f est une fonction de $\mathbb{R}$ dans $\mathbb{R}$ telle que $\mathbb{E}(|f(U)|) < +\infty$, pour tout $a > 0$:
$$\mathbb{E}\left(f(U)\right) = \mathbb{E}\left(\frac{f(V_a)}{aV_a^{a-1}}\right).$$
Comment utiliser cette relation pour calculer $\mathbb{E}\left(f(U)\right)$ à l'aide d'une méthode de Monte-Carlo ? Quelle fonction de a doit on alors minimiser pour obtenir une méthode optimale ?

3. On suppose que f est bornée. Montrer que, pour tout $0 < a < 2$:
$$\sigma_a^2 = \mathrm{Var}\left(\frac{f(V_a)}{aV_a^{a-1}}\right) = \mathbb{E}\left(\frac{f^2(U)}{aU^{a-1}}\right) - \mathbb{E}\left(f(U)\right)^2.$$

4. En utilisant le lemme de Fatou, montrer que, si $\mathbb{P}(f(U) \neq 0) > 0$, $\lim_{a\to 0^+} \sigma_a^2 = +\infty$. Puis que, si f est continue en 0 avec $f(0) \neq 0$, $\lim_{a\to 2^-} \sigma_a^2 = +\infty$.
On supposera, dans la suite, que f est bornée, continue en 0 avec $f(0) \neq 0$ (ce qui implique que $\mathbb{P}(f(U) \neq 0) > 0$).

5. Montrez que, pour $0 < a < 2$, σ_a^2 admet des dérivées d'ordre 1 et 2 par rapport à a qui s'écrivent sous la forme :
$$\frac{d\sigma_a^2}{da} = \mathbb{E}\left(f^2(U)g_1(a,U)\right) \text{ et } \frac{d^2\sigma_a^2}{da^2} = \mathbb{E}\left(f^2(U)g_2(a,U)\right),$$
g_1 et g_2 étant des fonctions de $\mathbb{R}^{+*} \times [0,1]$ dans $\mathbb{R}$ que l'on calculera.

6. En déduire que σ_a^2 est une fonction convexe sur l'intervalle $]0,2[$ qui atteint son minimum au point $\hat{a}$ solution unique de l'équation $\psi(a) = 0$ où
$$\psi(a) = \mathbb{E}\left(\frac{f^2(U)}{a^2U^{a-1}}\left(1 - a|\ln(U)|\right)\right).$$

7. Soit $(U_n, n \geq 1)$ une suite de variables aléatoires indépendantes suivant une loi uniforme sur $[0,1]$. On définit $\psi_n(a)$, pour $a > 0$, par
$$\psi_n(a) = \frac{1}{n}\sum_{i=1}^{n} \frac{f^2(U_i)}{a^2U_i^{a-1}}\left(1 - a|\ln(U_i)|\right)$$
Vérifier que, pour $a \in]0,2[$, $\psi_n(a)$ converge presque sûrement vers $\psi(a)$ lorsque n tends vers $+\infty$.

8. On pose $\tau = \inf\{k \geq 0, |f(U_k)| > 0\}$. Vérifier que $\mathbb{P}(\tau < +\infty) = 1$. Montrer que, pour $n \geq \tau$, $\psi_n(a)$ est une fonction continue et croissante de a vérifiant $\psi_n(0^-) = -\infty$ et $\psi_n(+\infty) = +\infty$. En déduire qu'il existe une unique $\hat{a}_n > 0$ tel que $\psi_n(a_n) = 0$. Proposer un algorithme permettant de calculer a_n et puis (sans preuve) une méthode d'approximation de $\hat{a}$.

Le cas d'une dimension plus grande que 1**.** Pour $d \geq 1$, on note $U = (U^1, \ldots, U^d)$ une variable aléatoire de loi uniforme sur $[0,1]^d$. Pour $a > 0$, on considère une famille de vecteurs aléatoires $V_a = (V_a^1, \ldots, V_a^d)$ composés de variables aléatoires indépendantes de loi $\beta(a, 1)$.

1. Montrer que, si f est une fonction de $[0,1]^d$ dans $\mathbb{R}$ telle que $\mathbb{E}(|f(U)|) < +\infty$, pour tout $a > 0$, on peut écrire :
$$\mathbb{E}\left(f(U)\right) = \mathbb{E}\left(l(a, V_a) f(V_a)\right),$$
l étant une fonction de $\mathbb{R}^{+*} \times [0,1]^d$ dans $\mathbb{R}$ que l'on précisera.

2. Pour f bornée et $0 < a < 2$, exprimer $\sigma_a^2 = \text{Var}\ \left(l(a, V_a) f(V_a)\right)$ à l'aide de la fonction l et du vecteur aléatoire U.

3. Expliquer comment l'on a intérêt à choisir a pour mettre en oeuvre une méthode de Monte-Carlo.

   On suppose que f est bornée, continue en 0 et telle que $f(0) \neq 0$. Vérifier que σ_a^2 est une fonction convexe, telle que $\sigma_{0^+}^2 = +\infty$ et $\sigma_{2^-}^2 = +\infty$ Montrer que la $\hat{a}$ optimal est la solution unique de $\frac{d\sigma_a^2}{da} = 0$. Montrer que cette condition peut s'écrire sous la forme
$$\frac{d\sigma_a^2}{da} = \mathbb{E}\left(H(a, U)\right) = 0.$$
   Proposer (sans preuve) une méthode d'approximation de $\hat{a}$.

Problem 15. Biaisage gaussien "prévisible"
Soit $G = (G_1, G_2, \ldots, G_n)$ un vecteur de gaussiennes, centrées, réduites, indépendantes.

On considère n fonctions $\lambda_i(g)$, $i = 1, \ldots, n$, telles que : $\lambda_i(g) = f_i(g_1, \ldots, g_{i-1})$, où les f_i sont des fonctions de $\mathbb{R}^{i-1}$ dans $\mathbb{R}$, supposées régulières et bornées (f_1 et donc λ_1 sont supposées constantes).

On note R_λ la transformation de $\mathbb{R}^n$ dans $\mathbb{R}^n$ qui associe à g, $g' = R_\lambda(g)$ dont les coordonnées sont calculées par, i croissant de 1 à n,
$$g'_i = g_i + f_i(g'_1, \ldots, g'_{i-1}) = g_i + \lambda_i(g').$$

2. Montrer que R_λ est une transformation bijective de $\mathbb{R}^n$ dans $\mathbb{R}^n$. On vérifiera que R_λ^{-1} son inverse est donné par $R_\lambda^{-1}(g')_i = g'_i - \lambda_i(g')$.

   Montrer que le déterminant de la matrice jacobienne de R_λ^{-1} est identiquement égal à 1.

On définit la variable aléatoire $M_n^\lambda(G)$ par
$$M_n^\lambda(G) = \exp\left(\sum_{i=1}^n \lambda_i(G) G_i - \frac{1}{2}\sum_{i=1}^n \lambda_i^2(G)\right)$$

3. Soit $\mathcal{F}_i = \sigma(G_1, \ldots, G_i)$ montrer que, pour $i \leq n$,
$$\mathbb{E}\left(\exp\left(\lambda_i(G) G_i - \frac{1}{2}\lambda_i^2(G)\right) | \mathcal{F}_{i-1}\right) = 1.$$
   En déduire que $\mathbb{E}(M_n^\lambda(G)) = 1$.

4. Vérifier que l'on définit une nouvelle probabilité $\mathbb{P}^{(\lambda)}$ en posant

$$\mathbb{P}^{(\lambda)}(A) = \mathbb{E}\left(M_n^\lambda(G)\mathbf{1}A\right).$$

et que l'on a ($\mathbb{E}^{(\lambda)}$ désigne l'espérance sous la probabilité $\mathbb{P}^{(\lambda)}$)

$$\mathbb{E}^{(\lambda)}(X) = \mathbb{E}\left(M_n^\lambda(G)X\right)$$

pour toute variable aléatoire $\mathbb{P}^{(\lambda)}$-intégrable X.

5. En utilisant le changement de variable $g = R_\lambda^{-1}(g')$, montrer que la loi de G sous $\mathbb{P}^{(\lambda)}$ est identique à celle de $R_\lambda(G)$ sous $\mathbb{P}$. En déduire une méthode de simulation de la loi de G sous $\mathbb{P}^{(\lambda)}$.

6. On note $X^\lambda = \phi(G)/M_n^\lambda(G)$. Vérifier que $\mathbb{E}^{(\lambda)}(X^\lambda) = \mathbb{E}(\phi(G))$ et montrer que la variance de X^λ sous $\mathbb{P}^{(\lambda)}$ est donnée par

$$\text{Var}^{(\lambda)}(X^\lambda) = \mathbb{E}\left(e^{-\sum_{i=1}^n \lambda_i(G)G_i + \frac{1}{2}\sum_{i=1}^n \lambda_i^2(G)}\phi(G)^2\right) - \mathbb{E}(\phi(G))^2$$

7. Comment peut on simuler une variable aléatoire dont la loi est celle de $X^\lambda = \phi(G)/M_n^\lambda(G)$ sous $\mathbb{P}^{(\lambda)}$? Quel est l'intérêt de pouvoir représenter $\mathbb{E}(\phi(G))$ sous la forme $\mathbb{E}^{(\lambda)}(X^\lambda)$?

8. Pour α un réel donné, on pose $u(\alpha) = \text{Var}^{(\alpha\lambda)}(X^{\alpha\lambda})$. Montrer que u est une fonction différentiable et que :

$$u'(\alpha) = \mathbb{E}\left(\left(\sum_{i=1}^n \alpha\lambda_i(G)^2 - \sum_{i=1}^n \lambda_i(G)G_i\right)e^{-\sum_{i=1}^n \alpha\lambda_i(G)G_i + \frac{1}{2}\sum_{i=1}^n \alpha^2\lambda_i(G)^2}\phi(G)^2\right).$$

On notera, en particulier, que $u'(0)$ vaut $-\mathbb{E}\left(\left(\sum_{i=1}^n \lambda_i(G)G_i\right)\phi(G)^2\right)$.

9. Montrer que u est deux fois différentiable et que u'' peut s'exprimer sous la forme

$$u''(\alpha) = \mathbb{E}\left(\left\{\sum_{i=1}^n \lambda_i^2(G) + \left(\sum_{i=1}^n \alpha\lambda_i(G)^2 - \sum_{i=1}^n \lambda_i(G)G_i\right)^2\right\} \times \right.$$
$$\left. \times\, e^{-\sum_{i=1}^n \alpha\lambda_i(G)G_i + \frac{1}{2}\sum_{i=1}^n \alpha^2\lambda_i(G)^2}\phi(G)^2\right)$$

10. On suppose dans la fin de ce problème que la fonction $\phi|\lambda|$ est non nulle sur un ensemble de mesure de Lebesgue non nulle. Montrer que u est strictement convexe et que $\lim_{\alpha\to\pm\infty} u(\alpha) = +\infty$.

11. Montrer que si $u'(0) \neq 0$, il existe un α tel que $u(\alpha) < u(0)$, puis que, pour ce α, $\alpha\lambda$ est un choix qui permet de réduire la variance de la méthode de Monte-Carlo.

12. Montrer, en utilisant la stricte convexité de u, que si $u'(0) = 0$, 0 réalise le minimum u. Interpréter ce résultat.

13. On suppose, pour cette question, que λ est un vecteur de fonctions paires (c'est à dire qui vérifie, pour tout g, $\lambda(-g) = \lambda(g)$ non nul et que la fonction ϕ^2 est paire. Montrer que, sous ces hypothèses, λ ne permet pas de réduire la variance.

Problem 16. Simulation directionnelle et stratification
Soit $G = (G_1, \ldots, G_d)$ un vecteur constitué de d gaussiennes centrées réduites indépendantes et H un fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}$ et on cherche à calculer la probabilité :

$$p = \mathbb{P}\left(H(G) \geq \lambda\right),$$

pour une valeur λ positive donnée.

1. Décrire la méthode de Monte-Carlo habituelle permettant d'estimer p et expliquer comment l'on peut évaluer l'erreur de la méthode.

2. On suppose que l'on cherche à approche une valeur de p de l'ordre de 10^{-10}. Donner une évaluation (grossière) du nombre minimal de tirages n permettant d'évaluer p à 20% près. Lorsque le calcul de H demande un temps de l'ordre d'une seconde, qu'en concluez vous ? (1 an $\approx 3 \times 10^7$ secondes)

3. On écrit le vecteur G sous la forme $G = RA$ où $R = |G|$ et $A = G/|G|$ prend ses valeurs dans la sphère unité de $\mathbb{R}^d$, $S_d = \{x \in \mathbb{R}^d, |x| = 1\}$.

   Montrer que R^2 suit une loi du χ^2 à d degrés de liberté et que A suit une loi invariante par rotation dont le support est inclu dans S_d (On admettra que cela implique que la loi de A est la loi uniforme sur la sphère).

4. A l'aide du changement de variable polaire généralisé $(g_1, \ldots, g_d) \to (r, \theta_1, \ldots, \theta_{d-1})$

$$\begin{cases} g_1 & = r\cos\theta_1 \\ g_2 & = r\sin\theta_1\cos\theta_2 \\ & \cdots \\ g_{d-1} & = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{d-2}\cos\theta_{d-1} \\ g_d & = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{d-2}\sin\theta_{d-1} \end{cases}$$

   montrer que R et A sont des variables aléatoires indépendantes.

5. On suppose, à partir de maintenant, que pour tout $a \in S_d$ la fonction $r \to H(ra)$ est strictement croissante, que $H(0) = 0$ et $\lim_{r\to+\infty} H(ra) = +\infty$.

   Soit $a \in S_d$, on pose $\phi(a) = \mathbb{P}(H(Ra) \geq \lambda)$. Calculer explicitement $\phi(a)$ en fonction de la fonction de répartition de la loi du χ^2 à d degrés de liberté et de la solution de $H(r^*a) = \lambda$.

6. Soit $(A_i, i \geq 0)$ une suite de variables aléatoires indépendantes et de loi uniforme sur la sphère unité. Montrer que $\lim_{n\to+\infty} \sum_{i=1}^n \phi(A_i) = p$, au sens de la convergence presque sûre.

7. Montrer que $\text{Var}\ (\phi(A_1)) \leq \text{Var}\ \left(\mathbf{1}_{H(G)\geq\lambda}\right)$ et comparer la vitesse de convergence de l'estimateur de la question 1 et celui de la question précédente.

8. Nous allons maintenant réduire la variance de l'estimateur précédent par une technique de stratification.

   On considère les 2^d quadrants définit, pour $\epsilon \in \{-1,+1\}^d$, par :

   $$C_\epsilon = \{x \in S_d, \epsilon_1 x_1 \geq 0, \ldots, \epsilon_d x_d \geq 0\}.$$

   Montrer que $\mathbb{P}(A_1 \in C_\epsilon \cap C_{\epsilon'}) = 0$, pour $\epsilon \neq \epsilon'$. En déduire la valeur de $\rho_\epsilon = \mathbb{P}(A_1 \in C_\epsilon)$?

9. On considère l'estimateur I_n stratifié dans les strates $(C_\epsilon, \epsilon \in \{-1,+1\}^d)$:

   $$I_n = \sum_{\epsilon \in \{-1,+1\}^d} \frac{\rho_\epsilon}{n_\epsilon} \sum_{k=1}^{n_\epsilon} \phi(A_k^{(\epsilon)}),$$

   où $n_\epsilon = n w_\epsilon(n)$ avec $w_\epsilon(n) \geq 0$ et $\sum_{\epsilon \in \{-1,+1\}^d} w_\epsilon(n) = 1$ et où les $(A_k^{(\epsilon)}, k \geq 0)$ sont des suites de variables aléatoires indépendantes suivant une loi uniforme sur C_ϵ, ces suites étant indépendantes entre elles.

   Calculer $\mathrm{Var}\ (I_n)$ en fonction de n et des $(w_\epsilon(n), \epsilon \in \{-1,+1\}^d)$, des $(\rho_\epsilon, \epsilon \in \{-1,+1\}^d)$ et des $(\sigma_\epsilon, \epsilon \in \{-1,+1\}^d)$ où $\sigma_\epsilon^2 = \mathrm{Var}\ (A_1^{(\epsilon)})$.

10. Montrer l'inégalité :

    $$\left(\sum_{i=1}^N \alpha_i a_i\right)^2 \leq \sum_{i=1}^N \alpha_i a_i^2,$$

    pour $(\alpha_i, a_i, 1 \leq i \leq N)$ des réels positifs tel que $\sum_{i=1}^N \alpha_i = 1$.

    En déduire que $\mathrm{Var}\ (I_n) \geq \frac{1}{n}\left(\sum_{\epsilon \in \{-1,+1\}^d)} \rho_\epsilon \sigma_\epsilon\right)^2$.

    Comment peut on choisir les $(w_\epsilon(n), \epsilon \in \{-1,+1\}^d)$ pour réaliser la borne inférieure précédente ?

11. Proposer une méthode de simulation selon la loi uniforme sur C_ϵ ?

12. On fixe une fois pour toute les w_ϵ (indépendement de n) par :

    $$w_\epsilon = \frac{\rho_\epsilon \sigma_\epsilon}{\sum_{\epsilon \in \{-1,+1\}^p} \rho_\epsilon \sigma_\epsilon}.$$

    On pose $n_\epsilon = [n w_\epsilon]$. Montrer que, au sens de la convergence presque sûre :

    $$\lim_{n \to +\infty} I_n = p.$$

13. Montrer que la limite en loi des variables aléatoires : $\sqrt{n}(I_n - p)$ est une gaussienne centrée de variance $\left(\sum_{\epsilon \in \{-1,+1\}^d)} \rho_\epsilon \sigma_\epsilon\right)^2$.

    Comment peut on utiliser ce résultat pour estimer l'erreur commise lorsque l'on estime p à l'aide de I_n ?

Problem 17. Méthode du "carré latin"

Le cas de la dimension 1 On considére un suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$, $U = (U_i, i \geq 1)$ et une permutation σ tirée uniformement sur l'ensemble des permutations de $\{1,\ldots,N\}$ et indépendante de la suite U. Pour N entier fixé et pour $1 \leq i \leq N$, on pose :

$$Y_i^N = \frac{i - U_i}{N} \text{ et } X_i^N = Y_{\sigma(i)}^N.$$

On considére une fonction mesurable bornée f et l'on cherche à comparer l'estimateur classique :

$$I_0(N) = \frac{1}{N}\left(f(U_1) + \cdots + f(U_N)\right),$$

avec :

$$I_1(N) = \frac{1}{N}\left(f(X_1^N) + \cdots + f(X_N^N)\right)$$

1. Quel est la limite de $I_0(N)$ lorsque N tends vers $+\infty$? Comment peut on estimer l'erreur commise lorsque l'on approxime cette limite par $I_0(N)$?
2. Pour un i fixé,$1 \leq i \leq N$, identifier la loi de X_i^N. En déduire que $\mathbb{E}\left(I_1(N)\right) = \mathbb{E}(f(U_1))$
3. Montrer que $I_1(N) = \frac{1}{N}\left(f(Y_1^N) + \cdots + f(Y_N^N)\right)$, et en déduire que

$$N\text{Var }\left(I_1(N)\right) = \int_0^1 f^2(s)ds - N\sum_{k=1}^{N}\left(\int_{(k-1)/N}^{k/N} f(s)ds\right)^2,$$

4. Vérifier que $N\text{Var }\left(I_1(N)\right) = \dfrac{N}{2}\displaystyle\sum_{i=1}^{N}\int_{(i-1)/N}^{i/N}\int_{(i-1)/N}^{i/N}(f(s) - f(s'))^2 dsds'$. Lorsque f est une fonction continue, montrer que $N\text{Var }\left(I_1(N)\right)$ tend vers 0 lorsque N tends vers $+\infty$. Que vaut $N\text{Var }\left(I_0(N)\right)$?
5. Pour f continue, en quel sens $I_1(N)$ convergente t'il vers $\mathbb{E}(f(U))$?
6. Quel estimateur vous parait préférable, $I_0(N)$ ou $I_1(N)$, du point de vue de sa vitesse de convergence en moyenne quadratique ?

   D'un point de vue pratique quels vous paraissent les inconvénients de $I_1(N)$ par rapport à $I_1(N)$?
7. Comment peut-on interpréter l'estimateur $I_0(N)$ en terme de méthode de stratification ?

Le cas de la dimension 2 On considère maintenant le cas bidimensionnel. U désigne un suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]^2$, $U = ((U_i^1, U_i^2), i \geq 1)$ et deux permutations indépendantes σ_1 et σ_2 tirée uniformement sur l'ensemble des permutations de $\{1,\ldots,N\}$ et indépendantes de la suite U. Pour N entier fixé et pour $1 \leq i,j \leq N$, on pose :

$$Y_{i,j}^N = \left(\frac{i - U_i^1}{N}, \frac{j - U_j^2}{N}\right) \text{ et } X_i^N = Y_{\sigma_1(i),\sigma_2(i)}^N.$$

On pose, comme dans la première partie, pour f une fonction bornée de $[0,1]^2$ dans $\mathbb{R}$:

$$I_0(N) = \frac{1}{N}\left(f(U_1) + \cdots + f(U_N)\right) \text{ et } I_1(N) = \frac{1}{N}\left(f(X_1^N) + \cdots + f(X_N^N)\right)$$

1. Montrer que
$$\mathbb{E}\left(f(X_i^N)\right) = \frac{1}{N^2}\sum_{\substack{1\le k\le N\\ 1\le l\le N}} \mathbb{E}\left(f(Y_{k,l})\right) = \mathbb{E}\left(f(U_1)\right)$$

2. Montrer que, pour $i \neq j$:
$$\mathbb{E}\left(f(X_i^N)f(X_j^N)\right) = \frac{1}{N^2(N-1)^2}\sum_{\substack{1\le k_1\neq k_2\le N\\ 1\le l_1\neq l_2\le N}} \mathbb{E}\left(f(Y_{k_1,l_1})\right)\mathbb{E}\left(f(Y_{k_2,l_2})\right),$$
puis que :
$$\mathbb{E}\left(f(X_i^N)f(X_j^N)\right) = \frac{1}{N^2(N-1)^2}\Big\{N^4\mathbb{E}(f(U_1))^2 - \sum_{\substack{1\le k\le N\\ 1\le l_1,l_2\le N}} \mathbb{E}\left(f(Y_{k,l_1})\right)\mathbb{E}\left(f(Y_{k,l_2})\right) - \sum_{\substack{1\le k_1,k_2\le N\\ 1\le l\le N}} \mathbb{E}\left(f(Y_{k_1,l})\right)\mathbb{E}\left(f(Y_{k_2,l})\right) + \sum_{\substack{1\le k\le N\\ 1\le l\le N}} \mathbb{E}\left(f(Y_{k,l})\right)^2\Big\}.$$

3. En déduire que, pour f continue :
$$\lim_{N\to+\infty}(N-1)\mathrm{Cov}\left(f(X_1^N), f(X_2^N)\right) = 2\mathbb{E}(f(U))^2 - \int_{[0,1]^3} f(x,y)f(x,y')dxdydy' - \int_{[0,1]^3} f(x,y)f(x',y)dxdx'dy.$$

4. Montrer que :
$$\mathrm{Var}\ \left(\mathbb{E}(f(U)|U_1)\right) = \int_{[0,1]^3} f(x,y)f(x,y')dxdydy' - \mathbb{E}\left(f(U)\right)^2,$$
$$\mathrm{Var}\ \left(\mathbb{E}(f(U)|U_2)\right) = \int_{[0,1]^3} f(x,y)f(x',y)dxdx'dy - \mathbb{E}\left(f(U)\right)^2.$$

5. En déduire que :
$$\lim_{N\to+\infty} N\mathrm{Var}\left(I_1(N)\right) = \mathrm{Var}\left(f(U)\right) - \mathrm{Var}\ \left(\mathbb{E}(f(U)|U_1)\right) - \mathrm{Var}\ \left(\mathbb{E}(f(U)|U_2)\right).$$

6. Quel est l'intérêt d'utiliser $I_1(N)$ plutôt que $I_0(N)$? Quels en sont les inconvénients ?

7. Vérifier que :
$$\mathrm{Var}\ \left\{f(U) - \mathbb{E}(f(U)|U_1) - \mathbb{E}(f(U)|U_2)\right\} = \mathrm{Var}\left(f(U)\right) - \mathrm{Var}\ \left(\mathbb{E}(f(U)|U_1)\right) - \mathrm{Var}\ \left(\mathbb{E}(f(U)|U_2)\right).$$

Problem 18. Annulation du bais d'une suite d'estimateurs On se donne une variable aléatoire Y à valeurs dans $\mathbb{R}$ de carré intégrable.

1. On suppose, tout d'abord, que l'on sait simuler exactement la variable aléatoire Y. Pour un ϵ réel positif, montrer que le nombre $N(\epsilon)$ de tirages aléatoires de Y indépendants permettant d'obtenir (avec une grande probabilité) une approximation de $\mathbb{E}(Y)$ avec une erreur d'environ ϵ croit comme Cte$/\epsilon^2$.

On suppose maintenant que l'on ne sait pas simuler exactement Y mais seulement qu'il existe une suite de variables aléatoires de carré intégrable $(Y_n, n \geq 0)$ facilement simulables[1] telles que

$$\lim_{n\to+\infty} \mathbb{E}\left((Y_n - Y)^2\right) = 0. \tag{1.38}$$

2. Montrer que le biais $b_n = \mathbb{E}(Y) - \mathbb{E}(Y_n)$ tend vers 0 lorsque n tend vers $+\infty$. Vérifier que $\lim_{n\to+\infty} \mathbb{E}((Y_n - Y)Y) = 0$. En déduire que $\lim_{n\to+\infty} \mathbb{E}(Y_n^2) = \mathbb{E}(Y^2)$ et en conclure que $\lim_{n\to+\infty} \mathrm{Var}\,(Y_n) = \mathrm{Var}\,(Y)$.

3. On se place dans un cas où l'on a une estimation a priori du bias du type $|b_n| \leq \frac{C}{n^\alpha}$, avec $\alpha > 0$. On choisit un $n > 0$ et le nombre de tirage N, $(Y_n^1, \ldots, Y_n^N)$, selon la loi de Y_n. On approxime $\mathbb{E}(Y)$ par $\bar{Y}^{n,N} = \frac{1}{N}\sum_{i=1}^N Y_n^i$. Montrer que, pour N grand, avec une probabilité proche de 1, on a

$$|\bar{Y}^{n,N} - \mathbb{E}(Y)| \leq R(n,N) := \frac{C}{n^\alpha} + 2\frac{K}{\sqrt{N}}.$$

4. On prend comme mesure de l'erreur $R(n,N)$ et l'on suppose que le coût de N tirages selon la loi de Y_n est donné par Coût$(n,N) = C_1 nN$. On s'impose une contrainte sur le coût total du calcul C_{total}. On s'intéresse à l'erreur minimale $R(n,N)$ que l'on peut obtenir sous la contrainte de coût Coût$(n,N) = C_{\text{total}}$, définie par :

$$\text{Error}(C_{\text{total}}) = \min_{n,N,\text{Cost}(n,N)=C_{\text{total}}} R(n,N)$$

Montrer que Error$(C_{\text{total}}) =$ Cte $\times\, C_{\text{total}}^{-\frac{\alpha}{2\alpha+1}}$. En déduire que le coût minimal nécessaire pour obtenir une précision ϵ croît comme Cte $\times\, \epsilon^{-\frac{2\alpha+1}{\alpha}}$. En comparant avec le résultat de la quoi question 1, expliquer en quoi la présence du biais dégrade la performance de l'algorithme de simulation.

Comment se débarasser du biais ? Nous allons voir que l'on peut construire, à partir de la suite $(Y_n, n \geq 0)$ une variable aléatoire simulable et sans biais $\tilde{Y}$ (c'est-à-dire telle que $\mathbb{E}() = \mathbb{E}(Y)$) au prix toutefois d'une augmentation de la variance.

On considère, pour cela, une variable aléatoire τ à valeurs dans $\mathbb{N}^*$ *indépendant* de la suite $(Y_n, n \geq 0)$ dont la loi est donnée par

$$\mathbb{P}(\tau = n) = p_n, \text{ avec } p_n > 0 \text{ pour tout } n \geq 1. \tag{1.39}$$

On note $\Delta Y_n = Y_n - Y_{n-1}$, pour $n \geq 1$ (lorsque $n = 1$, on pose $Y_0 = 0$). On définit $\tilde{Y}$ par

$$\tilde{Y} = \frac{\Delta Y_\tau}{p_\tau} = \sum_{n\geq 1} \frac{\Delta Y_n}{p_n} \mathbf{1}_{\{\tau=n\}}.$$

[1] C'est le cas lorsque l'on discrétise une équation différentielle stochastique.

5. Montrer que, sous l'hypothèse (1.39), $\mathbb{E}(|\tilde{Y}|) = \sum_{n\geq 1} \mathbb{E}(|\Delta Y_n|)$. En déduire que lorsque $\mathbb{E}(|\tilde{Y}|) < +\infty$, on a $\mathbb{E}(\tilde{Y}|Y_n, n \geq 1) = Y$ et $\mathbb{E}(\tilde{Y}) = \mathbb{E}(Y)$.

6. Donner un exemple de suite *déterministe* $(\Delta Y_n, n \geq 1)$ où $Y = \sum_{n\geq 1} \Delta Y_n$ converge mais où $\mathbb{E}(|\tilde{Y}|) = +\infty$, pour tout choix de p vérifiant l'hypothèse (1.39).

7. Montrer que l'on a
$$\mathbb{E}(\tilde{Y}^2) = \sum_{n\geq 1} \frac{\mathbb{E}\left((\Delta Y_n)^2\right)}{p_n},$$
et que lorsque $\mathbb{E}(\tilde{Y}^2) < +\infty$, $\mathrm{Var}\,(\tilde{Y}) \geq \mathrm{Var}\,(Y)$.

8. On cherche à approximer $\mathbb{E}(Y)$ où $Y = f(X_T)$ et $Y_n = f(X_T^{(n)})$ où X une solution d'une EDS et $X^{(n)}$ son approximation par un schéma d'Euler de pas $h_n = 1/2^n$. Montrer (en supposant les coefficients de l'EDS et f aussi réguliers que souhaité) que $\mathbb{E}\left(\left|f(X_T^{(n)}) - f(X_T)\right|^2\right) \leq C\rho^n$, ρ étant un réel positif inférieur à 1 que l'on précisera. En déduire que $\mathbb{E}((\Delta Y_n)^2) \leq 4C\rho^n$.

   Par ailleurs le coût d'une simulation de Y_n est donné par Cte $\times\, 2^n$. Montrer que, si l'on néglige le temps de simulation de τ, le coût moyen d'une simulation de $\tilde{Y}$ est proportionnel à $\sum_{n\geq 1} p_n 2^n$.

9. On suppose que, comme dans l'exemple précédent, $\mathbb{E}\left((\Delta Y_n)^2\right) \leq C\rho^n$, avec $\rho < 1$ et que le coût d'une simulation de Y_n est donné par par $c\alpha^n$ avec $\alpha > 1$.

   On suppose enfin que la loi de τ est une loi géométrique[2] de paramètre q avec $0 < q < 1$ (on note $q' = 1 - q$). Vérifier que, si $q' > \rho$, $\mathbb{E}((\tilde{Y})^2) < +\infty$ et que :
$$\mathbb{E}(\tilde{Y}^2) \leq \frac{C\rho q'}{q(q' - \rho)}.$$

   Par ailleurs, montrer que le coût moyen d'une simulation de $\tilde{Y}$, $C_{\text{moy}} = \mathbb{E}(c\alpha^\tau)$, est fini lorsque $\alpha q' < 1$ et qu'il alors égal à $c\alpha q/(1 - \alpha q')$.

   En supposant[3] que la performance de la simulation est mesurée par le produit $\mathbb{E}(C_{\text{moy}}) \times \mathbb{E}((\tilde{Y})^2)$ montrer que l'on doit avoir $\rho\alpha < 1$ pour espérer obtenir une méthode efficace et que, dans ce cas, le meilleur choix de q' est $q' = \sqrt{\rho/\alpha}$.

[2] i.e. pour tout $n \geq 1$, $p_n = q(1 - q)^n$.

[3] Voir, pour des élèments de justification, *Glynn P.W., Whit W., 1992, The asymptotic efficiency of simulation estimators, Oper. Res. 40(3):505-520.*

Chapter 2

Introduction to stochastic algorithms

2.1 A reminder on martingale convergence theorems

$\mathcal{F} = (\mathcal{F}_n, n \geq 0)$ denote an increasing sequence of σ-algebra of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 2.1.1. *A sequence of real random variable $(M_n, n \geq 0)$ is a $\mathcal{F}$-*martingale *if and only if, for all $n \geq 0$:*

- *M_n is $\mathcal{F}_n$-measurable*
- *M_n is integrable, $\mathbb{E}(|M_n|) < +\infty$.*
- *$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$.*

- When, for all $n \geq 0$, $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leq M_n$ the sequence is called a *super-martingale*.
- When, for all $n \geq 0$, $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \geq M_n$ the sequence is called a *sub-martingale*.

Definition 2.1.2. *An $\mathcal{F}$-stopping time is a random variable τ taking its values in $\mathbb{N} \cup \{+\infty\}$ such that, for all $n \geq 0$, $\{\tau \leq n\} \in \mathcal{F}_n$.*

Given a stopping time τ and a process $(M_n, n \geq 0)$, we can define a stopped process by $M_{n\wedge\tau}$. It is easy to check that a stopped martingale (resp. sub, super) remains an $\mathcal{F}$-martingale (resp. sub, super).

Exercise 19. Check it using the fact that :

$$M_{(n+1)\wedge\tau} - M_{n\wedge\tau} = \mathbf{1}_{\{\tau>n\}}(M_{n+1} - M_n).$$

Convergence of super-martingale Almost sure convergence of super-martingale can be obtained under weak conditions.

Theorem 2.1.3. *Let $(M_n, n \geq 0)$ be a positive super-martingale with respect to $\mathcal{F}$ (i.e. the conditional expectation is decreasing $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leq M_n$, then M_n converge almost surely to a random variable M_∞ when n goes to $+\infty$.*

For a proof see [Williams(1991)].

Remark 2.1.1. The previous result remain true if, for all n, $M_n \geq -a$, with $a \geq 0$ (as $M_n + a$ is a positive super-martingale).

To obtain L^p-convergence we need stronger assumptions.

Theorem 2.1.4. *Assume $(M_n, n \geq 0)$ is a martingale with respect to $\mathcal{F}$, bounded in L^p for a $p > 1$ (i.e. $\sup_{n\geq 0} \mathbb{E}(|M_n|^p) < +\infty$), then then M_n converge almost surely and in L^p to a random variable M_∞ when n goes to $+\infty$.*

Remark 2.1.2. The case $p = 1$ is a special case, if $(M_n, n \geq 1)$ is bounded in L^1, M_n converge to M_∞ almost surely but we need to add the uniform integrability of the sequence to obtain convergence in L^1.

For a proof of these theorems see for instance [Williams(1991)] chapter 11.

2.1.1 Consequence and examples of uses

We first remind a deterministic lemma know as Kronecker Lemma.

Lemma 2.1.3 (Kronecker Lemma)**.** *Let $(A_n, n \geq 1)$ be an* increasing *sequence of strictly positive real numbers, such that $\lim_{n\to+\infty} A_n = +\infty$.*

Let $(\epsilon_k, k \geq 1)$ be a sequence of real numbers, such that $S_n = \sum_{i=1}^n \epsilon_k/A_k$ converge when n goes to $+\infty$.

Then :

$$\lim_{n\to+\infty} \frac{1}{A_n} \sum_{i=1}^{n} \epsilon_k = 0.$$

Proof. As S_n converge, we can write $S_n = S_\infty + \eta_n$, η_n being a sequence converging to 0 when n goes to $+\infty$. Moreover, using Abel transform :

$$\sum_{k=1}^{n} \epsilon_k = \sum_{i=1}^{n} A_k \frac{\epsilon_k}{A_k} = A_n S_n - \sum_{k=1}^{n-1} S_k (A_{k+1} - A_k).$$

So :

$$\begin{aligned} \frac{1}{A_n} \sum_{k=1}^{n} \epsilon_k &= S_\infty + \eta_n - S_\infty \frac{A_n - A_1}{A_n} - \frac{1}{A_n} \sum_{k=1}^{n-1} \eta_k (A_{k+1} - A_k) \\ &= S_\infty \frac{A_1}{A_n} - \frac{1}{A_n} \sum_{k=1}^{n-1} \eta_k (A_{k+1} - A_k). \end{aligned} \tag{2.1}$$

The first two terms converge to 0 (as A_n converge to $+\infty$ and η_n to 0 when n goes to $+\infty$). The third one is a (generalized) Cesaro mean of a sequence which converges to 0 so it converges to 0. □

A proof of the strong law of large number As an application of martingale convergence theorem, we will give a short proof of the strong law of large numbers for a square integrable random variable X. Let $(X_n, n \geq 1)$ be a sequence of independent random variables following the law of X. Denote by $X'_n = X_n - \mathbb{E}(X)$.

Let $\mathcal{F}_n = \sigma(X_k, k \leq n)$ and M_n be :

$$M_n = \sum_{k=1}^{n} \frac{X'_k}{k}.$$

Note that, using independence and $\mathbb{E}(X'_k) = 0$, M_n is an $\mathcal{F}$-martingale. Moreover, using once again independence, we get :

$$\mathbb{E}(M_n^2) = \text{Var } (X) \sum_{k=1}^{n} \frac{1}{k^2} < K < +\infty.$$

So the martingale M is bounded in L^2, and using theorem 2.1.4 converge almost surely to M_∞. Using Kronecker lemma this implies that :

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} X'_k = 0,$$

or $\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}(X)$.

We can relax the L^2 hypothesis to obtain the full strong law of large numbers under the traditional L^1 condition. See the following exercise (and [Williams(1991)] for a solution if needed).

Exercise 20. Suppose that $(X_n, n \geq 1)$ are independent variables following the law of X, with $\mathbb{E}(|X|) < +\infty$. Define Y_n by :

$$Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}.$$

1. Prove that $\lim_{n\to+\infty} \mathbb{E}(Y_n) = E(X)$.

2. Prove that $\sum_{n=1}^{+\infty} \mathbb{P}(|X| > n) \leq \mathbb{E}(|X|)$, and deduce that

$$\mathbb{P}(\text{Exists } n_0(\omega), \text{ for all } n \geq n_0, X_n = Y_n) = 1.$$

3. Prove that :

$$\sum_{n \geq 1} \frac{\text{Var } (Y_n)}{n^2} \leq \mathbb{E}\left(|X|^2 f(|X|)\right),$$

where

$$f(z) = \sum_{n \geq \max(1,z)} \frac{1}{n^2} \leq \frac{2}{\max(1,z)}.$$

Deduce that $\sum_{n\geq 1} \text{Var } (Y_n)/n^2 \leq 2\mathbb{E}(|X|) < +\infty$.

4. Let $W_n = Y_n - \mathbb{E}(Y_n)$, prove that $\sum_{k\geq n} \frac{W_k}{k}$ converge when n goes to $+\infty$, and deduce, using Kronecker lemma, that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k\leq n} W_k = 0,$$

then deduce $\lim_{n\to+\infty} \frac{1}{n} \sum_{k\leq n} Y_n = \mathbb{E}(X)$.

5. Using the result of question 2, prove that $\lim_{n\to+\infty} \frac{1}{n} \sum_{k\leq n} X_n = \mathbb{E}(X)$

An extension of the super-martingale convergence theorem For the proof of the convergence of stochastic algorithms we will need on extension of the super-martingale convergence theorem 2.1.3 known as Robbins-Sigmund lemma.

Lemma 2.1.4 (Robbins-Sigmund lemma)**.** *Assume V_n, a_n, b_n, c_n are sequences of positive random variables adapted to $(\mathcal{F}_n, n \geq 0)$ and that, moreover, fpr all $n \geq 0$:*

$$\mathbb{E}\left(V_{n+1}|\mathcal{F}_n\right) \leq (1 + a_n)V_n + b_n - c_n,$$

then, on $\left\{\sum_{n\geq 1} a_n < +\infty, \sum_{n\geq 1} b_n < +\infty\right\}$, V_n converge to a random variable V_∞ and $\sum_{n\geq 1} c_n < +\infty$.

Proof. Let :

$$\alpha_n = \frac{1}{\prod_{i=1}^n (1 + a_k)}.$$

Then define $V'_n = \alpha_{n-1}V_n$, $b'_n = \alpha_n b_n$, $c'_n = \alpha_n c_n$. Clearly the hypothesis can be rewritten as :

$$\mathbb{E}\left(V'_{n+1}|\mathcal{F}_n\right) \leq V'_n + b'_n - c'_n.$$

This means, if $X_n = V'_n - \sum_{k=0}^{n-1}(b'_k - c'_k)$, that X_n is a super-martingale.

Now consider the stopping time τ_a :

$$\tau_a = \inf\left\{n \geq 0, \sum_{k=0}^{n-1}(b'_k - c'_k) \geq a\right\},$$

(the infimum is $+\infty$ if the set is empty).

τ_a is a stooping time such that, if $n \leq \tau_a$, then $X_n \geq -a$. So $X_{n\wedge\tau_a}$ is a super-martingale, bounded from below, so it converges. So we can conclude that $\lim_{n\to+\infty} X_n$ exists on the set $\cup_{a>0}\{\tau_a = +\infty\}$. But, as a_n is positive, α_n is a positive, decreasing sequence, so it converges to $\bar{\alpha}$ when n goes to $+\infty$. Moreover :

$$\alpha_n = \frac{1}{\prod_{i=1}^n (1 + a_k)} = e^{-\sum_{k=0}^n \log(1+a_k)} \geq e^{-\sum_{k=0}^n a_k}.$$

So on the set $\left\{\sum_{n\geq 1} a_n < +\infty, \sum_{n\geq 1} b_n < +\infty\right\}$, $\bar{\alpha} > 0$. It follows that $\sum_{n\geq 1} b'_n < +\infty$, and as, $c'_n > 0$ that, for all $n \geq 0$

$$\sum_{k=0}^n b'_k - \sum_{k=0}^n c'_k \leq \sum_{k=0}^n b'_k < +\infty.$$

So for $a > \sum_{k=0}^{n} b'_k$, $T_a = +\infty$, and we can conclude that, on the set

$$\left\{ \sum_{n\geq 1} a_n < +\infty, \sum_{n\geq 1} b_n < +\infty \right\},$$

X_n converge to X_∞ when n goes to ∞. Moreover :

$$0 \leq \sum_{k=0}^{n} c'_k = X_n - V'_n + \sum_{k=0}^{n} b'_k,$$

so, as $V_n \geq 0$, and X_n and $\sum_{k=0}^{n} b'_k$ converge when n goes to $+\infty$, we get that $\sum_{k\geq 1} c'_k$ is finite, then, as α_n converge to $\bar{\alpha} > 0$ that $\sum_{k\geq 1} c_k$ is finite. □

2.2 Almost sure convergence for some stochastic algorithms

2.2.1 Almost sure convergence for the Robbins-Monro algorithm

Theorem 2.2.1. *Let f be a function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Assume that :*

H1 *Hypothesis on the function f. f is continuous and there exists $x^* \in \mathbb{R}^d$, such that $f(x^*) = 0$ and $\langle f(x), x - x^* \rangle > 0$ for $x \neq x^*$.*

H2 *Hypothesis on the step size γ. $(\gamma_n, n \geq 1)$ is a decreasing sequence of positive real numbers such that $\sum_{n\geq 1} \gamma_n = +\infty$ and $\sum_{n\geq 1} \gamma_n^2 < +\infty$.*

H3 *Hypothesis on the sequence Y. $(\mathcal{F}_n, n \geq 0)$ is a filtration on a probability space and $(Y_n, n \geq 1)$ is sequence of random variables on this probability space such that*

H3.1 $\mathbb{E}\left(Y_{n+1} | \mathcal{F}_n\right) = f(X_n)$,

H3.2 $\mathbb{E}\left(|Y_{n+1} - f(X_n)|^2 | \mathcal{F}_n\right) \leq \sigma^2(X_n)$ *where*

$$s^2(x) = \sigma^2(x) + f^2(x) \leq K(1 + |x|^2).$$

Define the sequence $(X_n, n \geq 0)$, by $X_0 = x_0$, where x_0 is a point in $\mathbb{R}^d$ and, for $n \geq 0$

$$X_{n+1} = X_n - \gamma_n Y_{n+1}.$$

Then $\lim_{n\to+\infty} X_n = x^*$, a.s..

Remark 2.2.1.

- The main application of this algorithm arise when $f(x)$ can be written as

$$f(x) = \mathbb{E}\left(F(x, U)\right),$$

where U follows a known law and F is a function. In this case Y_n is defined as $Y_n = F(X_n, U_{n+1})$ where $(U_n, n \geq 1)$ is a sequence of independent random variables following the law of U. Clearly, if $\mathcal{F}_n = \sigma(U_1, \ldots, U_n)$, we have

$$\mathbb{E}\left(F(X_n, U_{n+1}) | \mathcal{F}_n\right) = f(X_n).$$

Moreover

$$\mathbb{E}\left(|F(X_n, U_{n+1}) - f(X_n)|^2 | \mathcal{F}_n\right) = \sigma^2(X_n),$$

where $\sigma^2(x) = \mathbb{E}\left(|F(x,U) - f(x)|^2\right)$. So **H3.2** is an hypothesis on behavior of the expectation and the variance of $F(x,U)$ when $|x|$ goes to $+\infty$.

- The hypothesis **H1** is fulfilled when $f(x) = \nabla V(x)$ where V is a strictly convex function and there exists x^* which minimize $V(x)$. This will be the case in most of our examples.

Proof. First note that hypothesis **H3** implies that

$$\begin{aligned} \mathbb{E}\left(Y_{n+1}^2 | \mathcal{F}_n\right) &= \mathbb{E}\left((Y_{n+1} - f(X_n))^2 | \mathcal{F}_n\right) + \mathbb{E}\left(f(X_n)^2\right) \\ &\leq s^2(X_n) \leq K(1 + |X_n|^2) \leq K'(1 + |X_n - x^*|^2). \end{aligned} \tag{2.2}$$

Let $V(x) = |x - x^*|^2$ and set $V_n = V(X_n)$. Clearly :

$$V_{n+1} = V_n + \gamma_n^2 Y_{n+1}^2 - 2\gamma_n \langle X_n - x^*, Y_{n+1} \rangle.$$

Taking the conditional expectation we obtain :

$$\mathbb{E}\left(V_{n+1} | \mathcal{F}_n\right) = V_n + \gamma_n^2 \mathbb{E}\left(Y_{n+1}^2 | \mathcal{F}_n\right) - 2\gamma_n \langle X_n - x^*, \mathbb{E}\left(Y_{n+1} | \mathcal{F}_n\right) \rangle,$$

and using hypothesis **H3** we get

$$\mathbb{E}\left(V_{n+1} | \mathcal{F}_n\right) \leq V_n + \gamma_n^2 s^2(X_n) - 2\gamma_n \langle X_n - x^*, f(X_n) \rangle.$$

So, using inequality 2.2, we obtain :

$$\begin{aligned} \mathbb{E}\left(V_{n+1} | \mathcal{F}_n\right) &\leq V_n + K\gamma_n^2 (1 + V_n) - 2\gamma_n \langle X_n - x^*, f(X_n) \rangle \\ &= V_n \left(1 + K\gamma_n^2\right) + K\gamma_n - 2\gamma_n \langle X_n - x^*, f(X_n) \rangle. \end{aligned} \tag{2.3}$$

So, using Robbins-Sigmund lemma, with $a_n = b_n = K\gamma_n^2$, and $c_n = \gamma_n \langle X_n - x^*, f(X_n) \rangle$ (which is positive because of hypothesis **H1**), we get (by hypotheses **H2** $\sum_{n \geq 1} \gamma_n^2 < +\infty$) that, both

- V_n converge to V_∞, almost surely,
- $\sum_{n \geq 1} \gamma_n \langle X_n - x^*, f(X_n) \rangle < +\infty$.

Obviously V_∞ is a positive random variable, and we only need to check that this random is equal to 0.

Assume that $\mathbb{P}(V_\infty > 0) > 0$, then on the set $\{V_\infty > 0\}$ we have $0 < V_\infty/2 \leq V_n \leq 3V_\infty/2$ for $n \geq n_0(\omega)$, so

$$\sum_{n \geq 1} \gamma_n \langle X_n - x^*, f(X_n) \rangle \geq \inf_{V_\infty/2 \leq |x - x^*|^2 \leq 3V_\infty} \langle x - x^*, f(x) \rangle \sum_{n \geq n_0} \gamma_n.$$

But $\sum_{n \geq n_0} \gamma_n = +\infty$ and $\inf_{V_\infty/2 \leq |x-x^*|^2 \leq 3V_\infty} \langle x - x^*, f(x) \rangle > 0$ (remind that f, and so $\langle x - x^*, f(x) \rangle$, are continuous and $V_\infty/2 \leq |x - x^*|^2 \leq 3V_\infty$ is a compact set). So on the set $\{V_\infty > 0\}$ we should have $\sum_{n \geq 1} \gamma_n \langle X_n - x^*, f(X_n) \rangle = +\infty$, but we know that this sum is almost surely finite. So we have proved that $\mathbb{P}(V_\infty > 0) = 0$, which prove the almost sure convergence of the algorithm. □

2.2.2 Almost sure convergence for the Kiefer-Wolfowitz algorithm

The Kiefer-Wolfowitz algorithm is a variant of the Robbins-Monro algorithm. Its convergence can be proved using the Robbins-Siegmund lemma.

Theorem 2.2.2. *Let ϕ be a function from $\mathbb{R}$ to $\mathbb{R}$, such that*

$$\phi(x) = \mathbb{E}\left(F(x, U)\right),$$

where U is a random variable taking its values in $\mathbb{R}^p$ and F is a function from $\mathbb{R} \times \mathbb{R}^p$ to $\mathbb{R}$.

We assume that

- *ϕ is a C^2 strictly convex function such that*

  $$|\phi''(x)| \leq K(1 + |x|),$$

  and there exist x^* which minimize ϕ on $\mathbb{R}$.

- *$(\gamma_n, n \geq 1)$ and $(c_n, n \geq 1)$ are such 2 positive, decreasing sequence of real numbers such that*

  $$\sum_{n\geq 1} \gamma_n = +\infty, \sum_{n\geq 1} \gamma_n c_n < +\infty, \sum_{n\geq 1} \frac{\gamma_n^2}{c_n^2} < +\infty,$$

- *$s^2(x) = \mathbb{E}\left(F^2(x, U)\right) \leq K(1 + |x|)$.*

- *$(U_n^1, n \geq 1)$ and $(U_n^1 2, n \geq 1)$ are 2 independent sequences of independent random variables following the of U.*

We define $(X_n, n \geq 0)$ by $X_0 = x_0 \in \mathbb{R}$ and, inductively

$$X_{n+1} = X_n - \gamma_n \frac{F(X_n + c_n, U_{n+1}^1) - f(X_n - c_n, U_{n+1}^2)}{2c_n}.$$

Then

$$\lim_{n\to+\infty} X_n = x^*, \textit{a.s.}$$

Proof. First note that because of the assumptions on ϕ, for $|c| \leq 1$,

$$|\phi(x + c) - \phi(x - c) - 2c\phi'(x + c)| \leq c^2 K\left(1 + |x - x^*|\right). \tag{2.4}$$

Define $V_n = |X_n - x^*|^2$, clearly

$$V_{n+1} = V_n + |X_{n+1} - X_n|^2 + 2\left(X_n - x^*\right)\left(X_{n+1} - X_n\right).$$

Moreover

$$|X_{n+1} - X_n|^2 \leq \frac{2\gamma_n^2}{4c_n^2}\left(\left|f(X_n + c_n, U_{n+1}^1\right|^2 + \left|f(X_n - c_n, U_{n+1}^2\right|^2.\right)$$

So :

$$\mathbb{E}\left(|X_{n+1} - X_n|^2 |\mathcal{F}_n\right) \leq \frac{\gamma_n^2}{2c_n^2}\left(s^2(X_n + c_n) + s^2(X_n - c_n)\right),$$

and

$$\begin{aligned}\mathbb{E}(V_{n+1}|\mathcal{F}_n) \leq V_n & \\ \mathbf{A1} \quad & + \frac{\gamma_n^2}{2c_n^2}\left(s^2(X_n+c_n)+s^2(X_n-c_n)\right) \\ \mathbf{A2} \quad & - \frac{\gamma_n}{c_n}(X_n-x^*)\left[\phi(X_n+c_n)-\phi(X_n-c_n)-2c_n\phi'(X_n)\right] \\ \mathbf{A3} \quad & - \gamma_n(X_n-x^*)\phi'(X_n).\end{aligned} \tag{2.5}$$

Now, assuming that n is large enough to have $c_n \leq 1$,

$$\begin{aligned}\mathbf{A1} &\leq \frac{\gamma_n^2}{c_n^2}K\left(1+|X_n+c_n|^2+|X_n-c_n|^2\right) \\ &\leq \frac{\gamma_n^2}{c_n^2}K'\left(1+|X_n|^2\right) \leq \frac{\gamma_n^2}{c_n^2}K''\left(1+|X_n-x^*|^2\right) = \frac{\gamma_n^2}{c_n^2}K''(1+V_n),\end{aligned} \tag{2.6}$$

Note that $K < K' < K''$ but that, as usual in this kind of proof, K'' is still denoted by K [1]. Moreover using (2.4), we obtain (using also $|x| \leq (1+x^2)/2$)

$$\mathbf{A2} \leq 2\,|X_n-x^*|\,\frac{\gamma_n}{2c_n}Kc_n^2 = K\gamma_n c_n\,|X_n-x^*| \leq K\gamma_n c_n\left(1+|X_n-x^*|^2\right) = K\gamma_n c_n(1+V_n).$$

Finally we obtain

$$\mathbb{E}(V_{n+1}|\mathcal{F}_n) \leq V_n(1+K\frac{\gamma_n^2}{c_n^2}+K\gamma_n c_n)+K\frac{\gamma_n^2}{c_n^2}+K\gamma_n c_n-\gamma_n(X_n-x^*)\phi'(X_n).$$

Now we can using the Robbins-Siegmund lemma, setting

- $a_n = b_n = K\frac{\gamma_n^2}{c_n^2}+K\gamma_n c_n$
- $c_n = \gamma_n(X_n-x^*)\phi'(X_n)$, which is positive because of the convexity of ϕ.

So, using the fact that $\sum_{n\geq 1} a_n = \sum_{n\geq 1} b_n < +\infty$, we obtain V_n converge to V_∞ and $\sum_{n\geq 1} c_n < +\infty$. Now using the same argument as in the proof the convergence of the Robbins-Monro algorithm, we can conclude that $\mathbb{P}(V_\infty = 0) = 1$. This ends the proof of the convergence of the algorithm. □

2.3 Speed of convergence of stochastic algorithms

2.3.1 Introduction in a simplified context

Robbins-Monro type algorithms are well known to cause problems of speed of convergence. We will see that these algorithms can lead to central limit theorem (convergence in $C/\sqrt{n}$) but not for an arbitrary choice of γ_n, in some sense, γ_n has to be large enough to have an optimal rate of convergence.

[1] K denote a "constant" which can change from line to line (and even in the same line)!

It is easy to show this in a simplified framework. We assume that

$$F(x,u) = cx + u,$$

where $c > 0$ and that U follows a Gaussian law with mean 0 and variance 1. The standard Robbins-Monro algorithm can be written as

$$X_{n+1} = X_n - \gamma_n \left(cX_n + U_{n+1}\right).$$

with $\gamma_n = \alpha/(n+1)$. In this case $f(x) = cx$ and, using theorem 2.2.1, we can prove that X_n converge, almost surely, to 0 when n goes to $+\infty$.

To obtain more explicit computations we replace the discrete dynamic by a continuous one

$$dX_t = -\gamma_t \left(cX_t dt + \sigma dW_t\right), X_0 = x.$$

where $\gamma_t = \frac{\alpha}{t+1}$. Using a standard way to solve this equation, we compute

$$d\left(e^{c\int_0^t \gamma_s ds} X_t\right) = e^{c\int_0^t \gamma_s ds}\left[c\gamma_t X_t dt - c\gamma_t X_t - \gamma_t \sigma dW_t\right] = -e^{c\int_0^t \gamma_s ds}\gamma_t \sigma dW_t.$$

But

$$e^{c\int_0^t \gamma_s ds} = e^{c\alpha\int_0^t \frac{1}{s+1} ds} = (t+1)^{c\alpha}.$$

So, solving the previous equation, we get

$$X_t = \frac{X_0}{(t+1)^{c\alpha}} - \frac{\sigma\alpha}{(t+1)^{c\alpha}} \int_0^t \frac{1}{(s+1)^{1-c\alpha}} dW_s.$$

An easy computation leads to

$$\mathbb{E}(X_t^2) = \frac{|X_0|^2}{(t+1)^{2c\alpha}} + \frac{\sigma^2\alpha^2}{2c\alpha - 1}\left[\frac{1}{t+1} - \frac{1}{(t+1)^{2c\alpha}}\right].$$

We can now guess the asymptotic behavior of X_t

- if $2\alpha x > 1$, $\mathbb{E}(X_t^2)$ behave as $C/(t+1)$, so we can hope a central limit behavior for the algorithm,
- if $2\alpha x < 1$, $\mathbb{E}(X_t^2)$ behave as $C/(t+1)^{2c\alpha}$ which is worse than the awaited central limit behavior.

We can check on this example (exercise), that, when $2\alpha x > 1$ $\sqrt{t}X_t$ converge *in distribution* to a Gaussian random variable, but that, when $2\alpha x < 1$, $t^{c\alpha}X_t$ converge *almost surely* to a random variable.

We will see in what follows, that we can fully justify on the discrete algorithm : when α is large enough, a central limit theorem is true and when α is too small an asymptotic convergence worse that a central limit theorem occurs.

The result on which relies the proof is the central limit theorem for sequence of martingales (see below).

Practical considerations When using a Robbins-Monro style algorithm, in order to have a good behavior, γ_n has to be chosen large enough in order to have a central limit theorem, but not too large in order to minimize the variance of the algorithm. One way to do this is to choose $\gamma_n = \alpha/n$ with α large enough, another way is to choose $\gamma_n = \alpha/n^\beta$, with $1/2 < \beta < 1$. But this last choice increase the variance of the algorithm.

2.3.2 L^2 and locally L^2 martingales

Definition 2.3.1. *Let $(\mathcal{F}_n, n \geq 0)$ be a filtration on a probability space. An $\mathcal{F}$-martingale $(M_n, n \geq 0)$ is called a $\mathcal{F}$-square integrable martingale if, for all $n \geq 0$, $\mathbb{E}(M_n^2) < +\infty$.*

in this case, we are able to define a very useful object, the *bracket* of the martingale. We will see that the bracket of a martingale give a good indication of the asymptotic behavior of the martingale. This object will be useful to write the central limit theorem for martingales.

Definition 2.3.2. *Assume $(M_n, n \geq 0)$ is a square integrable martingale. There exists a unique, previsible, increasing process $(\langle M \rangle_n, n \geq 0)$, equal at 0 at time 0 such that*

$$M_n^2 - \langle M \rangle_n$$

is a martingale. Moreover $\langle M \rangle_n$ can be defined by $\langle M \rangle_0 = 0$ and

$$\langle M \rangle_{n+1} - \langle M \rangle_n = \mathbb{E}\left((M_{n+1} - M_n)^2 | \mathcal{F}_n\right) = \mathbb{E}\left(M_{n+1}^2 | \mathcal{F}_n\right) - M_n^2.$$

Proof. Previsible means here that $\langle M \rangle_n$ is $\mathcal{F}_{n-1}$-measurable. So, it is easy to check that if $\langle M \rangle_n$ is previsible and $M_n^2 - \langle M \rangle_n$ is a martingale

$$\langle M \rangle_{n+1} - \langle M \rangle_n = \mathbb{E}\left(M_{n+1}^2 | \mathcal{F}_n\right) - M_n^2,$$

which proves the unicity of $\langle M \rangle$ adding $\langle M \rangle_0 = 0$. Moreover, using the martingale property of M, on can check that

$$\mathbb{E}\left(M_{n+1}^2 | \mathcal{F}_n\right) - M_n^2 = \mathbb{E}\left((M_{n+1} - M_n)^2 | \mathcal{F}_n\right) \geq 0,$$

which proves that $\langle M \rangle_n$ is increasing. □

The following theorem relate the almost sure asymptotic behavior of a square integrable martingale M_n to its bracket.

Theorem 2.3.3 (Strong law of large number for martingales)**.** *Let $(M_n, n \geq 0)$ be a square integrable martingale and denote by $(\langle M \rangle_n, n \geq 0)$ its bracket, then*

- *on $\{\langle M \rangle_\infty := \lim_{n \to +\infty} \langle M \rangle_n < +\infty\}$, M_n converge almost surely to a random variable denoted as M_∞.*
- *on $\{\langle M \rangle_\infty = +\infty\}$,*

$$\lim_{n \to +\infty} \frac{M_n}{\langle M \rangle_n} = 0, a.s..$$

Moreover, as soon as $a(t)$ is a positive, increasing function such that $\int_0^{+\infty} \frac{dt}{1+a(t)} < +\infty$

$$\lim_{n \to +\infty} \frac{M_n}{\sqrt{a(\langle M \rangle_n)}} = 0, a.s..$$

Proof. Define τ_p as

$$\tau_p = \inf \{n \geq 0, \langle M \rangle_{n+1} \geq p\}.$$

τ_p is a stopping time as $\langle M \rangle$ is previsible. Note that, by definition, $\langle M \rangle_{\tau_p \wedge n} \leq \langle M \rangle_{\tau_p} \leq p$.

So $M_n^{(p)} = M_{n \wedge \tau_p}$ is also a martingale and $\langle M^{(p)} \rangle_n = \langle M \rangle_{n \wedge \tau_p}$ (since $M^2_{n \wedge \tau_p} - \langle M \rangle_{n \wedge \tau_p}$ is a martingale).

Now remark that, for all $n \geq 0$

$$\mathbb{E}\left((M_n^{(p)})^2\right) = \mathbb{E}(M_0^2) + \mathbb{E}\left(\langle M^{(p)} \rangle_n\right) \leq \mathbb{E}(M_0^2) + p.$$

So $(M_{n \wedge \tau_p}, n \geq 0)$ is a martingale bounded in L^2 so it converges when n goes to $+\infty$. So on the set $\{\tau_p = +\infty\}$, M_n itself converges to a random variable M_∞. As this is true for every p, we have proved that M_n converge to M_∞ on the set $\cup_{p \geq 0} \{\tau_p = +\infty\}$. But $\{\langle M \rangle_\infty < p\} = \{\tau_p = +\infty\}$, so

$$\{\langle M \rangle_\infty < +\infty\} = \cup_{p \geq 0} \{\langle M \rangle_\infty < p\} \subset \cup_{p \geq 0} \{\tau_p = +\infty\}.$$

So, the set $\{\langle M \rangle_\infty < +\infty\}$, M_n converge to M_∞, which proves the first point.

For the second one, we will consider new martingale defined by

$$N_n = \sum_{k=1}^{n} \frac{M_k - M_{k-1}}{\sqrt{1 + a(\langle M \rangle_k)}}.$$

$(N_n, n \geq 0)$ is a martingale as it is defined as martingale transform of the martingale M (recall that $\langle M \rangle_k$ is $\mathcal{F}_{k-1}$-measurable, which is exactly what is needed to prove that N is a martingale transform).

Moreover it is easy to check that this martingale is still a square integrable martingale and that we can compute its bracket because

$$\langle N \rangle_{n+1} - \langle N \rangle_n := \mathbb{E}\left(\frac{(M_{n+1} - M_n)^2}{1 + a(\langle M \rangle_{n+1})} \middle| \mathcal{F}_n \right) = \frac{\langle M \rangle_{n+1} - \langle M \rangle_n}{1 + a(\langle M \rangle_{n+1})}.$$

But :

$$\langle N \rangle_n = \sum_{k=1}^{n} \frac{\langle M \rangle_k - \langle M \rangle_{k-1}}{1 + a(\langle M \rangle_k)} \leq \sum_{k=1}^{n} \int_{\langle M \rangle_{k-1}}^{\langle M \rangle_k} \frac{dt}{1 + a(t)} \leq \int_0^{+\infty} \frac{dt}{1 + a(t)} < +\infty.$$

So $\langle N \rangle_\infty < +\infty$, a.s., and, using first part of this theorem, N_n converge a.s. to N_∞. But using Kronecker lemma, as we know that

$$N_\infty = \sum_{k=1}^{\infty} \frac{M_k - M_{k-1}}{\sqrt{1 + a(\langle M \rangle_k)}}$$

we conclude that $\lim_{n \to +\infty} \frac{M_n}{\sqrt{1 + a(\langle M \rangle_n)}} = 0$, and so that $\lim_{n \to +\infty} \frac{M_n}{\sqrt{a(\langle M \rangle_n)}} = 0$. The first case is obtained for $a(t) = t^2$. □

Application to the strong law of large numbers Assume that $(X_n, n \geq 1)$ is a sequence of independent random variables following the law of X, such that $\mathbb{E}(|X|^2) < +\infty$. Define $X'_n = X_n - \mathbb{E}(X)$, then

$$S_n = X_1 + \cdots + X_n - n\mathbb{E}(X) = X'_1 + \cdots + X'_n,$$

is a martingale with respect to $\sigma(X_1, \ldots, X_n)$. As, using independence, $\mathbb{E}\left((S_{n+1} - S_n)^2 | \mathcal{F}_n\right) = \mathbb{E}((X'_{n+1})^2 | \mathcal{F}_n) = \text{Var}\ (X)$, $\langle S \rangle_n = n\text{Var}\ (X)$. So $\langle S \rangle_\infty = \infty$ and using the previous theorem we get $\lim_{n\to+\infty} S_n/n = 0$, which give the strong law of large numbers.

Moreover using $a(t) = t^{1+\epsilon}$, with $\epsilon > 0$, we get

$$\lim_{n\to+\infty} \frac{1}{n^{\epsilon/2}} \sqrt{n} \left\{ \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X) \right\} = 0, a.s..$$

So we obtain a useful information on the speed of convergence of $\frac{X_1+\cdots+X_n}{n}$ to $\mathbb{E}(X)$.

Nevertheless, to obtain the central limit theorem a different tool is needed.

Extension to martingales locally in L^2 We can extend the definition of the bracket for a larger class of martingale : the martingale locally in L^2. We now give some definitions.

Definition 2.3.4. *A process $(M_n, n \geq 0)$ is a local martingale, if there exists a sequence of stopping times $(\tau_p, p \geq 0)$ such that*

- *τ_p increase in p and, a.s., goes to ∞ when p goes to ∞.*
- *$(M_{n\wedge\tau_p}, n \geq 0)$ is a martingale.*

Definition 2.3.5. *A process $(M_n, n \geq 0)$ is a locally square integrable martingale, if it exists sequence of stopping times $(\tau_p, p \geq 0)$ such that*

- *τ_p increase in p and, a.s., goes to ∞ when p goes to ∞.*
- *$(M_{n\wedge\tau_p}, n \geq 0)$ is a martingale bounded in L^2.*

Proposition 2.3.1. *If $(M_n, n \geq 0)$ is a locally square integrable martingale, there exist a unique $\mathcal{F}_{n-1}$-adapted increasing process $(\langle M \rangle_n, n \geq 0)$, equal at 0 at time 0, such that $(M_n^2 - \langle M \rangle_n, n \geq 0)$ is a local martingale.*

If τ_p, increasing with p and, a.s., going to ∞ when p goes to ∞, is such that $M_n^{\tau_p} = M_{\tau_p \wedge n}$ is a square integrable martingale then $\langle M^{\tau_p} \rangle_n = \langle M \rangle_{\tau_p \wedge n}$.

Proof. The proof of this proposition is left as an exercise. □

The strong law of large number remains true and unchange for martingale locally in L^2.

Theorem 2.3.6. *Let $(M_n, n \geq 0)$ be a locally square integrable martingale and denote by $(\langle M \rangle_n, n \geq 0)$ its bracket, then*

- *on $\{\langle M \rangle_\infty := \lim_{n\to+\infty} \langle M \rangle_n < +\infty\}$, M_n converge almost surely to a random variable denoted as M_∞.*

- *on* $\{\langle M\rangle_\infty = +\infty\}$, *as soon as* $a(t)$ *is a positive, increasing function such that* $\int_0^{+\infty} \frac{dt}{1+a(t)} < +\infty$

$$\lim_{n\to+\infty} \frac{M_n}{\sqrt{a(\langle M\rangle_n)}} = 0, a.s..$$

Proof. The proof is almost identical to the square integrable case. it is left as an exercise. □

Exercise 21. Let $(X_n, n \geq 1)$ be a sequence of independent real random variables following the law of X, such that $\mathbb{E}(|X|) < +\infty$. Denote by $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and assume that $(\lambda_n, n \geq 0)$ is an $\mathcal{F}_n$-adapted sequence of random variables. define M_n by

$$M_n = \sum_{k=0}^{n-1} \lambda_k X_{k+1}.$$

- Prove that M is bounded in L^2 if and only if $\sum_{k=0}^{+\infty} \mathbb{E}(\lambda_k^2) < +\infty$.
- Prove that M is a L^2-martingale if and only if, for all $k \geq 0\ \mathbb{E}(\lambda_k^2) < +\infty$.
- Prove that M is a locally L^2 martingale if and only if, for all $k \geq 0\ |\lambda_k| < +\infty$.
- Proves there exists martingales which are not locally in L^2.

2.3.3 Central limit theorem for martingales

We begin by stating the central limit theorem for martingales.

Theorem 2.3.7. *Let* $(M_n, n \geq 0)$ *be a locally in* L^2 *martingale and* $a(n)$ *be a sequence of strictly positive real numbers increasing to* $+\infty$. *Assume that*

Bracket condition : $\lim_{n\to+\infty} \frac{1}{a(n)}\langle M\rangle_n = \sigma^2$ *in probability.* (2.7)

Lindeberg condition : *for all* $\epsilon > 0$,

$$\lim_{n\to+\infty} \frac{1}{a(n)} \sum_{k=1}^{n} \mathbb{E}\left((\Delta M_{k+1})^2 \mathbf{1}_{\left\{|\Delta M_{k+1}| \geq \epsilon\sqrt{a(n)}\right\}} \middle| \mathcal{F}_k \right) = 0, \text{ in probability.} \quad (2.8)$$

Then :

$$\frac{M_n}{\sqrt{a(n)}} \text{ converge in distribution to } \sigma G,$$

where G *is a Gaussian random variable with mean* 0 *and variance* 1.

Remark 2.3.2. Roughly speaking in order to obtain a Central limit theorem for a martingale, we need to get, first, an asymptotic *deterministic* estimate for $\langle M\rangle_n \approx a(n)$ when n goes to infinity and then, to check the Lindeberg condition.

Exercise 22 gives a proof in a simple case in which the role of the martingale hypothese is clearer than in the proof which is given below. For a complete discussion on martingale convergence theorem, we refer to [Hall and Heyde(1980)].

Proof. This proof is a slightly adapted version of the one of [Major(2013)].

We will need an extension of Lebesgue theorem (also known as Lebesgue theorem) which say that if X_n converge in probability[2] to X and $|X_n| \leq \hat{X}$ with $\mathbb{E}(\hat{X}) < +\infty$, then $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ (see exercise 24 for a proof).

We denote by $M^{(k)}$ the locally in L^2 martingale $M_j^{(k)} = M_j/\sqrt{a(k)}$ and by $\langle M^{(k)}\rangle$ its bracket. Then we introduce for each $k \geq 0$ the stopping time

$$\tau_k = \inf\left\{j \geq 0, \langle M^{(k)}\rangle_{j+1} > 2\sigma^2\right\}, \tag{2.9}$$

The random variable τ_k is a stopping time with respect to the σ-algebras $\mathcal{F}_j$, $j \geq 0$, since the random variable $\langle M^{(k)}\rangle_{j+1}$ is $\mathcal{F}_j$ measurable. Moreover $\mathbb{P}(\lim_{k\to+\infty} \tau_k = +\infty) = 1$, because $\mathbb{P}(\tau_k > j) = \mathbb{P}(\langle M\rangle_{j+1} \leq 2\sigma^2 a(k))$ and $a(k)$ tends to $+\infty$ when k goes to $+\infty$. Now introduce, the stopped process $M^{[k]}$ defined by

$$M_j^{[k]} = M_{j\wedge\tau_k}^{(k)}.$$

We can check that $M^{[k]}$ is a martingale bounded in L^2 as

$$\langle M^{[k]}\rangle_j = \langle M^{(k)}\rangle_{j\wedge\tau_k} \leq 2\sigma^2, \tag{2.10}$$

so $\mathbb{E}\left(\langle M^{[k]}\rangle_j\right) \leq 2\sigma^2 < +\infty$ and $\mathbb{E}\left((M_j^{[k]})^2\right) \leq \mathbb{E}\left((M_0^{(k)})^2\right) + 2\sigma^2 < +\infty$. So $M^{[k]}$ is an L^2 martingale whose braket can be computed as

$$\Delta\langle M^{[k]}\rangle_j := \langle M^{[k]}\rangle_j - \langle M^{[k]}\rangle_{j-1} = \mathbb{E}\left((\Delta M_j^{[k]})^2 \middle| \mathcal{F}_{j-1}\right), \text{ where } \Delta M_j^{[k]} := M_j^{[k]} - M_{j-1}^{[k]}.$$

Note that we can rewrite the Lindeberg condition (2.8) as

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{(k)})^2 \mathbf{1}_{\left\{|\Delta M_j^{(k)}| \geq \epsilon\right\}} \middle| \mathcal{F}_{j-1}\right) = 0, \text{ in probability.}$$

Moreover $|\Delta M_j^{[k]}| = \mathbf{1}_{\{j<\tau_k\}}|\Delta M_j^{(k)}| \leq |\Delta M_j^{(k)}|$, so

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}| \geq \epsilon\right\}} \middle| \mathcal{F}_{j-1}\right) = 0, \text{ in probability.}$$

Now, as $\mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}| \geq \epsilon\right\}} \middle| \mathcal{F}_{j-1}\right) \leq \mathbb{E}\left((\Delta M_j^{[k]})^2 \middle| \mathcal{F}_{j-1}\right) = \Delta\langle M^{[k]}\rangle_j$, taking expectation in the previous convergence and justifying it by the (extended) Lebesgue theorem (as $\langle M^{[k]}\rangle_k \leq 2\sigma^2$) we obtain a stronger Lindeberg condition for $M^{[k]}$

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}| \geq \epsilon\right\}}\right) = 0. \tag{2.11}$$

Now note that $\langle M^{[k]}\rangle_k = \frac{1}{a(k)}\langle M\rangle_{k\wedge\tau_k}$ and, using hyphothesis (2.7), that

$$\lim_{k\to+\infty} \mathbb{P}(\tau_k > k) = \lim_{k\to+\infty} \mathbb{P}\left(\langle M\rangle_k \leq 2\sigma^2 a(k)\right) = 1.$$

[2] which is a weaker asumption than the almost surely convergence usually assumed.

So we have $\lim_{k\to+\infty}\langle M^{[k]}\rangle_k = \sigma^2$, in probability. Now taking expectation and using again Lebesgue theorem we get a stronger bracket condition for $M^{[k]}$

$$\lim_{k\to+\infty} \mathbb{E}\left(\langle M^{[k]}\rangle_k\right) = \sigma^2. \tag{2.12}$$

Now let $S_k = M_k^{(k)}$ and $\bar{S}_k = M_k^{[k]} = M_{k\wedge\tau_k}^{(k)}$ for $k \geq 1$. With these notation, we want to prove that S_k converge in law to to a gaussian random variable. But $\bar{S}_k - S_k$ converges in probability to 0 when $k \to \infty$ (since $\bar{S}_k = S_k$ if $\tau_k > k$ and we have already seen that $\lim_{k\to+\infty}\mathbb{P}(\tau_k > k) = 1$.). So, using Slutzky lemma, it remains to prove the convergence in probability of $\bar{S}_k$ to a gaussian random variable, i.e.

$$\lim_{k\to\infty} \mathbb{E}(e^{it\bar{S}_k}) = e^{-\sigma^2t^2/2} \quad \text{for all real numbers } t. \tag{2.13}$$

And we will show that relation (2.13) follows from

$$\lim_{k\to\infty} \mathbb{E}(e^{itS_k+t^2U_k/2}) = 1 \quad \text{for all real numbers } t, \tag{2.14}$$

where $U_k = \langle M^{[k]}\rangle_k$. Indeed, U_k converge in probability to σ^2 if $k \to \infty$, and $0 \leq U_k \leq 2\sigma^2$ for all $k \geq 1$ because of (2.10). Hence $e^{itS_k+t^2U_k/2} - e^{itS_k+\sigma^2t^2/2}$ converge in probability to 0 for all real numbers t if $k \to \infty$, and

$$\left|e^{itS_k+t^2U_k/2} - e^{itS_k+\sigma^2t^2/2}\right| \leq \left|e^{t^2U_k/2} - e^{\sigma^2t^2/2}\right| \leq e^{t^2\sigma^2}$$

Hence by (extended) Lebesgue's theorem $\lim_{k\to\infty}\mathbb{E}(e^{itS_k+t^2U_k/2} - e^{itS_k+\sigma^2t^2/2}) = 0$. Formula (2.13) follows from this statement if we can prove (2.14).

For this we first show that

$$\left|\mathbb{E}\left(e^{itS_k+t^2U_k/2}\right) - 1\right| \leq e^{\sigma^2t}\sum_{j=1}^{k}\mathbb{E}\left|e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(\, e^{it\Delta M_j^{[k]}}\,|\mathcal{F}_{j-1}\right) - 1\right|. \tag{2.15}$$

Indeed, let us introduce the random variables $S_{k,j} = M_j^{[k]}$, $U_{k,j} = \langle M^{[k]}\rangle_j$, for $j \geq 1$ and $S_{k,0} = 0$, $U_{k,0} = 0$ for all indices $k \geq 1$. Then we have $S_{k,k} = S_k$, $U_{k,k} = U_k$, and

$$\begin{aligned}\mathbb{E}\left(e^{itS_k+t^2U_k/2} - 1\right) &= \sum_{j=1}^{k}\mathbb{E}\left(e^{itS_{k,j}+t^2U_{k,j}/2} - e^{itS_{k,j-1}+t^2U_{k,j-1}/2}\right)\\ &= \sum_{j=1}^{k}\mathbb{E}e^{itS_{k,j-1}+t^2U_{k,j-1}/2}\mathbb{E}\left[\left(e^{it\Delta M_j^{[k]}+t^2\Delta\langle M^{[k]}\rangle_j/2} - 1\middle|\,\mathcal{F}_{j-1}\right)\right].\end{aligned}$$

Since $e^{itS_{k,j-1}+t^2U_{k,j-1}/2}$ is bounded by $e^{\sigma^2t^2}$, it follows from the above identity that

$$\left|\mathbb{E}\left(e^{itS_k+t^2U_k/2} - 1\right)\right| \leq e^{\sigma^2t}\sum_{j=1}^{k}\mathbb{E}\left|\mathbb{E}\left(e^{it\Delta M_j^{[k]}+t^2\Delta\langle M^{[k]}\rangle_j/2} - 1\middle|\,\mathcal{F}_{j-1}\right)\right|,$$

and as $\mathbb{E}\left(e^{it\Delta M_j^{[k]}+t^2\Delta\langle M^{[k]}\rangle_j/2} - 1\middle|\,\mathcal{F}_{j-1}\right) = e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\,\mathcal{F}_{j-1}\right) - 1$, this implies the estimate (2.15).

To prove formula (2.14) with the help of inequality (2.15) we have to give an estimate for $\mathbb{E}\left|e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\mathcal{F}_{j-1}\right)-1\right|$.

The expression $e^{t^2\Delta\langle M^{[k]}\rangle_j/2}$ can be written in the form $e^{t^2\Delta\langle M^{[k]}\rangle_j/2}=1+\frac{t^2\Delta\langle M^{[k]}\rangle_j}{2}+\eta_{k,j}^{(1)}$ with an appropriate random variable $\eta_{k,j}^{(1)}$ which satisfies the inequality $|\eta_{k,j}^{(1)}|\le K_1(t)\Delta\langle M^{[k]}\rangle_j^2$ with some number $K_1(t)$ depending only on the parameter t, because $\langle M^{[k]}\rangle_j\le 2\sigma^2$ by formula (2.10). We can estimate the expression

$$\eta_{k,j}^{(2)}=\mathbb{E}\left(e^{it\Delta M_j^{[k]}}-1+\frac{t^2(\Delta M_j^{[k]})^2}{2}\middle|\mathcal{F}_{j-1}\right)$$

in a similar way. To do this let us fix a small number $\epsilon>0$, and show that the inequality

$$\left|e^{it\Delta M_j^{[k]}}-1-it\Delta M_j^{[k]}+\frac{t^2(\Delta M_j^{[k]})^2}{2}\right|\le\alpha(\Delta M_j^{[k]}),$$

holds with $\alpha(x)=t^2x^2\mathbf{1}_{\{|x|>\epsilon\}}+\frac{\epsilon}{6}|t|^3x^2\mathbf{1}_{\{|x|\le\epsilon\}}$. Indeed, we get this estimate by bounding the expression $\left|e^{itx}-1-itx+\frac{t^2x^2}{2}\right|$ by t^2x^2 if $|x|>\epsilon$ and by $\frac{|t|^3|x|^3}{6}\le\epsilon\frac{|t|^3x^2}{6}$ if $|x|\le\epsilon$. Using that $\mathbb{E}(\Delta M_j^{[k]}|\mathcal{F}_{j-1})=0$ and taking the conditional expectation in the last inequality with respect to $\mathcal{F}_{j-1}$ we get

$$\begin{aligned}|\eta_{k,j}^{(2)}|&\le\mathbb{E}\left(\left|e^{it\Delta M_j^{[k]}}-1-it\Delta M_j^{[k]}+\frac{t^2(\Delta M_j^{[k]})^2}{2}\right|\middle|\mathcal{F}_{j-1}\right)\\&\le\mathbb{E}\left(\alpha(\Delta M_j^{[k]})\middle|\mathcal{F}_{k,j}\right)\le t^2\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)+\frac{\epsilon}{6}|t|^3\Delta\langle M^{[k]}\rangle_j.\end{aligned}$$

Since $\langle M^{[k]}\rangle_j\le 2\sigma^2$, both $\eta_{k,j}^{(1)}$ and $\eta_{k,j}^{(2)}$ are bounded random variables (with a bound depending only on the parameter t), and the above estimates imply that

$$\begin{aligned}&\left|e^{t^2\Delta\langle M^{[k]}\rangle j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\mathcal{F}_{j-1}\right)-1\right|\\&\quad=\left|\left(1+\frac{t^2\Delta\langle M^{[k]}\rangle_j}{2}+\eta_{k,j}^{(1)}\right)\left(1-\frac{t^2\Delta\langle M^{[k]}\rangle_j}{2}+\eta_{k,j}^{(2)}\right)-1\right|\\&\quad\le t^4(\Delta\langle M^{[k]}\rangle_j)^2+K_3(t)\left(|\eta_{k,j}^{(1)}|+|\eta_{k,j}^{(2)}|\right)\\&\quad\le K_4(t)\left((\Delta\langle M^{[k]}\rangle_j)^2+\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)+\epsilon\Delta\langle M^{[k]}\rangle_j\right).\end{aligned}$$

Let us take the expectation of the left-hand side and right-hand side expression in the last inequality and sum up for all indices $j\ge 1$. The inequality obtained in such a way together with formula (2.15) imply that

$$\begin{aligned}|\mathbb{E}e^{itS_k+t^2U_k/2}-1|\le K_5(t)\Bigg(&\sum_{j=1}^k\mathbb{E}\left((\Delta\langle M^{[k]}\rangle_j)^2\right)\\&+\sum_{j=1}^k\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\right)+\epsilon\sum_{j=1}^k\mathbb{E}\{\Delta\langle M^{[k]}\rangle_j\}\Bigg).\end{aligned}\tag{2.16}$$

To estimate the first sum at the right-hand side of (2.16) let us make the following estimate:

$$\begin{aligned}
\mathbb{E}\left\{(\Delta\langle M^{[k]}\rangle_j)^2\right\} &= \mathbb{E}\left\{\left[\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)+\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\leq\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)\right]^2\right\}\\
&\leq 2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)^2\right\}+2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\leq\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)^2\right\}\\
&\leq 2\mathbb{E}\left\{\Delta\langle M^{[k]}\rangle_j\mathbb{E}\left(\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)\right\}+2\epsilon^2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\leq\epsilon\right\}}\middle|\mathcal{F}_{j-1}\right)\right\}\\
&\leq 4\sigma^2\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\right)+2\epsilon^2\mathbb{E}\left(\Delta\langle M^{[k]}\rangle_j\right).
\end{aligned}$$

Using this estimate and (2.16), we obtain

$$\left|\mathbb{E}e^{itS_k+t^2U_k/2}-1\right|\leq K_6(t)\left(\sum_{j=1}^k\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\epsilon\right\}}\right)+\epsilon\sum_{j=1}^k\mathbb{E}\left(\Delta\langle M^{[k]}\rangle_j\right)\right),$$

and relations (2.11), (2.12) imply formula (2.14). Thus we have proved the central limit theorem. □

Application to the standard case It is easy to recover the traditional central limit theorem using the previous corollary. For this, consider a sequence of independent random variables following the law of X such that $\mathbb{E}(|X|^2)<+\infty$. Let $a(n)=n$ and $M_n=X_1+\cdots+X_n-n\mathbb{E}(X)$. M is a martingale and its bracket $\langle M\rangle_n=n\mathrm{Var}\,(X)$ (so obviously $\langle M\rangle_n/n$ converge to $\mathrm{Var}\,(X)=\sigma^2$!). It remains to check the Lindeberg condition. But

$$\frac{1}{n}\sum_{k=1}^n\mathbb{E}\left(X_{k+1}^2\mathbf{1}_{\left\{|X_{k+1}|\geq\epsilon\sqrt{n}\right\}}\middle|\mathcal{F}_k\right)=\mathbb{E}\left(X^2\mathbf{1}_{\left\{|X|\geq\epsilon\sqrt{n}\right\}}\right).$$

But integrability of X^2 and Lebesgue theorem proves that $\mathbb{E}\left(X^2\mathbf{1}_{\left\{|X|\geq\epsilon\sqrt{n}\right\}}\right)$ converge to 0 when n goes to ∞.

2.3.4 A central limit theorem for the Robbins-Monro algorithm

We can now derive a central limit theorem for the Robbins-Monro algorithm. We will only deal with the uni-dimensional case. This part follows closely [Duflo(1997)] (or [Duflo(1990)] in french).

Theorem 2.3.8. *Let $F(x,u)$ be a function from $\mathbb{R}\times\mathbb{R}^p$ to $\mathbb{R}$ and U is a random variable taking its values in $\mathbb{R}^d$. We assume that*

- *$f(x)=\mathbb{E}\left(F(x,U)\right)$ is C^2.*
- *$f(x^*)=0$ and $\langle f(x),x-x^*\rangle>0$, for $x\neq x^*$.*
- *$f'(x^*)=c$, where $c>0$.*

- *If* $\sigma^2(x) = \text{Var } (F(x,U)^2)$, $s^2(x) = \sigma^2(x) + f^2(x) \leq K\,(1+|x|^2)$.
- *It exists* $\eta > 0$ *such that, for all* $x \in \mathbb{R}$

$$v^{2+\eta}(x) := \mathbb{E}\left(|F(x,U)|^{2+\eta}\right) < +\infty,$$

and $\sup_{n\geq 0} v^{2+\eta}(X_n) < +\infty$.

We consider a sequence $(U_n, n \geq 1)$ *of independent random variables following the law of* U *and* $\gamma = \frac{\alpha}{n}$. *We define* X_n *by*

$$X_{n+1} = X_n - \gamma_n F(X_n, U_{n+1}), X_o = x \in \mathbb{R}.$$

Then X_n *converge almost surely to* x^* *and*

- *if* $c\alpha > 1/2$, $\sqrt{n}(X_n - x^*)$ *converge* in distribution *to a zero mean Gaussian random variable with variance* $\sigma^2 = \alpha^2\sigma^2(x^*)/(2c\alpha - 1)$
- *if* $c\alpha < 1/2$, $n^{c\alpha}(X_n - x^*)$ *converge* almost surely *to a random variable.*

Remark 2.3.3. It is easy to optimize in α the asymptotic variance $\alpha^2\sigma^2(x^*)/(2c\alpha - 1)$ and to prove that the optimal choice is given by $\alpha = 1/c$.

The same type of TCL can be obtained when $\gamma_n = \alpha/n^\beta$, with $1/2 < \beta < 1$. In this case it can be proved that, for every α, $(X_n - x^*)/\sqrt{\gamma_n}$ converge in distribution to a gaussian random variable, wathever the value of α.

Proof. We begin the proof with a simple deterministic lemma.

Lemma 2.3.4. *Let* c, α *be strictly positive numbers. Assume that* $\gamma_n = \alpha/n$.

Define $\alpha_n = 1 - c\gamma_n$, $\beta_n = \alpha_1 \ldots \alpha_n$, *then there exists a strictly positive number* C *such that*

$$\lim_{n\to+\infty} \beta_n n^{c\alpha} = C \text{ and } \lim_{n\to+\infty} (\gamma_n/\beta_n) n^{1-c\alpha} = \frac{\alpha}{C}.$$

Proof. Using Taylor formula, we get, for $0 < x < 1$

$$0 \leq \log(1-x) + x \leq \frac{x^2}{2}\frac{1}{(1-x)^2}.$$

So if $n \geq n_0$, such that $c\gamma_n < 1/2$, we have

$$0 \leq \log(1 - c\gamma_n) + c\gamma_n \leq 2c^2\gamma_n^2.$$

From this we deduce that, if $s_n = \sum_{k=1}^n \gamma_k$,

$$\lim_{n\to+\infty} \beta_n e^{cs_n} = C_1 > 0.$$

But $\lim_{n\to+\infty} \sum_{k=1}^n 1/k - \log(n) = C_2$ where C_2 is the Euler constant. So we have

$$\lim_{n\to+\infty} \beta_n n^{\alpha c} = C_3 = C_1 e^{\alpha C_2}.$$

and from this we deduce that $\lim_{n\to+\infty}(\gamma_n/\beta_n)n^{1-c\alpha} = \frac{\alpha}{C_3}$. □

In what follows we will assume that $x^* = 0$. Our algorithm writes as

$$\begin{aligned} X_{n+1} &= X_n - \gamma_n F(X_n, U_{n+1}) \\ &= X_n(1 - c\gamma_n) - \gamma_n \left[F(X_n, U_{n+1}) - f(X_n)\right] - \gamma_n \left[f(X_n) - cX_n\right]. \end{aligned}$$

Now, denote

- $\Delta M_{n+1} = F(X_n, U_{n+1}) - f(X_n)$ (ΔM_{n+1} is a martingale increment),
- $R_n = f(X_n) - cX_n$ (R_n will be "small"),
- $\alpha_n = 1 - c\gamma_n$ and $\beta_n = \alpha_1 \dots \alpha_n$.

With these notations

$$\frac{X_{n+1}}{\beta_n} = \frac{X_n}{\beta_{n-1}} - \frac{\gamma_n}{\beta_n}\Delta M_{n+1} - \frac{\gamma_n}{\beta_n} R_n,$$

so

$$X_n = X_0\beta_{n-1} - \beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}\Delta M_{k+1} - \beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k} R_k.$$

The main point is now to estimate the bracket of N_n the martingale defined by

$$N_n = \sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}\Delta M_{k+1}.$$

From this we will deduce the asymptotic behavior of X_n

Clearly

$$\langle N\rangle_n = \sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_n).$$

But we already know that X_n converge to $x^* = 0$, so, by continuity of σ, $\lim_{n\to+\infty}\sigma^2(X_n) = \sigma^2(x^*) > 0$.

The case where $2c\alpha > 1$ In this case we are interested in the convergence in distribution of $\sqrt{n}(X_n - x^*)$ to a gaussian random variable. The main step will be to apply the central limit theorem for the martingale N. So we have to get an asymptotic estimate to its bracket $\langle N\rangle_n$.

For this, taking into account that (see previous lemma) $\frac{\gamma_k}{\beta_k} \approx \frac{1}{C}k^{c\alpha-1}$, we can prove that, for $2c\alpha > 1$, $\lim_{n\to+\infty}\langle N\rangle_n = +\infty$, and it is easy to check that

$$\lim_{n\to+\infty}\frac{\sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_n)}{\sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2}\sigma^2(x^*)} = 1, \text{a.s.}.$$

A more precise analysis (left as an exercise) shows that

$$n^{1-2c\alpha}\sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2} = (1+\epsilon_n)\frac{\alpha^2}{C^2(2c\alpha-1)},$$

and leads to

$$\lim_{n\to+\infty} \frac{\langle N\rangle_n}{n^{2c\alpha-1}} = \frac{\alpha^2\sigma^2(x^*)}{C^2(2c\alpha-1)}, \text{a.s..}$$

So we have the bracket condition needed to apply the central limit theorem for martingale N with $a(n) = n^{2c\alpha-1}$. It remains to check the Lindeberg condition. For this note that

$$\mathbb{E}\left(|X|^2\mathbf{1}_{\{|X|\geq\delta\}}\middle|\,\mathcal{B}\right) \leq \frac{\mathbb{E}\left(|X|^{2+\eta}\middle|\,\mathcal{B}\right)}{\delta^2}.$$

So we have

$$\begin{aligned}
&\frac{1}{a(n)}\sum_{k=1}^{n-1}\mathbb{E}\left(|\Delta N_{k+1}|^2\mathbf{1}_{\left\{|\Delta N_{k+1}|\geq\epsilon\sqrt{a(n)}\right\}}\middle|\,\mathcal{F}_k\right)\\
&\qquad\leq \frac{1}{a(n)}\sum_{k=1}^{n-1}\frac{1}{(\epsilon\sqrt{a(n)})^\eta}\mathbb{E}\left(|\Delta N_{k+1}|^{2+\eta}\middle|\,\mathcal{F}_k\right)\\
&\qquad= \frac{1}{a(n)^{1+\eta/2}}\frac{1}{\epsilon^\eta}\sum_{k=1}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^2\mathbb{E}\left(|F(X_k,U_{k+1})-f(X_k)|^{2+\eta}\middle|\,\mathcal{F}_k\right)\\
&\qquad\leq \frac{L(\omega)}{\epsilon^\eta}\frac{1}{a(n)^{1+\eta/2}}\sum_{k=1}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^2,
\end{aligned}$$

where $L(\omega) = \sup_{n\geq 0} v^{2+\eta}(X_n)$ (which is supposed to be a.s. finite by hypothesis). So we get

$$\begin{aligned}
&\frac{1}{a(n)}\sum_{k=1}^{n-1}\mathbb{E}\left(|\Delta N_{k+1}|^2\mathbf{1}_{\left\{|\Delta N_{k+1}|\geq\epsilon\sqrt{a(n)}\right\}}\middle|\,\mathcal{F}_k\right)\\
&\qquad\leq \frac{L(\omega)}{\epsilon^\eta}\frac{1}{a(n)^{1+\eta/2}}\sum_{k=1}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^2\\
&\qquad\approx Kn^{-(2c\alpha-1)(1+\eta/2)n^{(c\alpha-1)(2+\eta)+1}} = Kn^{-\eta/2}.
\end{aligned}$$

And this ends the proof that Lindeberg condition is fulfilled. So $N_n/\sqrt{a(n)}$ converge in distribution to a gaussian random variable with variance $\frac{\alpha^2\sigma^2(x^*)}{C^2(2c\alpha-1)}$, and so $\sqrt{n}\beta_n N_n$ converge to a gaussian random variable with variance $\frac{\alpha^2\sigma^2(x^*)}{(2c\alpha-1)}$.

To conclude (using Slutzky lemma) that $\sqrt{n}X_n$ converge in distribution to the same random variable we need to prove that $\sqrt{n}\beta_{n-1}X_0$ converge to 0 in probability (this is immediate) and that $\epsilon_n = \sqrt{n}\beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k$ also converge to 0 in probability. This part is heavily technical and will not be proved here [3].

[3]Note, though, that, $|R_n| \leq C|X_n|^2$, so if we can prove that $\mathbb{E}(|X_n|^2) \leq K/n$, we have

$$\sqrt{n}\beta_{n-1}\sum_{k=0}^{n-1}\mathbb{E}\left(\left|\frac{\gamma_k}{\beta_k}R_k\right|\right) \leq \sqrt{n}\beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{k\beta_k} \leq K/\sqrt{n}.$$

So, ϵ_n converge in L^1 (and so in probability) to 0.

The case where $2c\alpha < 1$ In this case, we will prove that $\frac{X_n - x^*}{\beta_n}$ converge almost surely to a random variable.

For this, we first check that

$$\langle N \rangle_n \approx \sigma^2(x^*) \sum_{k=0}^{n-1} \frac{\gamma_k^2}{\beta_k^2} \approx \sigma^2(x^*)\alpha^2 K \sum_{k=0}^{n-1} k^{2c\alpha-2} < +\infty.$$

So, using the strong law for martingale, N_n converge a.s. to N_∞. If remains to check that $\sum_{k=0}^{n-1} \frac{\gamma_k}{\beta_k} R_k$ converge a.s. to obtain the result of the theorem.

But, as $|R_k| \leq C|X_k|^2$, we have

$$\mathbb{E}\left(\left|\sum_{k=0}^{n-1} \frac{\gamma_k}{\beta_k} R_k\right|\right) \leq C\mathbb{E}\left(\sum_{k=0}^{n-1} \frac{\gamma_k}{\beta_k}|X_k|^2\right) \leq C\sum_{k=0}^{n-1} \frac{\gamma_k}{\beta_k}\mathbb{E}(|X_k|^2) \approx C\sum_{k=0}^{n-1} \frac{\mathbb{E}(|X_k|^2)}{k^{1-c\alpha}}.$$

So if we can prove there exist $\beta > c\alpha$ such that $\mathbb{E}(|X_n|^2) \leq \frac{C}{n^\beta}$, we will be able to deduce the absolute convergence of $\sum_{k=0}^{n-1} \frac{\gamma_k}{\beta_k} R_k$. □

Exercise 22. In this exercise, we prove the central limit theroem for martingale in a special case.

Let $(M_n, n \geq 0)$ be a martingale such that $\sup_{n\geq 0} |\Delta M_n| \leq K < +\infty$, where K is a constant. M is a square integrable martingale and, so, we can denote by $\langle M \rangle$ its bracket. Assume moreover that

$$\lim_{n\to+\infty} \frac{\langle M \rangle_n}{n} = \sigma^2, \text{a.s.} \tag{2.17}$$

where σ is a positive real number.

1. For λ real, let $\phi_j(\lambda) = \log \mathbb{E}\left(e^{\lambda \Delta M_j} \middle| \mathcal{F}_{j-1}\right)$, prove that

$$X_n = \exp\left(\lambda M_n - \sum_{j=1}^{n} \phi_j(\lambda)\right),$$

   is a martingale.

2. For $u \in \mathbb{R}$ and n large enough (to be able to define the complex logarithm) prove that

$$\exp\left(iu\frac{M_n}{\sqrt{n}} - \sum_{j=1}^{n} \phi_j(iu/\sqrt{n})\right)$$

   is a complex martingale and so

$$\mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} - \sum_{j=1}^{n} \phi_j(iu/\sqrt{n})\right)\right] = 1.$$

3. Using $|e^x - 1 - x - x^2/2| \leq |x|^3$, for $|x| \leq 1$ prove that, for n large enough

$$\left|\mathbb{E}\left(e^{i\frac{u}{\sqrt{n}}\Delta M_j} \middle| \mathcal{F}_{j-1}\right) - 1 + \frac{u^2}{2n}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathcal{F}_{j-1}\right)\right| \leq \frac{u^3}{n^{3/2}} K^3,$$

then that, using $|\log(1+x)-x| \leq |x|^2$, that for a c,

$$\left|\phi_j\left(\frac{iu}{\sqrt{n}}\right) - \frac{u^2}{2n}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathcal{F}_{j-1}\right)\right| \leq \frac{c}{n^{3/2}}$$

4. Prove, using 2.17
$$\lim_{n\to+\infty}\sum_{j=1}^{n}\phi_j\left(\frac{iu}{\sqrt{n}}\right) = -\frac{\sigma^2u^2}{2}, \text{a.s.}$$

5. Proves that
$$\lim_{n\to+\infty}\mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} - \sum_{j=1}^{n}\phi_j(iu/\sqrt{n})\right)\right] - \mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} + \frac{\sigma^2u^2}{2}\right)\right] = 0,$$
and deduce that $\lim_{n\to+\infty}\mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}}\right)\right] = \exp\left(\frac{\sigma^2u^2}{2}\right)$.

6. Conclude that $\frac{M_n}{\sqrt{n}}$ converge in distribution to a gaussian random variable.

7. Generalize the result when
$$\lim_{n\to+\infty}\frac{\langle M\rangle_n}{a(n)} = \sigma^2, \text{a.s.} \tag{2.18}$$

Exercise 23. Assume that $(u_n, n \geq 0)$ and $(b_n, n \geq 0)$ are two sequence of positive real numbers, $c > 0$ such that, for all $n \geq 0$

$$u_{n+1} \leq u_n\left(1 - \frac{c}{n}\right) + b_n.$$

1. Let $\beta_n = 1/\left(\prod_{i=1}^{n-1}(1-\frac{c}{k})\right)$. Prove that
$$u_n \leq \frac{K}{\beta_n} + \frac{1}{\beta_n}\sum_{k=1}^{n-1}\beta_{k+1}b_k \leq \frac{K}{\beta_n} + \sum_{k=1}^{n-1}b_k$$

2. Assume that $b_k = \frac{C}{k^\alpha}$, with $\alpha > 1$, prove that
$$u_n \leq \frac{K}{n^c} + \frac{K}{n^{\alpha-1}} \leq \frac{K}{n^{\inf(c,\alpha-1)}}.$$

Exercise 24. We assume that X_n converge in probability to X and that $|X_n| \leq \hat{X}$ with $\mathbb{E}\left(\hat{X}\right) < +\infty$. We want to prove that $\lim_{n\to+\infty}\mathbb{E}(X_n) = \mathbb{E}(X)$.

1. Let K be a positive real number and define $\phi_K(x)$ by $\phi_K(x) = (-K)\mathbf{1}_{\{x<-K\}} + x\mathbf{1}_{\{|x|\leq K\}} + K\mathbf{1}_{\{K<x\}}$. Prove that $|\phi_K(x) - \phi_K(y)| \leq |x-y|$.

2. Prove that
$$\mathbb{E}\left(|X_n - X|\right) \leq \mathbb{E}\left(|\phi_K(X_n) - \phi_K(X)|\right) + 2\mathbb{E}(\hat{X}\mathbf{1}_{\left\{\hat{X}\geq K\right\}}).$$

3. Prove that $\lim_{n\to\infty} \mathbb{E}(\hat{X}\mathbf{1}_{\left\{\hat{X}\geq K\right\}}) = 0$.

4. Prove for a given K that,

$$\mathbb{E}\left(|\phi_K(X_n) - \phi_K(X)|\right) \leq 2K\mathbb{P}\left(|X_n - X| \geq \epsilon\right) + \epsilon,$$

and deduce the extended Lebesgue theorem.

Exercise 25. We assume that X_n converge in distribution to X and that f is a continuous function such that $\mathbb{E}(|f(X_n)|) < +\infty$, for all $n \geq 1$ and satisfies a property of *equi-integrability*

$$\lim_{A\to+\infty} \mathbb{E}\left(|f(X_n)|\mathbf{1}_{\{|f(X_n)|>A\}}\right) = 0$$

We want to prove that $\lim_{n\to+\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$.

1. Prove that $\sup_{n\geq 1} \mathbb{E}(|f(X_n)|) < +\infty$ and, considering the continuous function $f(x) \wedge K$, that $\mathbb{E}(|f(X)|) \leq \sup_{n\geq 1} \mathbb{E}(|f(X_n)|)$.

2. Prove that there exists a family of continuous functions $\phi_\delta^A(x)$, $x \in \mathbb{R}$ such that, $\phi_\delta^A(x) \leq \mathbf{1}_{\{x>A\}}$ and for every $x \in \mathbb{R}$, $\phi_\delta^A(x)$ converge to $\mathbf{1}_{\{x>A\}}$.

3. Prove that

$$\begin{aligned}|\mathbb{E}(f(X_n) - \mathbb{E}(f(X))| \leq \\ \left|\mathbb{E}\left(f(X_n)\left[1-\phi_\delta^A(X_n)\right]\right) - \mathbb{E}\left(f(X)\left[1-\phi_\delta^A(X)\right]\right)\right| \\ + \mathbb{E}\left(|f(X_n)|\,\mathbf{1}_{\{|f(X_n)|>A\}}\right) + \mathbb{E}\left(|f(X)|\,\mathbf{1}_{\{|f(X)|>A\}}\right),\end{aligned}$$

and conclude.

4. Find an alternative way to recover the result of the exercise 24.

Problem 26. Une méthode de Monte-Carlo adaptative

On considère une fonction f, mesurable et bornée, de $\mathbb{R}^p$ dans $\mathbb{R}$ et X une variable aléatoire à valeur dans $\mathbb{R}^p$.

On s'intéresse à un cas où l'on sait représenter $\mathbb{E}\left(f(X)\right)$ sous la forme

$$\mathbb{E}\left(f(X)\right) = \mathbb{E}\left(H(f;\lambda,U)\right), \tag{2.19}$$

où $\lambda \in \mathbb{R}^n$, U est une variable aléatoire à valeur dans $[0,1]^d$ suivant une loi uniforme et où, pour tout $\lambda \in \mathbb{R}^n$, $H(f;\lambda,U)$ est une variable aléatoire de carré intégrable qui prend des valeurs réelles. La question 2 montrer que cela est généralement possible.

Le but de ce problème est de montrer que l'on peut, dans ce cas, faire varier λ au cours des tirages tout en conservant les propriétés de convergence d'un algorithme de type Monte-Carlo.

1. On note ϕ la fonction de répartition d'une gaussienne centrée réduite et ϕ^{-1} son inverse. On suppose que X est une variable aléatoire réelle de loi gaussienne centrée réduite.

   Montrer que, si $\lambda \in \mathbb{R}$, H définie par :

$$H(f;\lambda,U) = e^{-\lambda G - \frac{\lambda^2}{2}} f\left(G+\lambda\right), \text{ où } G = \phi^{-1}(U)$$

   permet de satisfaire la condition (2.19).

2. Soit X une variable aléatoire à valeur dans $\mathbb{R}^p$, de loi arbitraire, que l'on peut obtenir par une méthode de simulation: cela signifie qu'il existe une fonction ψ de $[0,1]^d$ dans $\mathbb{R}^p$ telle que, si $U = (U_1, \ldots, U_d)$ suit une loi uniforme sur $[0,1]^d$, la loi de $\psi(U)$ est identique à celle de X.

   On pose, pour $i = 1, \ldots, d$, $G_i = \phi^{-1}(U_i)$ et l'on considère $\lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$. Proposer à partir de $(G_1 + \lambda_1, \ldots, G_d + \lambda_d)$ un choix de variable aléatoire $H(f; \lambda, U)$ qui permet de satisfaire l'égalité (2.19).

3. On considère une suite $(U^n, n \geq 1)$ de variables aléatoires indépendantes uniformément distribuées sur $[0,1]^d$. On note $\mathcal{F}_n = \sigma\left(U^k, 1 \leq k \leq n\right)$.

   Soit λ un vecteur fixé, comment peut on utiliser $H(f; \lambda, U)$ pour estimer $\mathbb{E}\left(f(X)\right)$? Comment peut-on estimer l'erreur commise sur cette estimation ?

   Quel est le critère pertinent pour choisir λ ?

On suppose que λ n'est plus constant mais évolue au fils du temps et est donné par une suite $(\lambda_n, n \geq 0)$ de variables aléatoires $\mathcal{F}_n$-mesurable (λ_0 est supposée constante). On suppose que

$$\begin{array}{l} \text{pour tout } \lambda \in \mathbb{R}^p,\, s^2(\lambda) = \mathrm{Var}\left(H(f; \lambda, U)\right) < +\infty, \\ \text{et } s^2(\lambda) \text{ est une fonction continue de } \mathbb{R}^d \text{ dans } \mathbb{R}. \end{array} \tag{2.20}$$

4. On pose

$$M_n = \sum_{i=0}^{n-1} \left[H(f; \lambda_i, U_{i+1}) - \mathbb{E}(f(X))\right].$$

   Montrer que, si H est *uniformément bornée*[4], $(M_n, n \geq 0)$ est une $\mathcal{F}_n$-martingale dont le crochet $\langle M \rangle_n$ s'exprime sous la forme $\langle M \rangle_n = \sum_{i=0}^{n-1} s^2(\lambda_i)$.

5. Plus généralement montrer que, sous l'hypothèse (2.20), $(M_n, n \geq 0)$ est une martingale localement dans L^2 de crochet toujours donné par $\langle M \rangle_n = \sum_{i=0}^{n-1} s^2(\lambda_i)$ (on pourra utiliser la famille de temps d'arrêt $\tau_A = \inf\{n \geq 0, |\lambda_n| > A\}$ et vérifier que $(M_{t \wedge \tau_A}, n \geq 0)$ est une martingale bornée dans L^2).

6. En utilisant la loi forte de grand nombre pour les martingales localement dans L^2, montrer que si

$$\lim_{n \to +\infty} \frac{\sum_{i=0}^{n-1} s^2(\lambda_i)}{n} = c, \tag{2.21}$$

   où c est une constante strictement positive, alors

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} H(f; \lambda_i, U_{i+1}) = \mathbb{E}(f(X)).$$

   Interpréter ce résultat en terme de méthode de Monte-Carlo. Vérifier que, si λ_n converge presque sûrement vers λ^* tel que $s^2(\lambda^*) > 0$, le résultat de cette question est vrai.

7. Quelle hypothèse faudrait-t'il ajouter à (2.21) pour obtenir un théorème central limite dans la méthode précédente (ne pas chercher à la vérifier) ? Enoncer le résultat que l'on obtiendrait alors.

[4] i.e., il existe K réel positif tel que pour tout λ et u, $|H(f; \lambda, u)| \leq K < +\infty$

Bibliography

[Abramovitz and Stegun(1970)] M. Abramovitz and A. Stegun. *Handbook of Mathematical Functions*. Dover, 9th edition, 1970.

[Antonov and Saleev(1980)] I.A. Antonov and V.M. Saleev. An economic method of computing lp_τ-sequences. *USSR Comput. Maths. Math. Phys*, 19:252–256, 1980.

[Barraquand and Martineau(1995)] J. Barraquand and D. Martineau. Numerical valuation of high dimensional multivariate american securities. *J. of Finance and Quantitative Analysis*, 30:383–405, 1995.

[Broadie and Glassermann(1997a)] M. Broadie and P. Glassermann. Pricing american-style securities using simulation. *J.of Economic Dynamics and Control*, 21:1323–1352, 1997a.

[Broadie and Glassermann(1997b)] M. Broadie and P. Glassermann. A stochastic mesh method for pricing high-dimensional american options. *Working Paper*, Columbia University:1–37, 1997b.

[Clément et al.(2002)] Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for American option pricing. *Finance Stoch.*, 6 (4):449–471, 2002.

[Cochran(1953)] William G. Cochran. *Sampling techniques*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953.

[Cohort(2001)] P. Cohort. Monte–carlo methods for pricing american options. Documentation for the premia software, INRIA and ENPC, February 2001. see `http://cermics.enpc.fr/~premia`.

[Cohort(2004)] Pierre Cohort. Limit theorems for random normalized distortion. *Ann. Appl. Probab.*, 14(1):118–143, 2004.

[Devroye(1986)] Luc Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York, 1986.

[Duflo(1990)] Marie Duflo. *Méthodes récursives aléatoires*. Techniques Stochastiques. Masson, Paris, 1990.

[Duflo(1997)] Marie Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

[Faure(1981)] H. Faure. Discrépance de suites associées à un système de numération (en dimension 1). *Bull. Soc. Math. France*, 109:143–182, 1981.

[Faure(1982)] H. Faure. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arith.*, 41:337–351, 1982.

[Fox(1988)] B.L. Fox. Algorithm 659:implementating Sobol's quasirandom sequence generator. *ACM Transaction on Mathematical Software*, 14(1):88–100, 1988.

[Fox et al.(1992)] B.L. Fox, P. Bratley, and H. Neiderreiter. Implementation and test of low discrepancy sequences. *ACM Trans. Model. Comput. Simul.*, 2(3):195–213, July 1992.

[Fu et al.(2001)] M.C. Fu, S.B. Laprise, D.B. Madan, Y. Su, and R. Wu. Pricing american options: a comparison of monte carlo simulation approaches. *Journal of Computational Finance*, 4(3), Spring 2001.

[Glasserman et al.(1999)] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path dependent opions. *Mathematical Finance*, 9(2):117–152, April 1999.

[Hall and Heyde(1980)] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

[Hammersley and Handscomb(1979)] J. Hammersley and D. Handscomb. *Monte Carlo Methods*. Chapman and Hall, London, 1979.

[Kalos and Whitlock(2008)] Malvin H. Kalos and Paula A. Whitlock. *Monte Carlo methods*. Wiley-Blackwell, Weinheim, second edition, 2008.

[Knuth(1998)] Donald E. Knuth. *The art of computer programming. Vol. 2*. Addison-Wesley, Reading, MA, 1998. Seminumerical algorithms, Third edition [of MR0286318].

[Ksas(2000)] Frédéric Ksas. *Méthodes de Monte-Carlo et suites à discrépance faible appliquées au calcul d'options en finance*. PhD thesis, Université d'Evry, Juillet 2000.

[L'Ecuyer(1990)] P. L'Ecuyer. Random numbers for simulation. *Communications of the ACM*, 33(10), Octobre 1990.

[Longstaff and Schwartz(1998)] F.A. Longstaff and E.S. Schwartz. Valuing american options by simulations:a simple least-squares approach. *Working Paper Anderson Graduate School of Management University of California*, 25, 1998.

[Major(2013)] Péter Major. Central limit theorems for martingales. Technical report, Mathematical Institute of the Hungarian Academy of Sciences, 2013. `http://www.math-inst.hu/~major/probability/martingale.pdf`.

[Niederreiter(1995)] H. Niederreiter. New developments in uniform pseudorandom number and vector generation. In P.J.-S. Shiue and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 87–120, Heidelberg, New York, 1995. Springer-Verlag.

[Niederreiter(1992)] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[Pages and Bally(2000)] G. Pages and V. Bally. A quantization algorithm for solving multi–dimensional optimal stopping problems. *Technical report 628, université Paris 6*, 2000.

[Pages et al.(2001)] G. Pages, V. Bally, and Printems. A stochastic optimisation method for nonlinear problems. *Technical report number 02, University Paris XII*, 2001.

[Pages(1995)] J.C. Fort G. Pages. About the a.s. convergence of the Kohonen algorithm with a general neighborhood function. *The Annals of Applied Probability*, 5(4), 1995.

[Pagès and Xiao(1997)] G. Pagès and Y.J. Xiao. Sequences with low discrepancy and pseudo random numbers: theoretical results and numerical tests. *J. Statist. Comput. Simul*, 56: 163–188, 1997.

[Press et al.(1992)] W. H. Press, Saul A. Teukolsky, W. T. Vetterling, and Brian P. Flannery. *Numerical receipes in C*. Cambridge University Press, Cambridge, second edition, 1992. The art of scientific computing.

[Radovic et al.(1996)] Igor Radovic, Ilya M. Sobol', and Robert F. Tichy. Quasi-Monte Carlo methods for numerical integration: Comparison of different low discrepancy sequences. *Monte Carlo Methods Appl.*, 2(1):1–14, 1996.

[Ripley(2006)] Brian D. Ripley. *Stochastic simulation*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006. Reprint of the 1987 original, Wiley-Interscience Paperback Series.

[Roman(1992)] S. Roman. *Coding and information theory*. Number 134 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1992.

[Roth(1954)] K. F. Roth. On irregularities of distribution. *Mathematika*, 1:73–79, 1954.

[Rubinstein(1981)] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley Series in Probabilities and Mathematical Statistics, 1981.

[Sabin and Gray(1986)] M.J. Sabin and R.M. Gray. Global convergence and empirical consistency of the generalised lloyd algorithm. *IEEE Transactions on Information Theory*, 32: 148–155, March 1986.

[Sobol'(1967)] I.M. Sobol'. On the distribution of points in a cube and the approximate evaluation od integral. *USSR Computational Math. and Math. Phys.*, 7:86–112, 1967.

[Sobol'(1976)] I.M. Sobol'. Uniformly distributed sequences with an additional unifrom property. *USSR Computational Math. and Math. Phys.*, pages 236–242, 1976.

[Tsitsiklis and Roy(2000)] J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *Working Paper*, MIT:1–22, 2000.

[Williams(1991)] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.