Monte Carlo Methods
and Stochastic Algorithms

BERNARD LAPEYRE AND BENJAMIN JOURDAIN

November 8, 2023

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 measurable 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 $\mathbb{E}(|X_1|) < +\infty$. Then one has :*

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

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

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

The Central Limit Theorem precises the asymptotic distribution of $\sqrt{n}\varepsilon_n$.

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 denote the standard deviation of X_1, that is*

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

When $\sigma > 0$,

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

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

Note that when $\sigma = 0$, then $\mathbb{P}(\varepsilon_n = 0) = 1$.

Remark 1.1.2. Let us suppose that $\sigma > 0$. By definition of the convergence in distribution, for each continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[f(\frac{\sqrt{n}}{\sigma}\varepsilon_n)]$ converges to $\mathbb{E}[f(G)]$ as $n \to \infty$. This convergence extends to bounded measurable functions f such that $\mathbb{P}(G \in \mathscr{D}_f) = 0$, where $\mathscr{D}_f$ denotes the set of points where f is not continuous. It follows that for all $c_1 < c_2$

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

In practice, one applies the following approximate rule, for n large enough, the law of ε_n is close to the Gaussian law 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 :

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

How to estimate the variance The previous result shows that it is crucial to estimate the standard deviation σ of the random variable. It is easy to do this by using the same samples as for the expectation. Let X be a square integrable (i.e. such that $\mathbb{E}(X_1^2) < \infty$) 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

$$\bar{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} \bar{X}_n^2.$$

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 $\sum_{i=1}^{n} X_i^2$. Since

$$\mathbb{E}(\bar{X}_n^2) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbb{E}(X_i X_j) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \frac{2}{n^2} \sum_{1 \leq j < i \leq n} \mathbb{E}(X_i)\mathbb{E}(X_j) = \frac{\mathbb{E}(X_1^2)}{n} + \frac{n-1}{n} (\mathbb{E}(X_1))^2,$$

$$\mathbb{E}(\bar{\sigma}_n^2) = \frac{n}{n-1} \left(\mathbb{E}(X_1^2) - \frac{\mathbb{E}(X_1^2)}{n} - \frac{n-1}{n} (\mathbb{E}(X_1))^2 \right) = \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 = \sigma^2,$$

and the estimator $\bar{\sigma}_n^2$ is unbiased. Moreover, the Strong Law of Large numbers implies that $\lim_{n\to+\infty} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2, \bar{X}_n\right) = (\mathbb{E}(X_1^2), \mathbb{E}(X_1))$ a.s. so that, since $\lim_{n\to+\infty} \frac{n}{n-1} = 1$, $\lim_{n\to+\infty} \bar{\sigma}_n^2 = \sigma^2$, almost surely. 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, $\mathbb{E}(X)$ belongs to the (random) interval given by

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

In fact, when $\sigma > 0$, $\sqrt{n} \frac{\varepsilon_n}{\bar{\sigma}_n} = \frac{\sigma}{\bar{\sigma}_n} \times \frac{\sqrt{n}\varepsilon_n}{\sigma}$ with the first factor converging a.s. to 1 and the second one converging in law to a centred Gaussian random variable G with variance 1 as $n \to \infty$. By

Slutsky's theorem which applies since the limit of the first factor is deterministic, $\left(\frac{\sigma}{\sigma_n}, \frac{\sqrt{n}\varepsilon_n}{\sigma}\right)$ converges in law to $(1, G)$ as $n \to \infty$. Since the convergence in law is transferred through continuous functions and the product is continuous from $\mathbb{R}^2$ to $\mathbb{R}$, $\sqrt{n}\frac{\varepsilon_n}{\sigma_n} = \frac{\sigma}{\sigma_n} \times \frac{\sqrt{n}\varepsilon_n}{\sigma}$ converges in law to $1 \times G = G$.

$\sqrt{n}\frac{\varepsilon_n}{\sigma_n}$ converges in law to G. So that

$$\lim_{n\to\infty} \mathbb{P}\left(\bar{X}_n - \frac{1.96\bar{\sigma}_n}{\sqrt{n}} \leq \mathbb{E}(X) \leq \bar{X}_n + \frac{1.96\bar{\sigma}_n}{\sqrt{n}}\right) = \mathbb{P}(|G| \leq 1.96) \simeq 0.95.$$

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 y_0. The integer y_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 by setting $(u_n = \frac{y_n+1}{m+1})_{n\in\mathbb{N}}$ where the sequence $(y_n)_{n\in\mathbb{N}}$ evolves starting from y_0 according to the following inductive formula:

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

In practice, the seed is set to y_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

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

Using the cumulative distribution function in simulation The simplest method of simulation relies on the use of the quantile function. The cumulative distribution function and quantile function of a real random variable X are respectively given by

$$\forall x \in \mathbb{R},\ F(x) = \mathbb{P}(X \leq x) \text{ and } \forall u \in (0,1),\ F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

When the cumulative distribution function F is continuous and increasing (which holds when X admits a density with respect to the Lebesgue measure which is a.e. positive under this measure), then it is invertible with inverse F^{-1}. In general, the quantile function F^{-1} is the left-continuous pseudo-inverse of the cumulative distribution function F. Note that

$$\forall u \in (0,1),\ \forall x \in \mathbb{R},\ F^{-1}(u) \leq x \Leftrightarrow u \leq F(x). \tag{1.2}$$

Indeed, by definition of F^{-1}, $u \leq F(x) \Rightarrow F^{-1}(u) \leq x$. Conversely, if $F^{-1}(u) \leq x$, then since F is non-decreasing, $F(F^{-1}(u)) \leq F(x)$. One easily concludes since, by right-continuity of F at $F^{-1}(u)$ and definition of $F^{-1}(u)$, $u \leq F(F^{-1}(u))$.

Proposition 1.2.1. *Let U be a random variable uniformly distributed on $[0,1]$. Then $F^{-1}(U)$ has the same law as X.*

Proof. For $x \in \mathbb{R}$, by (1.2), $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$ so that, since $F(x) \in [0,1]$,

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

So $F^{-1}(U)$ and X have the same cumulative distribution function and, hence, the same law. □

Simulation according to an exponential law 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 $F(x) = (1 - e^{-\lambda x})\mathbf{1}_{\mathbb{R}_+}(x)$, so the equation $F(x) = u$ can be solved as $x = -\frac{1}{\lambda}\ln(1-u)$. If U is uniformly distributed on $[0,1]$, then $-\frac{1}{\lambda}\ln(1-U)$ follows the exponential law with parameter λ and so does $-\frac{1}{\lambda}\ln(U)$ since $1-U$ has the same distribution as U.

Simulation according to a Cauchy distribution The density of the Cauchy distribution with parameter $a > 0$ is $\frac{a}{\pi(x^2+a^2)}$ so that the associated cumulative distribution function and quantile functions are $F(x) = \frac{1}{\pi}\arctan(\frac{x}{a}) + \frac{1}{2}$ and $F^{-1}(u) = a\tan(\pi(u - \frac{1}{2}))$. Hence, when U is uniformly distributed on $[0,1]$, $a\tan(\pi(U - \frac{1}{2}))$ follows the Cauchy distribution with parameter a and so does $a\tan(\pi U)$ since $\pi(U - \frac{1}{2})$ and πU are respectively uniformly distributed on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[0, \pi]$ and the function tan is periodic with period π.

Remark 1.2.2. 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. When $0 < \mathbb{P}(X \in]a,b]) = F(b) - F(a)$, 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 $X \in]a,b]$. Indeed, for $y \in]a,b]$,

$$\mathbb{P}(Y \leq y) = \mathbb{P}(F(a) + (F(b) - F(a))U \leq F(y)) = \mathbb{P}\left(U \leq \frac{F(y) - F(a)}{F(b) - F(a)}\right) = \frac{F(y) - F(a)}{F(b) - F(a)}.$$

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

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

The most widely used simulation method of a Gaussian law is the Box-Muller method. This method is based upon the following result.

Proposition 1.2.3. *Let U_1 and U_2 be two independent random variables which are uniformly distributed on* $[0,1]$. *Let X and Y be defined by*

$$\begin{aligned} X &= \sqrt{-2\ln U_1}\cos(2\pi U_2), \\ Y &= \sqrt{-2\ln U_1}\sin(2\pi U_2). \end{aligned}$$

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

Proof. We now that $R = \sqrt{-2\ln U_1}$ follows the exponential distribution with parameter $\frac{1}{2}$ and is independent from $\Theta = 2\pi U_2$ which is uniformly distributed on $[0,2\pi]$. Thus for $\varphi : \mathbb{R}^2 \to \mathbb{R}$ measurable and bounded,

$$\mathbb{E}[\varphi(X,Y)] = \mathbb{E}[\varphi(\sqrt{R}\cos\Theta, \sqrt{R}\sin\Theta)] = \frac{1}{4\pi} \int_{r=0}^{+\infty} \int_{\theta=0}^{2\pi} \varphi(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta) e^{-\frac{r}{2}} d\theta dr.$$

We perform the change of variables $(x,y) = (\sqrt{r}\cos\theta, \sqrt{r}\sin\theta)$ which is a C^1 diffeomorphism from $(0,+\infty)\times(0,2\pi)$ to $\mathbb{R}^2\setminus\mathbb{R}_+\times\{0\}$ with Jacobian matrix

$$\frac{D(x,y)}{D(r,\theta)} = \begin{pmatrix} \frac{\cos\theta}{2\sqrt{r}} & -\sqrt{r}\sin\theta \\ \frac{\sin\theta}{2\sqrt{r}} & \sqrt{r}\cos\theta \end{pmatrix} \text{ having determinant } \frac{\cos^2\theta+\sin^2\theta}{2} = \frac{1}{2}.$$

Also using that $r = x^2+y^2$, we conclude that

$$\mathbb{E}[\varphi(X,Y)] = \frac{1}{2\pi}\int_{\mathbb{R}^2}\varphi(x,y)e^{-\frac{x^2+y^2}{2}}dxdy.$$

□

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

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

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

The covariance matrix $C\in\mathbb{R}^{d\times d}$ is symmetric positive semi-definite (since, for each $v\in\mathbb{R}^d$, $v.Cv = \mathbb{E}\left((v.X)^2\right)\ge 0$). 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 $A = (a_{ij}, 1\le i,j\le n)$.

Moreover one can compute a square root of a given positive definite symmetric matrix by specifying that $a_{ij} = 0$ for $i<j$ (i.e. A is a lower triangular matrix). Then one has

$$c_{ij} = \sum_{k=1}^{i\wedge j} a_{ik}a_{jk}.$$

When the diagonal coefficients of A are chosen non-negative, A is uniquely obtained by computing its first column

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

and then for j increasing from 2 to d, knowing the $j-1$ first columns, the j-th column is obtained by

$$
\begin{aligned}
a_{jj} &:= \sqrt{c_{jj} - \sum_{k=1}^{j-1} a_{jk}^2}, \qquad (1.3)\\
\text{For } 1 < j < i \le d & \qquad (1.4)\\
a_{ij} &:= \frac{c_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}}{a_{jj}},\\
\text{For } 1 < i < j &\\
a_{ij} &:= 0. \qquad (1.5)
\end{aligned}
$$

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 $X = AG$.

Note that it is not clear that $c_{jj} - \sum_{k=1}^{j-1} a_{jk}^2 \ge 0$ so that the square-root in (1.3) is well defined. This is ensured by the existence for any $C \in \mathbb{R}^{d\times d}$ symmetric positive semi-definite of a lower triangular matrix A such that $AA^* = C$ which can be proved by induction on the dimension d. In dimension $d = 1$, the existence is clear. Let us check that existence in dimension d implies existence in dimension $d+1$. Let $\tilde{C} \in \mathbb{R}^{(d+1)\times(d+1)}$ be symmetric positive semi-definite and $C \in \mathbb{R}^{d\times d}, \mathbf{b} \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$ be defined by $\tilde{C} = \begin{pmatrix} \alpha & \mathbf{b}^* \\ \mathbf{b} & C \end{pmatrix}$. One has

$$\forall (x, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}^d,\ 0 \le f(x,\mathbf{y}) := (x, \mathbf{y}^*)\tilde{C} \begin{pmatrix} x \\ \mathbf{y} \end{pmatrix} = \alpha x^2 + 2(\mathbf{b}^*\mathbf{y})x + \mathbf{y}^* C \mathbf{y}.$$

In particular, $\alpha \ge 0$ (choice $\mathbf{y} = 0$) and the symmetric matrix C is positive semi-definite (choice $x = 0$).

When $\alpha = 0$, then the non-negativity of $\mathbb{R} \ni x \mapsto f(x, \mathbf{b}) = 2x|\mathbf{b}|^2 + \mathbf{b}^* C \mathbf{b}$ implies that $\mathbf{b} = \mathbf{0}$ the nul vector in $\mathbb{R}^d$. Let $A \in \mathbb{R}^{d\times d}$ denote a lower-triangular matrix such that $AA^* = C$. The matrix $\tilde{A} = \begin{pmatrix} 0 & \mathbf{0}^* \\ \mathbf{0} & A \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$ is lower-triangular and such that $\tilde{C} = \tilde{A}\tilde{A}^*$.

When $\alpha > 0$, then for $\mathbf{y} \in \mathbb{R}^d$, the non-negativity of the quadratic function $\mathbb{R} \ni x \mapsto f(x,\mathbf{y}) = \alpha x^2 + 2(\mathbf{b}^*\mathbf{y})x + \mathbf{y}^* C \mathbf{y} = (x, \mathbf{y}^*)\tilde{C} \begin{pmatrix} x \\ \mathbf{y} \end{pmatrix}$ implies non-positivity of its discriminant $\Delta = 4(\mathbf{b}^*\mathbf{y})^2 - 4\alpha \mathbf{y}^* C \mathbf{y}$ so that $\mathbf{y}^* \left(C - \frac{1}{\alpha}\mathbf{b}\mathbf{b}^*\right) \mathbf{y} \ge 0$. We deduce that the Schur complement $C - \frac{1}{\alpha}\mathbf{b}\mathbf{b}^*$ is positive semi-definite. Let $A \in \mathbb{R}^{d\times d}$ denote a lower-triangular matrix such that $AA^* = C - \frac{1}{\alpha}\mathbf{b}\mathbf{b}^*$. The matrix $\tilde{A} = \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{1}{\sqrt{\alpha}}\mathbf{b} & A \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$ is lower-triangular and such that $\tilde{A}\tilde{A}^* = \tilde{C}$.

Discrete law Consider a random variable X taking values in a countable set $\{x_k, k \in \mathbb{N}\}$. 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.

Some specific techniques are preferred for the usual distributions on $\mathbb{N}$.

Binomial law with parameter $(n,p) \in \mathbb{N}^* \times [0,1]$ If $(U_i)_{1\leq i\leq n}$ are independent and uniformly distributed on $[0,1]$ then the random variables $\mathbf{1}_{\{U_i \leq p\}}$ are independent Bernoulli random variables with parameter p so that their sum $\sum_{i=1}^n \mathbf{1}_{\{U_i \leq p\}}$ is distributed according to the binomial law with parameter (n,p).

Geometric distribution with parameter $p \in (0,1]$ Let for $x \in \mathbb{R}$, $\lceil x \rceil$ denote the integer such that $\lceil x \rceil - 1 < x \leq \lceil x \rceil$ and X be distributed according to the exponential law with parameter $\lambda > 0$. Then for $n \in \mathbb{N}^*$

$$\mathbb{P}(\lceil X \rceil = n) = \mathbb{P}(n-1 < X \leq n) = \int_{n-1}^{n} \lambda e^{-\lambda x} dx = e^{-\lambda(n-1)}(1-e^{-\lambda}).$$

Thus $\lceil X \rceil$ is distributed according to the geometric law with parameter p when $1 - e^{-\lambda} = p$ i.e. $\lambda = -\ln(1-p)$. With the simulation of the exponential law, we conclude that so does $\lceil \frac{\ln U}{\ln(1-p)} \rceil$ when U is uniformly distributed on $[0,1]$.

Poisson distribution with parameter $\theta > 0$ If $(U_i)_{i\geq 1}$ are i.i.d. according to the uniform law on $[0,1]$, then

$$\nu := \inf\left\{ n \in \mathbb{N} : \prod_{i=1}^{n+1} U_i < e^{-\theta} \right\}.$$

is distributed according to the Poisson random distribution with parameter θ. Indeed, $\mathbb{P}(\nu = 0) = \mathbb{P}\left(U_1 < e^{-\theta}\right) = e^{-\theta}$.

And for $n \in \mathbb{N}^*$, we have

$$\begin{aligned}\mathbb{P}(\nu = n) = \mathbb{P}\left(\prod_{i=1}^{n} U_i \geq e^{-\theta} > \prod_{i=1}^{n+1} U_i\right) &= \mathbb{P}\left(\sum_{i=1}^{n} \frac{-\ln U_i}{\theta} \leq 1 < \sum_{i=1}^{n+1} \frac{-\ln U_i}{\theta}\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} \frac{-\ln U_i}{\theta} \leq 1\right) - \mathbb{P}\left(\sum_{i=1}^{n+1} \frac{-\ln U_i}{\theta} \leq 1\right).\end{aligned}$$

Since the random variables $-\frac{1}{\theta}\ln(U_i)$ are i.i.d. according to the exponential distribution with parameter θ, for $k \in \mathbb{N}^*$, $\sum_{i=1}^{k} \frac{-\ln U_i}{\theta}$ follows the gamma distribution with parameter (k, θ) and density $\frac{\theta^k x^{k-1}}{(k-1)!}e^{-\theta x}\mathbf{1}_{\{x>0\}}$. For $n \in \mathbb{N}^*$, since, by integration by parts,

$$\mathbb{P}\left(\sum_{i=1}^{n+1}\frac{-\ln U_i}{\theta} \leq 1\right) = \int_0^1 \frac{\theta^{n+1}x^n}{n!}e^{-\theta x}dx = \left[-\frac{\theta^n x^n}{n!}e^{-\theta x}\right]_0^1 + \int_0^1 \frac{\theta^n x^{n-1}}{(n-1)!}e^{-\theta x}dx$$
$$= -\frac{\theta^n}{n!}e^{-\theta} + \mathbb{P}\left(\sum_{i=1}^{n}\frac{-\ln U_i}{\theta} \leq 1\right),$$

we conclude that $\mathbb{P}(\nu = n) = \frac{\theta^n}{n!}e^{-\theta}$.

Rejection sampling Let us suppose that we want to sample according to a distribution μ_Y such that $\mu_Y(dy) = p(y)\mu_X(dy)$ with μ_X a distribution according to which we already know how to sample and p a density ($\int p(y)\mu_X(dy) = 1$) with values in $[0,M]$ with $M < \infty$ (note that if $\mu_Y \neq \mu_X$, then $M > 1$). Let $(X_i,U_i)_{i\geq 1}$ be i.i.d. with X_1 distributed according to μ_X independent from U_1 uniformly distributed on $[0,1]$ and

$$\nu = \inf\left\{i \geq 1 : U_i \leq \frac{p(X_i)}{M}\right\}.$$

Then for $n \in \mathbb{N}^*$ and φ a measurable and bounded function, by the i.i.d. property,

$$\mathbb{E}\left[\mathbf{1}_{\{\nu=n\}}\varphi(X_\nu)\right] = \mathbb{E}\left[\prod_{i=1}^{n-1}\mathbf{1}_{\{U_i>\frac{p(X_i)}{M}\}} \times \mathbf{1}_{\{U_n\leq\frac{p(X_n)}{M}\}}\varphi(X_n)\right] = \mathbb{P}\left(U_1 > \frac{p(X_1)}{M}\right)^{n-1}\mathbb{E}\left[\mathbf{1}_{\{U_1\leq\frac{p(X_1)}{M}\}}\varphi(X_1)\right].$$

By the tower property of the conditional expectation then the freezing Lemma,

$$\mathbb{E}\left[\mathbf{1}_{\{U_1\leq\frac{p(X_1)}{M}\}}\varphi(X_1)\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{U_1\leq\frac{p(X_1)}{M}\}}\varphi(X_1)\middle|X_1\right]\right] = \mathbb{E}\left[\varphi(X_1)\mathbb{E}\left[\mathbf{1}_{\{U_1\leq\frac{p(X_1)}{M}\}}\middle|X_1\right]\right]$$
$$= \mathbb{E}\left[\varphi(X_1)\frac{p(X_1)}{M}\right] = \frac{1}{M}\int\varphi(x)p(x)\mu_X(dx) = \frac{1}{M}\int\varphi(y)\mu_Y(dy).$$

In the same way $\mathbb{P}\left(U_1 > \frac{p(X_1)}{M}\right) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{U_1>\frac{p(X_1)}{M}\}}\middle|X_1\right]\right] = \mathbb{E}\left[1-\frac{p(X_1)}{M}\right] = 1-\frac{1}{M}$. Thus

$$\mathbb{E}\left[\mathbf{1}_{\{\nu=n\}}\varphi(X_\nu)\right] = \left(1-\frac{1}{M}\right)^{n-1}\frac{1}{M}\int\varphi(y)\mu_Y(dy),$$

so that X_ν is distributed according to μ_Y and independent from ν which follows the geometric distribution with parameter $1-\frac{1}{M}$.

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

We have shown in the preceding section 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

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

using a random variable Y with lower variance than $g(X)$. Then we approximate $\mathbb{E}(X)$ by the empirical mean $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ of random variables i.i.d. according to the law of Y.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 $\mathbb{E}(X)$ as

$$\mathbb{E}(X) = \mathbb{E}(X - (Z - \mathbb{E}(Z))),$$

where Z is a square integrable random variable with positive variance such that $\mathbb{E}(Z)$ can be explicitly computed and which is in some sense close to X so that Var $(X-Z)$ is smaller than Var (X). In these circumstances, we use a Monte-Carlo method to estimate $\mathbb{E}(X-Z)$, and we add the value of $\mathbb{E}(Z)$.

Nothing guarantees that Var $(X-Z) \leq$ Var (X) but one can achieve variance reduction by introducing some multiplicative parameter $\alpha \in \mathbb{R}$. The function

$$v(\alpha) := \text{Var }(X-\alpha Z) = \text{Cov}(X-\alpha Z, X-\alpha Z) = \text{Var }(X) - 2\alpha\text{Cov}(X,Z) + \alpha^2\text{Var }(Z)$$

attains its minimum equal to Var $(X) - \frac{(\text{Cov}(X,Z))^2}{\text{Var }(Z)} = \text{Var }(X)(1-\text{Corr}(X,Z)^2)$ at $\alpha_\star = \frac{\text{Cov}(X,Z)}{\text{Var }(Z)}$. Let $((X_i,Z_i))_{i\geq 1}$ be independent copies of (X,Z). By the strong law of large numbers, the estimator $\hat{\alpha}_n = \frac{\frac{1}{n}\sum_{i=1}^n X_iZ_i - \bar{X}_n \times \bar{Z}_n}{\frac{1}{n-1}\sum_{i=1}^n Z_i^2 - (\bar{Z}_n)^2}$ converges a.s. to $\alpha_\star$ as $n\to\infty$ and $\bar{X}_n - \hat{\alpha}_n(\bar{Z}_n - \mathbb{E}(Z))$ converges a.s. to $\mathbb{E}(X)$. Moreover,

$$\sqrt{n}\,(\bar{X}_n - \hat{\alpha}_n(\bar{Z}_n - \mathbb{E}(Z)) - \mathbb{E}(X)) = (1, -\hat{\alpha}_n)\sqrt{n}\begin{pmatrix} \bar{X}_n - \mathbb{E}(X) \\ \bar{Z}_n - \mathbb{E}(Z) \end{pmatrix}$$

where $\sqrt{n}\begin{pmatrix} \bar{X}_n - \mathbb{E}(X) \\ \bar{Z}_n - \mathbb{E}(Z) \end{pmatrix}$ converges in law to $W \sim \mathcal{N}_2\left(0, \begin{pmatrix} \text{Var }(X) & \text{Cov}(X,Z) \\ \text{Cov}(X,Z) & \text{Var }(Z) \end{pmatrix}\right)$ as $n\to\infty$. With Slutsky's lemma and the continuity of the scalar product on $\mathbb{R}^2$, we conclude that$\sqrt{n}\,(\bar{X}_n - \hat{\alpha}_n(\bar{Z}_n - \mathbb{E}(Z)) - \mathbb{E}(X))$ converges in distribution to $(1, -\alpha^\star)W \sim \mathcal{N}_1(0, v(\alpha_\star))$.

Let us illustrate the control variates approach 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 $\bar{S}_T = \frac{1}{T}\int_0^T S_s ds$. 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,

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

Moreover unlike the arithmetic mean $\bar{S}_T$ the distribution of which is unknown, the geometric mean $\tilde{S}_T = S_0 \exp\left(\frac{1}{T}\int_0^T \left(\sigma W_t + (r - \frac{\sigma^2}{2})tdt\right)\right)$ has a log-normal distribution. Indeed $\frac{1}{T}\int_0^T \left(\sigma W_t + (r - \frac{\sigma^2}{2})tdt\right)$ is normal with expectation $(r - \frac{\sigma^2}{2})\frac{T}{2}$ and variance

$$\mathbb{E}\left[\left(\frac{\sigma}{T}\int_0^T W_t dt\right)^2\right] = \frac{\sigma^2}{T^2}\int_{t=0}^T \int_{s=0}^T \mathbb{E}[W_s W_t] ds dt = \frac{2\sigma^2}{T^2}\int_{t=0}^T \int_{s=0}^t s ds dt = \frac{\sigma^2 T}{3}.$$

Hence an explicit formula is available for $\mathbb{E}\left[(K - \tilde{S}_T)_+\right]$ and we can use $(K - \tilde{S}_T)_+$ as a control variate when computing $\mathbb{E}\left[(K - \bar{S}_T)_+\right]$.

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, $1 \leq i \leq d$), be positive real numbers. We want to compute a put option on a basket

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

where $X = a_1 S_T^1 + \cdots + a_d S_T^d$. For $\left(p_i = \frac{a_i x_j}{\sum_{j=1}^d a_j x_j}\right)_{1 \leq i \leq d}$, the idea is to approximate

$$X = \left(\sum_{j=1}^d a_j x_j\right) \times \left(\sum_{i=1}^d p_i e^{\left(r - \frac{1}{2}\sum_{j=1}^p \sigma_{ij}^2\right)T + \sum_{j=1}^p \sigma_{ij} W_T^j}\right)$$

by the log-normal random variable obtained by replacing the rithmetic mean by the corresponding geometric mean

$$Y = \left(\sum_{j=1}^d a_j x_j\right) e^{\sum_{i=1}^d p_i \left(\left(r - \frac{1}{2}\sum_{j=1}^p \sigma_{ij}^2\right)T + \sum_{j=1}^p \sigma_{ij} W_T^j\right)}.$$

As we can compute an explicit formula for

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

we can 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 $\mathbb{R}^d$, then

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

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

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

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

$$\text{Var}\,(Z) = \int_{\mathbb{R}^d}\left(\frac{g(x)f(x)}{\tilde{f}(x)}\right)^2\tilde{f}(x)dx - \mathbb{E}(g(X))^2 \geq \left(\int_{\mathbb{R}^d}|g(x)|f(x)dx\right)^2 - \mathbb{E}(g(X))^2,$$

by the Cauchy-Schwarz inequality. The lower bound is attained for $\tilde{f}_\star(x) = \frac{|g(x)|f(x)}{\mathbb{E}[|g(X)|]}$ and is even equal to zero when g has constant sign. The computation of the normalizing constant $\mathbb{E}[|g(X)|]$ of $\tilde{f}_\star(x)$ being as difficult as (and, when g has constant sign, equivalent to) the computation of $\mathbb{E}[g(X)]$, this optimal choice cannot in general be used in practice. Nevertheless, this leads to the following heuristic approach : choose $\tilde{f}(x)$ as a good approximation of $|g(x)|f(x)$ such that the distribution with density $\tilde{f}(x)/\int_{\mathbb{R}^d}\tilde{f}(x)dx$ 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

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

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 :

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

This equality can be rewritten as

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

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 since the option will unlikely 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 9 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)Glasserman, Heidelberger, and Shahabuddin].

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

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

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

Moreover, denote by I_t the value of the index

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

where $a_1, \dots, a_d$ is a given set of positive numbers such that $\sum_{i=1}^{d} a_i = 1$. 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

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

there exists a function ϕ such that

$$h = \phi\left(G_1, \dots, G_p\right),$$

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

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

where $G = (G_1, \dots, G_p)$ is a p-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 $m = (m_1, \dots, m_p)$,

$$\mathbb{E}\left(\phi(G)\right) = \mathbb{E}\left(\phi(G+m) e^{-m.G - \frac{|m|^2}{2}} \right), \tag{1.6}$$

where $m.G = \sum_{i=1}^{p} m_i G_i$ and $|m|^2 = \sum_{i=1}^{p} m_i^2$. In view of 1.6, the variance $V(m)$ of the random variable

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

is

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

Let us suppose that $\mathbb{P}(\phi(G) \neq 0) > 0$ and $\forall \lambda \in \mathbb{R}^d$, $\mathbb{E}\left(\phi^2(G)e^{\lambda.G}\right) < \infty$. Since $m.G \leq \frac{|m|^2}{4} + |G|^2$, one has $e^{-m.G+\frac{|m|^2}{2}} \geq e^{\frac{|m|^2}{4}-|G|^2}$ so that $V(m) \geq e^{\frac{|m|^2}{4}}\mathbb{E}\left(\phi^2(G)e^{-|G|^2}\right) - \mathbb{E}^2(\phi(G))$. We deduce that $\lim_{|m|\to+\infty} V(m) = +\infty$. On the other hand, one can interchange the expectation and derivatives with respect to m to obtain

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

$$\nabla^2 V(m) = \mathbb{E}\left(\phi^2(G)e^{-m.G+\frac{|m|^2}{2}}(I_d + (m-G)(m-G)^*)\right).$$

Hence $V(m)$ is a strictly convex function which goes to $+\infty$ with $|m|$. We conclude that there is a unique $m_\star \in \mathbb{R}^d$ such that $V(m_\star) = \inf_{m\in\mathbb{R}^d} V(m)$ and $m_\star$ is characterized by $\nabla V(m_\star) = 0$. One may approximate $m_\star$ by either solving this equation using a stochastic algorithm or by minimizing $\frac{1}{n}\sum_{i=1}^n \phi^2(G_i)e^{-m.G_i+\frac{|m|^2}{2}}$ where the G_i are i.i.d. copies of G.

The reader is also refered to [Glasserman et al.(1999)Glasserman, Heidelberger, and Shahabuddin] for an almost optimal way to choose the parameter m.

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 $U_1,\ldots,U_n$ 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

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

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 $\text{Cov}(g(U_1),g(1-U_1)) \leq 0$. If g is a monotonic function this is always true and thus the Monte-Carlo method using antithetic variables is better than the standard one. Indeed, when f and g are two functions with the same monotony, $(f(U_1)-f(U_2))(g(1-U_1)-g(1-U_2)) \leq 0$ so that, by taking the expectation and using that U_1 and U_2 are i.i.d.,

$$2\left(\mathbb{E}[f(U_1)g(1-U_1)]-\mathbb{E}[f(U_1)]\mathbb{E}[g(1-U_1)]\right) \leq 0 \text{ and } \text{Cov}(g(U_1),g(1-U_1)) \leq 0.$$

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

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

When g is monotonic in each of its variables, $\text{Cov}(g(U_1,\ldots,U_d),g(1-U_1,\ldots,1-U_d)) \leq 0$. Indeed, one can prove by induction on d that when f and g are monotonic in each of their variables with the same monotony for a given variable,

$$\mathbb{E}[f(U_1,\ldots,U_d)g(1-U_1\ldots,1-U_d)] \leq \mathbb{E}[f(U_1,\ldots,U_d)]\mathbb{E}[g(1-U_1\ldots,1-U_d)].$$

We have treated the case $d=1$ above. Let us suppose that the property holds at rank d. One has

$$\mathbb{E}[f(U_1\ldots,U_d,U_{d+1})g(1-U_1\ldots,1-U_d,1-U_{d+1})] = \mathbb{E}[H(U_{d+1})],$$

where, $H(u_{d+1}) = \mathbb{E}[f(U_1,\ldots,U_d,u_{d+1})g(1-U_1,\ldots,1-U_d,1-u_{d+1})]$ by the freezing Lemma. When f and g are monotonic in each of their variables with the same monotony for a given variable, by the property at rank d, $H(u_{d+1}) \leq F(u_{d+1})G(1-u_{d+1})$ with $F(x) = \mathbb{E}[f(U_1,\ldots,U_d,x)]$ and $G(x) = \mathbb{E}[g(1-U_1,\ldots,1-U_d,x)]$ having the same monotony. Hence, using the property at rank 1 for the second inequality, we obtain

$$\begin{aligned}\mathbb{E}[H(U_{d+1})] &\leq \mathbb{E}[F(U_{d+1})G(1-U_{d+1})] \leq \mathbb{E}[F(U_{d+1})]\mathbb{E}[G(1-U_{d+1})]\\ &= \mathbb{E}[f(U_1\ldots,U_d,U_{d+1})]\mathbb{E}[g(1-U_1\ldots,1-U_d,1-U_{d+1})].\end{aligned}$$

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

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

The same computations as before prove that the estimator I_{2n} is better than the crude one if and only if $\text{Cov}(g(X),g(T(X))) \leq 0$. 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 $g(x)=(\lambda e^{\sigma x}-K)_+$.

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

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

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

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

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

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 efficient simulation according to the law of X^i is possible, one can use a Monte-Carlo method to approximate each conditional expectation $\mathscr{E}_i = \mathbb{E}(g(X)|X \in D_i)$ by

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

where $(X_1^i, \ldots, X_{n_i}^i)$ are independent copies of X^i. When the numbers $p_i = \mathbb{P}(X \in D_i)$ can be explicitly computed, an estimator $\tilde{\mathscr{E}}$ of $\mathscr{E}$ is given by

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

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

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

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

Fix the total number of simulations $\sum_{i=1}^{I} n_i = n$ and denote by $q_i = \frac{n_i}{n}$ the proportion of simulations affected to stratum i. Then the above variance writes

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

This lower bound is attained for

$$q_i^\star = \frac{p_i \sigma_i}{\sum_{j=1}^{I} p_j \sigma_j},\ i \in \{1, \ldots, I\}.$$

Note that this variance is smaller than the one obtained without stratification. Indeed, using again the convexity inequality $\sum_{i=1}^{I} p_i a_i^2 \geq \left(\sum_{i=1}^{I} p_i a_i\right)^2$, we obtain

$$\begin{aligned}
\text{Var }(g(X)) &= \mathbb{E}\left(g(X)^2\right) - \mathbb{E}\left(g(X)\right)^2 \\
&= \sum_{i=1}^{I} p_i \mathbb{E}\left(g^2(X)|X \in D_i\right) - \left(\sum_{i=1}^{I} p_i \mathbb{E}\left(g(X)|X \in D_i\right)\right)^2 \\
&= \sum_{i=1}^{I} p_i \text{Var }\left(g(X)|X \in D_i\right) + \sum_{i=1}^{I} p_i \mathbb{E}\left(g(X)|X \in D_i\right)^2 - \left(\sum_{i=1}^{I} p_i \mathbb{E}\left(g(X)|X \in D_i\right)\right)^2 \\
&\geq \textstyle\sum_{i=1}^{I} p_i \sigma_i^2,
\end{aligned}$$

where the right-hand side is n times the variance of the stratified estimator for the choice $q_i = p_i$ for $i \in \{1, \ldots, I\}$.

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 q_i may *increase* the variance. A common way to circumvent this difficulty is to choose $q_i = p_i$ for $i \in \{1, \ldots, I\}$. 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 need 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 cumulative distribution function of a standard 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)Glasserman, Heidelberger, and S The computation of an European basket option in a multidimensional Black-Scholes model can be expressed as

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

for some function g and for $G = (G_1, \ldots, G_d)$ a vector of independent standard Gaussian random variables. Choose a vector $u \in \mathbb{R}^d$ such that $|u| = 1$. Then

$$\text{Var}\,(< u, G >) = u^* I_d u = |u|^2 = 1,$$

so that $< u, G >= u_1 G_1 + \cdots + u_d G_d$ is also a standard Gaussian random variable. Then choose a partition $(B_i, 1 \leq i \leq I)$ of $\mathbb{R}$ such that

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

and define the strata by setting

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

This can be done by setting

$$B_i =]\mathcal{N}^{-1}((i-1)/I), \mathcal{N}^{-1}(i/I)],$$

where $\mathcal{N}$ is the distribution function of a standard Gaussian random variable and $\mathcal{N}^{-1}$ is its inverse. By page 6, when $U \sim \mathcal{U}[0,1]$, the random variable $\mathcal{N}^{-1}\left(\frac{i-1}{I} + \frac{U}{I}\right)$ follows the law a standard Gaussian random variable conditioned to be in B_i and so does $\mathcal{N}^{-1}\left(\frac{i-U}{I}\right)$. Now for $j \in \{1, \cdots, d\}$,

$$\begin{aligned}\text{Cov}(G_j - < u, G > u_j, < u, G >) &= \text{Cov}\left(G_j, \sum_{k=1}^{d} u_k G_k\right) - \text{Var}\,(< u, G >) u_j \\ &= \sum_{k=1}^{d} u_k \text{Cov}(G_j, G_k) - u_j = 0.\end{aligned}$$

Since $(G - < u, G > u, < u, G >)$ is a Gaussian random vector as a linear transform of the Gaussian random vector G, we deduce that $G - < u, G > u$ and $< u, G >$ are independent. Hence when U is independent of G, then $\left(G - < u, G > u, \mathcal{N}^{-1}\left(\frac{i-U}{I}\right)\right)$ follows the conditional law of $(G - < u, G > u, < u, G >)$ given $< u, G >\in B_i$ and

$$G + \left(\mathcal{N}^{-1}\left(\frac{i-U}{I}\right) - < u, G >\right) u$$

follows the conditional law of G given $< u, G > \in B_i$.

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)Glasserman, Heidelberger, and Shahabuddin].

1.3.5 Mean value or conditioning

This method uses the well known fact that conditioning reduces the variance. Indeed, for any 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

$$0 \leq \mathbb{E}\left[(Z - \mathbb{E}[Z|Y])^2\right] = \mathbb{E}[Z^2] - \mathbb{E}\left[\mathbb{E}[Z|Y]^2\right] = \mathbb{E}[Z^2] - \mathbb{E}[\phi^2(Y)],$$

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

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 $Z = g(X,Y)$, where X and Y are independent random variables. In this case, we have

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

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

A basic example. Suppose that we want to compute $\mathbb{P}(X \leq Y)$ 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

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

where F is the cumulative distribution function of X. The variance reduction can be significant, especially when the probability $\mathbb{P}(X \leq Y)$ 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 $(x_k)_{k\geq 1}$ 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. It is not difficult to choose n points such that the approximation is good for a fixed value of n : $(x_k^n = \frac{2k-1}{2n})_{1\leq k\leq n}$ is such a good choice in dimension

$d=1$ and there even exist Gauss points $(y_k^n)_{1\le k\le n}$ with companion weights $(\omega_k^n)_{1\le k\le n}\in[0,1]^n$ summing to 1 such that $\sum_{k=1}^n \omega_k^n g(y_k^n)$ is equal to $\int_0^1 g(u)du$ for each polynomial g with degree not greater than $2n-1$. But constructing a sequence $(x_k)_{k\ge 1}$ with good properties for all values of n is not so easy.

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. By definition, if x and y are two points in $[0,1]^d$, $x\le y$ if and only if $x_i\le y_i$, for all $1\le i\le d$.

Definition 1.4.1. *A sequence $(x_n)_{n\ge 1}$ 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 $[0,y]=\{z\in[0,1]^d, 0\le z\le y\}$.

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 $[0,1]^d$*

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

4. *$\forall m=(m_1,\ldots,m_d)\in\mathbb{Z}^d$, $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n e^{i2\pi\langle m,x_k\rangle}=\int_{[0,1]^d} e^{i2\pi\langle m,x\rangle}dx=\mathbf{1}_{\{m=0\}}$.*

Remark 1.4.1.
- Property 3 is the weak convergence of the probability measure $\frac{1}{n}\sum_{k=1}^n\delta_{x_k}$ to the Lebesgue measure on $[0,1]^d$. Property 4 is the characterization of this weak convergence in terms of Fourier transform known as Weyl's criterion.
- If $(U_n)_{n\ge 1}$ is a sequence of independent random variables with uniform law on $[0,1]$, the random sequence

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

is almost surely uniformly distributed. The strong law of large numbers ensures that

$$a.s., \forall m\in\mathbb{Z}^d,\ \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n e^{i2\pi\langle m,U_k\rangle}=\mathbb{E}\left[e^{i2\pi\langle m,U_1\rangle}\right]=\mathbf{1}_{\{m=0\}},$$

where we could interchange a.s. and $\forall m\in\mathbb{Z}^d$ since $\mathbb{Z}^d$ is countable.

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 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 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. Indeed when for $U \sim \mathscr{U}[0,1]^d$, $\phi(U)$ and $\psi(U)$ have the same distribution then $\mathbb{E}[f(\phi(U))] = \mathbb{E}[f(\psi(U))]$ which also writes $\int_{[0,1]^d} f(\phi(u))du = \int_{[0,1]^d} f(\psi(u))du$, Var $(\phi(U)) =$ Var $(\psi(U))$ but in general $V(f \circ \phi)$ is not equal to $V(f \circ \psi)$.

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 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.

In dimension $d = 1$, when g is C^1 with bounded variation, $V(g) = \int_0^1 |g'(u)|du$ and since $g(u) = g(1) - \int_0^1 \mathbf{1}_{u \le v} g'(v)dv$, we have

$$\begin{aligned} \left| \int_0^1 g(u)dy - \frac{1}{n}\sum_{k=1}^n g(x_k) \right| &= \left| \int_0^1 \left(\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{x_k \le v} - \mathbf{1}_{u \le v} \right) g'(v)dv \right| \\ &\le \int_0^1 |g'(v)|dv \sup_{v\in[0,1]} \left| \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{x_k \le v} - \mathbf{1}_{u \le v} \right| = V(g) \times D_n^*(x). \end{aligned}$$

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)].

Irrational translation of the torus These sequences are defined by

$$x_n = (\{n\alpha_1\}, \ldots, \{n\alpha_d\}), n \ge 1 \tag{1.7}$$

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 for each $m \in \mathbb{Z}^d \setminus \{0\}$, $\langle m, \alpha \rangle \notin \mathbb{Z}$ or equivalently $e^{i2\pi\langle m,\alpha\rangle} \neq 1$. Note that this condition implies that the α_i are irrational numbers. Since for $m \in \mathbb{Z}^d$, $k\langle m, \alpha\rangle - \langle m, x_k\rangle = \langle m, (\lfloor k\alpha_1 \rfloor, \cdots, \lfloor k\alpha_d \rfloor)\rangle \in \mathbb{Z}$, $e^{i2\pi\langle m,x_k\rangle} = e^{i2\pi k\langle m,\alpha\rangle}$ and

$$\forall m \in \mathbb{Z}^d \setminus \{0\},\ \frac{1}{n}\sum_{k=1}^{n} e^{i2\pi\langle m,x_k\rangle} = \frac{1}{n}\sum_{k=1}^{n} e^{i2\pi k\langle m,\alpha\rangle} = \frac{e^{i2\pi\langle m,\alpha\rangle}}{n} \times \frac{e^{i2\pi n\langle m,\alpha\rangle} - 1}{e^{i2\pi\langle m,\alpha\rangle} - 1} \to_{n\to\infty} 0.$$

Hence, by point 4 in Definition 1.4.1 , the sequence $(x_n)_{n\geq 1}$ is uniformly distributed.

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

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

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

The Van Der Corput sequence Let p be an integer, $p \geq 2$ and n a positive integer. We denote by $a_0, a_1, \dots, a_r$ 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

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

Note that since $p^r \leq a_0 + a_1 p + \cdots + a_r p^r \leq (p-1) \times (1 + p + \cdots + p^r) = p^{r+1} - 1$, $r = \lfloor \frac{\ln n}{\ln p} \rfloor$, $a_r = \lfloor \frac{n}{p^r} \rfloor$ and for $r - 1 \geq j \geq 0$, $a_j = \lfloor \frac{n - \sum_{i=j+1}^{r} a_i p^i}{p^j} \rfloor$. Using standard notations, n can be writen as

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

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

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

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

Note that if $v(n) = \min\{i \geq 0 : a_i(n) < p - 1\}$ then $n + 1 = (1 + a_{v(n)}(n))p^{v(n)} + \sum_{j=v(n)+1}^{r(n)} a_j(n) p^j$ so that

$$\phi_p(n+1) - \phi_p(n) = \frac{1}{p^{v(n)+1}} - (p-1)\sum_{j=0}^{v(n)-1} \frac{1}{p^{j+1}} = \frac{1}{p^{v(n)+1}} + \frac{1}{p^{v(n)}} - 1.$$

Since $1 - \frac{1}{p^{v(n)}} = \sum_{j=0}^{v(n)-1} \frac{p-1}{p^{j+1}} \leq \phi_p(n) < \sum_{j\in\mathbb{N}} \frac{p-1}{p^{j+1}} - \frac{1}{p^{v(n)+1}} = 1 - \frac{1}{p^{v(n)+1}}$, we deduce that $\phi_p(n+1) = \psi_p(\phi_p(n))$ for the Kakutani transform $\psi_p : [0,1) \to [0,1)$ defined by

$$\psi_p(x) = \sum_{k\geq 0} \mathbf{1}_{[1-\frac{1}{p^k}, 1-\frac{1}{p^{k+1}})}(x)\left(x + \frac{1}{p^k} + \frac{1}{p^{k+1}} - 1\right).$$

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

$$x_n^d = (\phi_{p_1}(n), \cdots, \phi_{p_d}(n)) \tag{1.8}$$

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 p be an odd integer greater or equal to d. Now define a function T on the set of numbers x such that

$$x = \sum_{j=0}^{r} \frac{a_j}{p^{j+1}},$$

where each a_j belongs to $\{0, \ldots, p-1\}$, by

$$T(x) = \sum_{j=0}^{r} \frac{b_j}{p^{j+1}},$$

where

$$b_j = \sum_{i=j}^{r} \left(\begin{array}{c} i \\ j \end{array} \right) a_i \bmod p,$$

and $\left(\begin{array}{c} i \\ j \end{array} \right) = \frac{i!}{j!(i-j)!}$ denotes the binomial coefficient. The Faure sequence is then defined as follows

$$x_n = \Big(\phi_p(n-1), T(\phi_p(n-1)), \cdots, T^{d-1}(\phi_p(n-1))\Big), \tag{1.9}$$

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

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 $(M_1,\dots,M_q)\in\mathbb{N}^q$, 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 $\left\{0,\dots,2^N-1\right\}$ 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)Press, Teukolsky, Vetterling, and Flannery], see also [Fox(1988)]. Empirical studies indicate that Sobol sequences are among the most efficient low discrepancy sequences (see [Fox et al.(1992)Fox, Bratley, and Neiderreiter] and [Radovic et al.(1996)Radovic, Sobol', and Tichy] for numerical comparisons of sequences).

1.5 Exercises and problems

1.5.1 Exercises

Exercise 1. Convergence in probability of the empirical mean without integrability.
Let $(X_j)_{j\geq 1}$ be a sequence of i.i.d. random variables s.t. $\lim_{n\to\infty} n\mathbb{P}(|X_1|>n)=0$.

1. Remark that $X_1^2 1_{\{|X_1|\leq n\}} \leq \sum_{k=1}^n (\sum_{j=1}^k 2j) 1_{\{k-1<|X_1|\leq k\}}$ and show that
$$\mathbb{E}[X_1^2 1_{\{|X_1|\leq n\}}] \leq 2\sum_{j=1}^n j\mathbb{P}(|X_1|>j-1).$$
2. Deduce that $\lim_{n\to\infty} \text{Var}\left(\frac{1}{n}\sum_{j=1}^n X_j 1_{\{|X_j|\leq n\}}\right)=0$.

We suppose the existence of $x\in\mathbb{R}$ such that $\lim_{n\to\infty}\mathbb{E}[X_1 1_{\{|X_1|\leq n\}}]=x$.

3. For $\varepsilon>0$ and n large enough so that $|\mathbb{E}[X_1 1_{\{|X_1|\leq n\}}]-x|\leq \frac{\varepsilon}{2}$, check that
$$\mathbb{P}\left(|\bar{X}_n-x|\geq\varepsilon\right)\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^n X_j 1_{\{|X_j|\leq n\}}-\mathbb{E}[X_1 1_{\{|X_1|\leq n\}}]\right|\geq\frac{\varepsilon}{2}\right)+n\mathbb{P}(|X_1|>n).$$
Conclude that $(\bar{X}_n)_{n\geq 1}$ converges in probability to x as $n\to\infty$.
4. Give an example of a non integrable random variable X_1 with symmetric distribution (i.e. X_1 and $-X_1$ have the same law, which implies that $\mathbb{E}[X_1 1_{\{|X_1|\leq n\}}]=0$) such that $\lim_{n\to\infty} n\mathbb{P}(|X_1|>n)=0$.

Exercise 2. No convergence in probability in the CLT.
Let $(X_j)_{j\geq 1}$ be a sequence of i.i.d. square integrable random variables.

1. Show that, as $n\to\infty$, $\left(\sqrt{n}(\bar{X}_n-\mathbb{E}(X_1)), \frac{1}{\sqrt{n}}\sum_{j=n+1}^{2n}(X_j-\mathbb{E}(X_1))\right)$ converges in distribution to (Y_1,Y_2) with Y_1 and Y_2 i.i.d. and precise their common distribution.
2. Deduce that $\sqrt{2n}(\bar{X}_{2n}-\mathbb{E}(X_1))-\sqrt{n}(\bar{X}_n-\mathbb{E}(X_1))$ converges in distribution to $\sqrt{2-\sqrt{2}}Y_1$.
3. Show that when a sequence $(Z_n)_{n\geq 1}$ converges in probability to some limit Z, then $Z_{2n}-Z_n$ converges in probability to 0.
4. Conclude that, when $\text{Var}\,(X_1)>0$, $\sqrt{n}(\bar{X}_n-\mathbb{E}(X_1))$ does not converge in probability.

Exercise 3. Various rates of convergence of the empirical mean.
Let $(Z_j)_{j\geq 1}$ be a sequence of random variables i.i.d. according to the symmetric Pareto distribution with parameter $\alpha\in]0,2[$, the density of which is $\frac{\alpha}{2|z|^{\alpha+1}}1_{\{|z|\geq 1\}}$.

1. When $\alpha>1$, what is the behaviour of the sequence $\bar{Z}_n=\frac{1}{n}\sum_{j=1}^n Z_j$ as $n\to+\infty$?
2. Check that the common characteristic function Φ of the random variables Z_j satisfies
$$\Phi(u)-1=\alpha|u|^\alpha\int_{|u|}^{+\infty}\frac{\cos(t)-1}{t^{\alpha+1}}dt$$

3. Give an equivalent of $\Phi(u)-1$ as $u\to 0$. Deduce that $n^{\frac{\alpha-1}{\alpha}}\bar{Z}_n$ converges in distribution. What is the rate of convergence to 0 of the empirical mean $\bar{Z}_n$ when $1<\alpha<2$?

Exercise 4. Simulation according to the beta distribution.
Let $a,b>0$ and (U,V) uniformy distributed on $D=\{(u,v)\in\mathbb{R}^2,\ u>0,\ v>0,\ u^{\frac{1}{a}}+v^{\frac{1}{b}}<1\}$ ((U,V) has the density $1_{\{(u,v)\in D\}}/|D|$ where $|D|$ is the area of D).

1. Compute the distribution of $(S,X)=\left(U^{\frac{1}{a}}+V^{\frac{1}{b}},\frac{U^{\frac{1}{a}}}{U^{\frac{1}{a}}+V^{\frac{1}{b}}}\right)$.

2. Check that X follows the beta distribution with parameters a and b, the density of which is $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mathbf{1}_{\{0<x<1\}}x^{a-1}(1-x)^{b-1}$ where, for $c>0$, $\Gamma(c)=\int_0^{+\infty}x^{c-1}e^{-x}dx$. Are the random variables S and X independent? Compute $|D|$.

Exercise 5. Simulation according $\mathscr{N}_1(0,1)$. Let $((X_i,Y_i))_{i\geq 1}$ be i.i.d. with X_1 and Y_1 independent exponential random variables with parameter 1. Let ε be independent of this sequence and such that $\mathbb{P}(\varepsilon=1)=\mathbb{P}(\varepsilon=-1)=\frac{1}{2}$. We set

$$N=\inf\{i\geq 1:\ 2Y_i\geq(1-X_i)^2\}\ \text{ and }\ Z=\varepsilon X_N.$$

1. What is the distribution of N? Compute $\mathbb{E}(N)$.

2. What is the distribution of X_N? Deduce that of Z.

3. Deduce a way to simulate according to $\mathscr{N}_1(0,1)$.

Exercise 6. : Simulation according to the gamma districution
We recall that for $a,\theta>0$, the density of the distribution $\Gamma(a,\theta)$ is $p_{a,\theta}(z)=\frac{\theta^a z^{a-1}}{\Gamma(a)}e^{-\theta z}1_{\{z>0\}}$ where for $a>0$, $\Gamma(a)=\int_0^{+\infty}x^{a-1}e^{-x}dx$. We suppose that $a>1$ and we set $f(z)=z^{a-1}e^{-z}1_{\{z>0\}}$ and $\mathscr{D}_a=\{(x,y)\in\mathbb{R}_+^2:0\leq x\leq\sqrt{f(\frac{y}{x})}\}$.

1. Compute $\sup_{z>0}f(z)$ and $\sup_{z>0}z^2f(z)$. Deduce that $\mathscr{D}_a\subset[0,x_a]\times[0,y_a]$ where $x_a=\left(\frac{a-1}{e}\right)^{\frac{a-1}{2}}$ et $y_a=\left(\frac{a+1}{e}\right)^{\frac{a+1}{2}}$.

2. Let $(X,Y)\sim\mathscr{U}(\mathscr{D}_a)$ be uniformly distributed on $\mathscr{D}_a$ i.e. with density $\frac{1}{|\mathscr{D}_a|}\mathbf{1}_{\{0\leq y\}}\mathbf{1}_{\{0\leq x\leq\sqrt{f(\frac{y}{x})}\}}$ where $|\mathscr{D}_a|$ denotes the area of $\mathscr{D}_a$.
What is the distribution of (X,W) where $W=\frac{Y}{X}$? What is that of W? Deduce that $|\mathscr{D}_a|=\frac{\Gamma(a)}{2}$. Conclude that $Z\overset{\text{def}}{=}\frac{W}{\theta}\sim\Gamma(a,\theta)$.

3. How to simulate according to the distributions $\mathscr{U}(\mathscr{D}_a)$ and $\Gamma(a,\theta)$?

4. To simulate according to $\Gamma(a,1)$, we do not need the constant $\int_{\mathbb{R}}f(x)dx=\Gamma(a)$ which permits to normalize f into $p_{a,1}$. Does replacing f by cf where $c\in]0,+\infty[$ change the efficiency of the above method?

Exercise 7. Letac's bound for the rejection sampling
Let p be a probability density on the interval $[0,1]$ according to which one wants to simulate by a rejection algorithm using a sequence $((U_i,X_i))_{i\geq 1}$ of i.i.d. random vectors with $U_i\sim\mathscr{U}[0,1]$. More precisely, we suppose the existence of an acceptation set $\mathscr{A}$ such that $\mathbb{P}((U_1,X_1)\in\mathscr{A})>0$ and the conditional law of U_1 given $(U_1,X_1)\in\mathscr{A}$ has the density p. Let $N=\min\{i\geq 1:(U_i,X_i)\in\mathscr{A}\}$ and B be a Borel subset of $[0,1]$.

1. What is the distribution of N? And that of U_N?
2. Check that for $n \in \mathbb{N}^*$, $\mathbb{P}(U_n \in B, N \geq n) = \mathbb{P}(U_n \in B)\mathbb{P}(N \geq n)$.
3. Deduce that $\mathbb{P}(U_N \in B) \leq \mathbb{P}(U_1 \in B)\mathbb{E}(N)$.
4. Conclude that $\mathbb{E}(N) \geq \sup\{\rho \geq 0 : \int_0^1 1_{\{p(u)\geq\rho\}} du > 0\}$.

Exercise 8. Let Z be a Gaussian random variable and K a positive real number.

1. Let $d = \frac{\mathbb{E}(Z)-\log(K)}{\sqrt{\text{Var }(Z)}}$, prove that
$$\mathbb{E}\left(\mathbf{1}_{\{Z\geq\log(K)\}}e^Z\right) = e^{\mathbb{E}(Z)+\frac{1}{2}\text{Var }(Z)}\mathcal{N}\left(d+\sqrt{\text{Var }(Z)}\right).$$
2. Prove the formula (Black and Scholes formula)
$$\mathbb{E}\left(\left(e^Z - K\right)_+\right) = e^{\mathbb{E}(Z)+\frac{1}{2}\text{Var }(Z)}\mathcal{N}\left(d+\sqrt{\text{Var }(Z)}\right) - K\mathcal{N}(d),$$

Exercise 9. Let λ and K be two positive real numbers 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 $m \leq m_0 = \ln(K/\lambda)/\sigma$.

Exercise 10. 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 $[0,1]$.

1. Prove that $\frac{1}{N}\sum_{i=1}^N h\left((i-1+U_i)/N\right)$ has a lower variance than $\frac{1}{N}\sum_{i=1}^N h(U_i)$.
2. Interpret this result in terms of a stratification method.

Exercise 11. Let X and Y be independent real random variables. Let F and G be the cumulative distribution functions of X and G respectively. We want to compute by a Monte-Carlo method the probability
$$\theta = \mathbb{P}(X+Y \leq t).$$

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 method. Why does this method reduce the variance?

Exercise 12. 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 positive real numbers. This exercise studies various methods to compute the price of an index option given by $p = \mathbb{P}(Z > t)$.

1. In this question, we assume that (X_1,X_2) is a Gaussian vector with mean 0 such that Var $(X_1) =$ Var $(X_2) = 1$ and $\mathrm{Cov}(X_1,X_2) = \rho$, with $|\rho| \le 1$. 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

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

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

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

Propose a new Monte-Carlo method which allows to compute p.

4. Assuming now that X_1 and X_2 are two independent random variables with respective cumulative distribution functions $F_1(x)$ and $F_2(x)$. 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 Var $(X_1) =$ Var $(X_2) = 1$ and $\mathrm{Cov}(X_1,X_2) = \rho$, with $|\rho| \le 1$. Prove that $p = \mathbb{E}[1 - F_2(\phi(X_1))]$ where F_2 is the cumulative distribution function of X_2 and ϕ a function to be computed.

   Deduce a variance reduction method computing p.

Exercise 13. Let G be a centred Gaussian random variable with unit variance.

1. For $m \in \mathbb{R}$, we set $L^m = \exp\left(-mG - \frac{m^2}{2}\right)$. Show that $\mathbb{E}(L^m f(G+m)) = \mathbb{E}(f(G))$ for each measurable and bounded function $f : \mathbb{R} \to \mathbb{R}$.

Let X^m be an integrable random variable such that $\mathbb{E}(X^m f(G+m)) = \mathbb{E}(f(G))$ for each measurable and bounded function $f : \mathbb{R} \to \mathbb{R}$.

2. Prove that $\mathbb{E}(X^m|G) = L^m$. Is it better to approximate $\mathbb{E}(f(G))$ with the empirical mean corresponding to $\mathbb{E}(X^m f(G+m))$ or the one corresponding to $\mathbb{E}(L^m f(G+m))$?

3. Show that the variance of $L^m f(G+m)$ writes

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

and that it is minimal when $m = \frac{\mathbb{E}(Gf^2(G-m))}{2\mathbb{E}(f^2(G-m))}$. What is the optimal value of m when $f(x) = x$ for each $x \in \mathbb{R}$?

For two positive real numbers p_1 and p_2 with sum equal to 1 and two real numbers m_1 and m_2, we set :

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

Let also $\mu(f) = \mathbb{E}(l(G)f(G))$ for f measurable and bounded.

4. Show that
$$\mu(f) = \int_{\mathbb{R}} f(x)p(x)dx,$$
for some probability density p .

5. Propose how to simulate $\tilde{G}$ according to the density p.

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

7. For $p_1 = p_2 = 1/2$, $m_1 = -m_2 = m$ and $f(x) = x$, show that
$$\text{Var}\,(l^{-1}(\tilde{G})\tilde{G}) = \mathbb{E}\left(\frac{e^{m^2/2}G^2}{\cosh(mG)}\right).$$
Denoting by $v(m)$ this variance, check that $v'(0) = 0$ and $v''(0) < 0$.

How to choose m to reduce the variance when computing $\mathbb{E}(G)$?

Exercise 14. Let X be a square integrable random variable and Y a square integrable control variate such that $\mathbb{E}(Y) = 0$.

1. For $\lambda \in \mathbb{R}$, compute Var $(X - \lambda Y)$ and deduce $\lambda^\star$ which minimizes this variance. What happens when X and Y are independent?

2. Let $((X_i, Y_i), i \geq 1)$ be a sequence i.i.d. according to the distribution of (X, Y) and for $n \geq 1$,
$$\lambda_n^\star = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n}\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum_{i=1}^n X_i)^2}.$$
Show that $\lambda_n^\star$ converges almost surely to $\lambda^\star$ as $n \to +\infty$.

3. Using Slutsky's lemma, show that $\frac{1}{\sqrt{n}}(\lambda^\star - \lambda_n^\star)\sum_{i=1}^n Y_i$ converges to 0 in probability and deduce that
$$\sqrt{n}\left(\frac{1}{n}(X_1 - \lambda_n^\star Y_1 + \cdots + X_n - \lambda_n^\star Y_n) - \mathbb{E}(X)\right)$$
converges in distribution to a centred Gaussian random variable with variance Var $(X - \lambda^\star Y)$.

Chapter 2

Introduction to stochastic algorithms

2.1 A reminder on martingale convergence theorems

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

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

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

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

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 $\mathscr{F}$-martingale (resp. sub, super).

Exercise 15. 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 non-negative super-martingale with respect to $\mathscr{F}$, 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 $a < +\infty$ (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* $\mathscr{F}$*, 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 and 12.

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* non-decreasing *sequence of strictly positive real numbers, such that* $\lim_{n\to+\infty} A_n = +\infty$.

Let $(\varepsilon_k, k \geq 1)$ *be a sequence of real numbers, such that* $S_n = \sum_{k=1}^n \varepsilon_k / A_k$ *converges to some limit* S_∞ *when* n *goes to* $+\infty$*.*

Then :

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

Proof. Under the convention $S_0 = 0$, we have $\varepsilon_k = A_k(S_k - S_{k-1})$ for $k \in \mathbb{N}^*$ so that

$$\sum_{k=1}^n \varepsilon_k = \sum_{k=1}^n A_k(S_k - S_{k-1}) = A_n S_n + \sum_{k=1}^{n-1} A_k S_k - \sum_{\ell=1}^{n-1} A_{\ell+1} S_\ell = A_n S_n - \sum_{k=1}^{n-1} S_k \left(A_{k+1} - A_k\right).$$

As a consequence,

$$\frac{1}{A_n} \sum_{k=1}^n \varepsilon_k = S_n - \frac{1}{A_n} \sum_{k=1}^{n-1} S_k \left(A_{k+1} - A_k\right)$$

where the first term in the right-hand side converges to S_∞ as $n \to \infty$ and $\frac{1}{A_n} \sum_{k=1}^{n-1} S_k \left(A_{k+1} - A_k\right)$ also converges to S_∞ as a (generalized) Cesaro mean of a sequence converging to S_∞.

□

A proof of the strong law of large numbers 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 $\mathscr{F}_n = \sigma(X_k, k \leq n)$ 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 $\mathscr{F}$-martingale. Moreover, using once again independence, we get :

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

So the martingale M is bounded in L^2, and using theorem 2.1.4 converge almost surely to M_∞. Using the Kronecker lemma with $A_k = k$ and $\varepsilon_k = X'_k$, we deduce that almost surely

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} X'_k = 0 \text{ and } \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \mathbb{E}(X).$$

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 16. 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 $\lim_{n\to+\infty} \mathbb{E}(Y_n) = \mathbb{E}(X)$.
2. Prove that $\sum_{n=1}^{+\infty} \mathbb{P}(|X| > n) = \mathbb{E}(\lceil |X| \rceil) - 1$, and deduce that
$$\mathbb{P}(\exists n_0(\omega),\, \forall n \geq n_0,\, X_n = Y_n) = 1.$$
3. Check that $\text{Var}\,(Y_n) \leq \mathbb{E}\left(|X|^2 \mathbf{1}_{\{|X|\leq n\}}\right)$ and prove that :
$$\sum_{n\geq 1} \frac{\text{Var}\,(Y_n)}{n^2} \leq \mathbb{E}\left(|X|^2 f(|X|)\right),$$
where
$$f(z) = \sum_{n\geq \max(1,z)} \frac{1}{n^2} \leq \sum_{n\geq \max(1,z)} \left(\frac{2}{n} - \frac{2}{n+1}\right) \leq \frac{2}{\max(1,z)}.$$
Deduce that $\sum_{n\geq 1} \text{Var}\,(Y_n)/n^2 \leq 2\mathbb{E}\left(|X|\right) < +\infty$.
4. Let $W_n = Y_n - \mathbb{E}(Y_n)$, prove that $\sum_{k\leq n} \frac{W_k}{k}$ converge when n goes to $+\infty$, and deduce, using Kronecker lemma, that
$$a.s.,\ \lim_{n\to+\infty} \frac{1}{n}\sum_{k\leq n} W_k = 0,$$
then deduce that $a.s.$, $\lim_{n\to+\infty} \frac{1}{n}\sum_{k\leq n} Y_n = \mathbb{E}(X)$.
5. Using the result of question 2, prove that $a.s.$, $\lim_{n\to+\infty} \frac{1}{n}\sum_{k\leq n} X_n = \mathbb{E}(X)$

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

Lemma 2.1.4 (Robbins-Sigmund lemma)**.** *Assume $(V_n)_{n\geq 0}$, $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$, $(c_n)_{n\geq 0}$ are sequences of non-negative random variables adapted to $(\mathscr{F}_n, n \geq 0)$ such that for all $n \in \mathbb{N}$, $\mathbb{E}[|V_n| + |b_n|] < +\infty$ and that, moreover, for all $n \geq 0$:*

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

then, on $\left\{\sum_{n\geq 0} a_n < +\infty, \sum_{n\geq 0} b_n < +\infty\right\}$, V_n converges to a random variable V_∞ and $\sum_{n\geq 0} c_n < +\infty$.

Proof. Let :

$$\alpha_n = \frac{1}{\prod_{k=0}^{n}(1+a_k)} \text{ for } n \geq 0 \text{ and } \alpha_{-1} = 1.$$

The non-increasing sequence of $[0,1]$-valued random variables $(\alpha_n)_{n\geq -1}$ converges a.s. to a non-negative limit α_∞. Since

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

$\alpha_\infty > 0$ on $\{\sum_{k\geq 0} a_k < +\infty\}$.

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

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

Since $0 \leq \mathbb{E}\left(V'_{n+1}|\mathscr{F}_n\right)$, $0 \leq V'_n \leq V_n$ and $0 \leq b'_n \leq b_n$, this implies that $c'_n \leq V_n + b_n$ so that c'_n, V'_n and b'_n are integrable. Moreover, the sequence $(M_n = V'_n - \sum_{k=0}^{n-1}(b'_k - c'_k))_{n\geq 0}$ is a supermartingale.

Now consider for $a > 0$ the stopping time τ_a :

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

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

When $\tau_a \geq n$, then $\sum_{k=0}^{n-1}(b'_k - c'_k) \leq a$ and $M_n \geq -a$. The supermartingale $(M_{n\wedge\tau_a})_{n\geq 0}$ being bounded from below by $-a$, it converges a.s. by Remark 2.1.1. Since $\sum_{k=0}^{n}(b'_k - c'_k) \leq \sum_{k\geq 0} b_k$, $\tau_a = +\infty$ when $a \geq \sum_{k\geq 0} b_k$. So we can conclude that $M_\infty = \lim_{n\to+\infty} M_n$ exists a.s. on the set $\left\{\sum_{k\geq 0} b_k < +\infty\right\} \subset \cup_{a\in\mathbb{N}^*}\{\tau_a = +\infty\}$. Since $0 \leq b'_k \leq b_k$ and $0 \leq c'_k$ with

$$\sum_{k=0}^{n-1} c'_k = M_n - V'_n + \sum_{k=0}^{n-1} b'_k \leq M_n + \sum_{k\geq 0} b_k,$$

$\sum_{k\geq 0} b'_k + \sum_{k\geq 0} c'_k < +\infty$ on the set $\left\{\sum_{k\geq 0} b_k < +\infty\right\}$. Hence on this set $V'_\infty = \lim_{n\to+\infty} V'_n$ exists a.s.. On $\{\sum_{k\geq 0} a_k < +\infty, \sum_{k\geq 0} b_k < +\infty\}$, we conclude that V_n converges a.s. to $\frac{V'_\infty}{\alpha_\infty}$ and $\sum_{k\geq 0} c_k \leq \frac{1}{\alpha_\infty}\sum_{k\geq 0} c'_k < +\infty$. □

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 $(\mathscr{F}_n, n \geq 0)$ be a filtration and $(X_n, n \geq 0)$ an $\mathscr{F}_n$-adapted $\mathbb{R}^d$-valued process starting from $X_0 = x_0$, where $x_0 \in \mathbb{R}^d$ and evolving inductively by*

$$\forall n \in \mathbb{N},\ X_{n+1} = X_n - \gamma_n Y_{n+1}.$$

Under

H1 *Hypothesis on the step size sequence:* $(\gamma_n, n \geq 0)$ *is a sequence of positive real numbers such that* $\sum_{n\geq 0} \gamma_n = +\infty$ *and* $\sum_{n\geq 0} \gamma_n^2 < +\infty$.

H2 *Hypothesis on the sequence of random increments:* $(Y_n, n \geq 1)$ *is a sequence of* $\mathbb{R}^d$*-valued random vectors such that for all* $n \in \mathbb{N}$

H2.1 $\mathbb{E}(Y_{n+1}|\mathscr{F}_n) = f(X_n)$ *where* $f : \mathbb{R}^d \to \mathbb{R}^d$ *is continuous and such that*

$$\exists x^\star \in \mathbb{R}^d,\ f(x^\star) = 0 \text{ and } \forall x \in \mathbb{R}^d \setminus \{x^\star\},\ \langle f(x), x - x^\star \rangle > 0.$$

H2.2 $\mathbb{E}\left(|Y_{n+1} - f(X_n)|^2 |\mathscr{F}_n\right) \leq \sigma^2(X_n)$ *where* $s^2(x) = \sigma^2(x) + |f(x)|^2$ *is such that*

$$\exists K < \infty,\ \forall x \in \mathbb{R}^d,\ s^2(x) \leq K(1 + |x - x^\star|^2).$$

Then $\lim_{n\to+\infty} X_n = x^\star$, *a.s.*.

Remark 2.2.1.

- The main application of this algorithm arise when $f(x)$ can be written as

$$f(x) = \mathbb{E}(F(x,U)),$$

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 $\mathscr{F}_n = \sigma(X_0, U_1, \ldots, U_n)$, we have that the sequence $(X_n, n \geq 0)$ is $\mathscr{F}_n$-adapted and

$$\mathbb{E}(F(X_n, U_{n+1})|\mathscr{F}_n) = f(X_n).$$

Moreover

$$\mathbb{E}\left(|F(X_n, U_{n+1}) - f(X_n)|^2 |\mathscr{F}_n\right) = \sigma^2(X_n),$$

where $\sigma^2(x) = \mathbb{E}\left(|F(x,U) - f(x)|^2\right)$. So **H2.2** is an hypothesis on behavior of the expectation and the variance of $F(x,U)$ when $|x|$ goes to $+\infty$.

- The assumption on f in hypothesis **H2.1** is fulfilled when $f(x) = \nabla v(x)$ for a strictly convex C^1 function v minimal at point $x^\star$.

- The non-negativity of $\langle f(x), x - x^\star\rangle$ for each $x \in \mathbb{R}^d$ implies that for $\varepsilon > 0$, $-\langle f(x^\star - \varepsilon f(x^\star)), f(x^\star)\rangle \geq 0$ so that, when f is also continuous at $x^\star$, $|f(x^\star)|^2 = 0$. Hence in **H2.1**, the fact that $f(x^\star) = 0$ is a consequence of the other assumptions made on the function f.

Proof. First note that hypothesis **H2** implies that

$$\begin{aligned}\mathbb{E}\left(|Y_{n+1}|^2|\mathscr{F}_n\right) &= \mathbb{E}\left(|Y_{n+1} - f(X_n)|^2|\mathscr{F}_n\right) + |f(X_n)|^2 \\ &\leq s^2(X_n) \leq K(1 + |X_n - x^\star|^2).\end{aligned} \tag{2.1}$$

Let $V_n = |X_n - x^\star|^2$. Clearly :

$$V_{n+1} = V_n + \gamma_n^2 |Y_{n+1}|^2 - 2\gamma_n \langle X_n - x^\star, Y_{n+1}\rangle.$$

Taking the conditional expectation we obtain :

$$\mathbb{E}(V_{n+1}|\mathscr{F}_n) = V_n + \gamma_n^2 \mathbb{E}\left(|Y_{n+1}|^2|\mathscr{F}_n\right) - 2\gamma_n \langle X_n - x^\star, \mathbb{E}(Y_{n+1}|\mathscr{F}_n)\rangle,$$

and, using inequality (2.1), we deduce :

$$\begin{aligned}\mathbb{E}\left(V_{n+1}|\mathscr{F}_n\right) &\leq V_n + K\gamma_n^2\left(1+V_n\right) - 2\gamma_n\langle X_n - x^\star, f(X_n)\rangle \\ &= V_n\left(1+K\gamma_n^2\right) + K\gamma_n^2 - 2\gamma_n\langle X_n - x^\star, f(X_n)\rangle.\end{aligned} \tag{2.2}$$

Since $\langle X_n - x^\star, f(X_n)\rangle$ is non-negative by hypothesis **H2.1**, we deduce that $\mathbb{E}[V_{n+1}] \leq \left(1+K\gamma_n^2\right)\mathbb{E}[V_n] + K\gamma_n^2$, so that, by an obvious inductive reasoning, V_n is integrable for each $n \in \mathbb{N}$. So, using Robbins-Sigmund lemma with $a_n = b_n = K\gamma_n^2$ and $c_n = \gamma_n\langle X_n - x^\star, f(X_n)\rangle$, we get (by hypothesis **H1**, $\sum_{n\geq 1}\gamma_n^2 < +\infty$) that, both

- V_n converge to V_∞, almost surely,
- $\sum_{n\geq 1}\gamma_n\langle X_n - x^\star, f(X_n)\rangle < +\infty$.

Obviously V_∞ is a non-negative 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 $0 < V_\infty/2 \leq V_n \leq 2V_\infty$ for $n \geq n_0(\omega)$, so

$$\sum_{n\geq 1}\gamma_n\langle X_n - x^\star, f(X_n)\rangle \geq \inf_{V_\infty/2\leq|x-x^\star|^2\leq 2V_\infty}\langle x - x^\star, f(x)\rangle \sum_{n\geq n_0}\gamma_n.$$

But $\sum_{n\geq n_0}\gamma_n = +\infty$ and $\inf_{V_\infty/2\leq|x-x^\star|^2\leq 2V_\infty}\langle x - x^\star, f(x)\rangle > 0$ (remind that f, and so $\langle x - x^\star, f(x)\rangle$, are continuous and $V_\infty/2 \leq |x-x^\star|^2 \leq 2V_\infty$ is a compact set). So on the set $\{V_\infty > 0\}$ we should have $\sum_{n\geq 1}\gamma_n\langle X_n - x^\star, f(X_n)\rangle = +\infty$, but we know that this sum is almost surely finite. So we have proved that $\mathbb{P}(V_\infty > 0) = 0$, which gives 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 there exist $x^\star$ which minimizes ϕ on $\mathbb{R}$ and*

  $$\exists K < +\infty, \forall x\in\mathbb{R}, \left|\phi''(x)\right| \leq K(1+|x-x^\star|) \text{ and } s^2(x) := \mathbb{E}\left(F^2(x,U)\right) \leq K(1+(x-x^\star)^2).$$

- *$(\gamma_n, n\geq 0)$ and $(\delta_n, n\geq 0)$ are bounded sequences of positive numbers such that*

  $$\sum_{n\geq 0}\gamma_n = +\infty, \sum_{n\geq 0}\gamma_n\delta_n < +\infty, \sum_{n\geq 0}\frac{\gamma_n^2}{\delta_n^2} < +\infty,$$

- *$((U_n^1, U_n^2), n \geq 1)$ is a sequence of independent random vectors with U_n^1 and U_n^2 following the law of U.*

We define $(X_n, n \geq 0)$ by $X_0 = x_0 \in \mathbb{R}$ and, inductively

$$X_{n+1} = X_n - \gamma_n \frac{F(X_n + \delta_n, U_{n+1}^1) - F(X_n - \delta_n, U_{n+1}^2)}{2\delta_n}.$$

Then

$$\lim_{n \to +\infty} X_n = x^\star, a.s..$$

Proof. Define $V_n = (X_n - x^\star)^2$, clearly

$$V_{n+1} = V_n + (X_{n+1} - X_n)^2 + 2(X_n - x^\star)(X_{n+1} - X_n).$$

Since

$$(X_{n+1} - X_n)^2 \leq \frac{2\gamma_n^2}{4\delta_n^2}\left(F^2(X_n + \delta_n, U_{n+1}^1) + F^2(X_n - \delta_n, U_{n+1}^2)\right),$$

we have

$$\begin{aligned}\mathbb{E}\left((X_{n+1} - X_n)^2 | \mathscr{F}_n\right) &\leq \frac{\gamma_n^2}{2\delta_n^2}\left(s^2(X_n + \delta_n) + s^2(X_n - \delta_n)\right) \\ &\leq \frac{K\gamma_n^2}{2\delta_n^2}\left(2 + (X_n + \delta_n - x^\star)^2 + (X_n - \delta_n - x^\star)^2\right) \\ &= \frac{K\gamma_n^2}{\delta_n^2}(X_n - x^\star)^2 + K\frac{\gamma_n^2}{\delta_n^2} + K\gamma_n^2,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(V_{n+1}|\mathscr{F}_n) \leq &\left(1 + \frac{K\gamma_n^2}{\delta_n^2}\right)V_n + K\frac{\gamma_n^2}{\delta_n^2} + K\gamma_n^2 \\ &- \frac{\gamma_n}{\delta_n}(X_n - x^\star)\left[\phi(X_n + \delta_n) - \phi(X_n - \delta_n) - 2\delta_n\phi'(X_n)\right] \\ &- 2\gamma_n(X_n - x^\star)\phi'(X_n).\end{aligned} \tag{2.3}$$

Now, by the growth assumption on ϕ'',

$$\left|\phi(x + \delta) - \phi(x - \delta) - 2\delta\phi'(x)\right| = \left|\int_{x-\delta}^{x+\delta}\int_x^y \phi''(z)dzdy\right| \leq K\delta^2(1 + |x - x^\star| + |\delta|).$$

Therefore

$$\begin{aligned}|(X_n - x^\star)\left[\phi(X_n + \delta_n) - \phi(X_n - \delta_n) - 2\delta_n\phi'(X_n)\right]| &\leq K\delta_n^2|X_n - x^\star|(1 + |X_n - x^\star| + \delta_n) \\ &\leq 2K\delta_n^2(X_n - x^\star)^2 + \frac{K(\delta_n^2 + \delta_n^4)}{2}\end{aligned}$$

using that $|X_n - x^\star|(1 + \delta_n) \leq \frac{(X_n - x^\star)^2 + 1}{2} + \frac{(X_n - x^\star)^2 + \delta_n^2}{2}$. Finally we obtain

$$\mathbb{E}(V_{n+1}|\mathscr{F}_n) \leq V_n\left(1 + \frac{K\gamma_n^2}{\delta_n^2} + 2K\gamma_n\delta_n\right) + K\frac{\gamma_n^2}{\delta_n^2} + K\gamma_n^2 + \frac{K\gamma_n(\delta_n + \delta_n^3)}{2} - 2\gamma_n(X_n - x^\star)\phi'(X_n).$$

Note that since $x^\star$ minimizes ϕ, $\phi'(x^\star) = 0$ and, by strict convexity of ϕ, $(x - x^\star)\phi'(x) = (x - x^\star)(\phi'(x) - \phi'(x^\star))$ is positive for $x \in \mathbb{R} \setminus \{x^\star\}$. By continuity, for each $\varepsilon \in (0,1)$, $\inf_{\varepsilon < |x - x^\star| \leq \frac{1}{\varepsilon}}(x - x^\star)\phi'(x) > 0$. Setting

- $a_n = K\frac{\gamma_n^2}{\delta_n^2} + 2K\gamma_n\delta_n$,
- $b_n = K\frac{\gamma_n^2}{\delta_n^2} + K\gamma_n^2 + \frac{K\gamma_n(\delta_n+\delta_n^3)}{2}$,
- $c_n = 2\gamma_n(X_n - x^\star)\phi'(X_n)$,

we have that a_n, b_n, c_n are non-negative, $\sum_{n\geq 1} a_n = K\left(\sum_{n\geq 1}\frac{\gamma_n^2}{\delta_n^2} + 2\sum_{n\geq 1}\gamma_n\delta_n\right) < +\infty$ and since the sequence $(\delta_n)_{n\geq 1}$ is bounded by some finite constant C,

$$\sum_{n\geq 1}\gamma_n^2 \leq C^2\sum_{n\geq 1}\frac{\gamma_n^2}{\delta_n^2} < +\infty \text{ and } \sum_{n\geq 1}\gamma_n\delta_n^3 \leq C^2\sum_{n\geq 1}\gamma_n\delta_n < +\infty$$

so that $\sum_{n\geq 1} b_n < +\infty$. Last $\mathbb{E}[V_{n+1}] \leq \left(1 + \frac{K\gamma_n^2}{\delta_n^2} + 2K\gamma_n\delta_n\right)\mathbb{E}[V_n] + K\frac{\gamma_n^2}{\delta_n^2} + K\gamma_n^2 + \frac{K\gamma_n(\delta_n+\delta_n^3)}{2}$, which, combined with an obvious inductive reasoning, ensures that V_n is integrable for each $n \in \mathbb{N}$. By the Robbins-Sigmund lemma, we obtain that V_n converges 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 suffer from problems of speed of convergence. We will see that these algorithms can lead to central limit theorems (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.

It is easy to show this in a simplified framework for time-steps of the form $\left(\gamma_n = \frac{\alpha}{n+1}\right)_{n\geq 0}$. We assume that

$$F(x,u) = cx + u,$$

where $c > 0$ and that U follows a Gaussian law with mean 0 and variance $\sigma^2 > 0$. 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 converges 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

$$d\left(e^{c\int_0^t \gamma_s ds}X_t\right) = e^{c\int_0^t \gamma_s ds}\left[c\gamma_t X_t dt - c\gamma_t X_t dt - \gamma_t\sigma dW_t\right] = -e^{c\int_0^t \gamma_s ds}\gamma_t\sigma dW_t.$$

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

$$X_t = \frac{x}{(t+1)^{c\alpha}} - \frac{\sigma\alpha}{(t+1)^{c\alpha}} \int_0^t \frac{1}{(s+1)^{1-c\alpha}} dW_s.$$

An easy computation leads to

$$\text{Var}\,(X_t) = \frac{\sigma^2\alpha^2}{(t+1)^{2c\alpha}} \int_0^t \frac{ds}{(s+1)^{2-2c\alpha}} = \frac{\sigma^2\alpha^2}{(t+1)^{2c\alpha}} \times \frac{(t+1)^{2c\alpha-1}-1}{2c\alpha-1}$$

with the last factor equal to $\ln(t+1)$ when $2c\alpha = 1$. Hence $X_t \sim \mathcal{N}_1\left(\frac{x}{(t+1)^{c\alpha}}, \frac{\sigma^2\alpha^2}{(t+1)^{2c\alpha}} \times \frac{(t+1)^{2c\alpha-1}-1}{2c\alpha-1}\right)$.

We can now deduce the asymptotic behavior of X_t as $t \to +\infty$

- if $2c\alpha > 1$, $\sqrt{t}X_t$ converges *in distribution* to a Gaussian random variable distributed according to $\mathcal{N}_1\left(0, \frac{\sigma^2\alpha^2}{2c\alpha-1}\right)$,
- if $2c\alpha < 1$, $t^{c\alpha}X_t$ converges *in distribution* to a Gaussian random variable distributed according to $\mathcal{N}_1\left(x, \frac{\sigma^2\alpha^2}{1-2c\alpha}\right)$, which is worse than the awaited central limit behavior.

We can check on this example (see Exercise 18), that when $2c\alpha < 1$, $t^{c\alpha}X_t$ converges in L^2 to a random variable.

We will see in what follows, that we can fully describe the asymptotic behavior of 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 the proof relies is the central limit theorem for 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 $\gamma_n = \frac{\alpha}{n+1}$ with α large enough, another way is to choose $\gamma_n = \frac{\alpha}{(n+1)^\beta}$, with $1/2 < \beta < 1$. But since $n^{\beta/2}(X_n - x^\star)$ converges in law to some Gaussian random variable, this last choice increases 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 gives 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 that $(M_n, n \geq 0)$ is a square integrable martingale. There exists a unique, predictable, non-decreasing 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

$$\langle M \rangle_{n+1} - \langle M \rangle_n = \mathbb{E}\left((M_{n+1} - M_n)^2 | \mathcal{F}_n\right) = \mathbb{E}\left(M_{n+1}^2 | \mathcal{F}_n\right) - M_n^2.$$

Predictable means here that $\langle M\rangle_n$ is $\mathscr{F}_{n-1}$-measurable.

Proof. If $\langle M\rangle_{n+1}$ is $\mathscr{F}_n$-measurable, then

$$M_n^2 - \langle M\rangle_n = \mathbb{E}\left(M_{n+1}^2 - \langle M\rangle_{n+1}|\mathscr{F}_n\right) \Longleftrightarrow \langle M\rangle_{n+1} - \langle M\rangle_n = \mathbb{E}\left(M_{n+1}^2|\mathscr{F}_n\right) - M_n^2.$$

This implies that $(\langle M\rangle_n = \sum_{k=1}^n (\mathbb{E}\left(M_k^2|\mathscr{F}_{k-1}\right) - M_{k-1}^2), n \geq 0)$ is the unique predictable process equal to 0 at time 0 such that $(M_n^2 - \langle M\rangle_n, n \geq 0)$ is a martingale.
Moreover, using the martingale property of M, one can check that

$$\mathbb{E}\left(M_{n+1}^2|\mathscr{F}_n\right) - M_n^2 = \mathbb{E}\left((M_{n+1} - M_n)^2|\mathscr{F}_n\right) \geq 0,$$

which proves that $\langle M\rangle_n$ is non-decreasing. □

The following theorem relates the almost sure asymptotic behavior of a square integrable martingale M_n to the one of its bracket.

Theorem 2.3.3 (Strong law of large number for martingales). *Let $(M_n, n \geq 0)$ be a square integrable martingale with bracket $(\langle M\rangle_n, n \geq 0)$ and let $\langle M\rangle_\infty = \lim_{n\to+\infty}\langle M\rangle_n$. Then*

- *on $\{\langle M\rangle_\infty < +\infty\}$, M_n converge almost surely to a random variable denoted as M_∞.*
- *on $\{\langle M\rangle_\infty = +\infty\}$,*

$$\lim_{n\to+\infty} \frac{M_n}{\langle M\rangle_n} = 0, \text{ a.s. and, more generally, } \lim_{n\to+\infty} \frac{M_n}{\sqrt{a(\langle M\rangle_n)}} = 0, \text{ a.s.,}$$

as soon as $a(\cdot)$ is a non-negative, non-decreasing function such that $\int_0^{+\infty} \frac{dt}{1+a(t)} < +\infty$.

Remark 2.3.1. The first statement is an extension of the a.s. convergence property of martingales bounded in L^2 and its proof relies on this result. Indeed if $(M_n, n \geq 0)$ is a martingale bounded in L^2, then $\mathbb{E}(\langle M\rangle_n) = \mathbb{E}(M_n^2) - \mathbb{E}(M_0^2) \leq \sup_{k\in\mathbb{N}} \mathbb{E}(M_k^2)$. By monotone convergence, one deduces that $\mathbb{E}(\langle M\rangle_\infty) \leq \sup_{k\in\mathbb{N}} \mathbb{E}(M_k^2) < +\infty$ so that $\mathbb{P}(\langle M\rangle_\infty < \infty) = 1$.

Proof. The random variable

$$\tau_p = \inf\{n \geq 0, \langle M\rangle_{n+1} > p\}$$

is a stopping time as $\langle M\rangle$ is predictable. So $M_n^{(p)} := M_{n\wedge\tau_p}$ and $M_{n\wedge\tau_p}^2 - \langle M\rangle_{n\wedge\tau_p}$ are also martingales . Note that, by definition, $\langle M\rangle_{n\wedge\tau_p} \leq \langle M\rangle_{\tau_p} \leq p$.

Now remark that, for all $n \geq 0$

$$\mathbb{E}\left((M_n^{(p)})^2\right) = \mathbb{E}(M_0^2) + \mathbb{E}\left(\langle M\rangle_{n\wedge\tau_p}\right) \leq \mathbb{E}(M_0^2) + p.$$

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 $\{\langle M\rangle_\infty \leq p\} = \{\tau_p = +\infty\}$, so

$$\{\langle M\rangle_\infty < +\infty\} = \cup_{p\geq 0}\{\langle M\rangle_\infty \leq p\} = \cup_{p\geq 0}\{\tau_p = +\infty\}.$$

So, on the set $\{\langle M\rangle_\infty < +\infty\}$, M_n converge to M_∞, which proves the first point.

For the second one, we will consider the process

$$N_n = \sum_{k=1}^{n} \frac{M_k - M_{k-1}}{\sqrt{1+a(\langle M\rangle_k)}}.$$

Since $\sqrt{1+a(\langle M\rangle_k)} \geq 1$ and $M_k - M_{k-1}$ is square integrable, so are $\frac{M_k-M_{k-1}}{\sqrt{1+a(\langle M\rangle_k)}}$ and, in turn, N_n. Since $\langle M\rangle$ is predictable,

$$\mathbb{E}\left(\frac{M_k - M_{k-1}}{\sqrt{1+a(\langle M\rangle_k)}}\middle|\mathscr{F}_{k-1}\right) = \frac{1}{\sqrt{1+a(\langle M\rangle_k)}}\mathbb{E}\left(M_k - M_{k-1}|\mathscr{F}_{k-1}\right) = 0$$

and $(N_n, n \geq 0)$ is a square integrable martingale. We can compute its bracket :

$$\langle N\rangle_{n+1} - \langle N\rangle_n = \mathbb{E}\left(\frac{(M_{n+1}-M_n)^2}{1+a(\langle M\rangle_{n+1})}\middle|\mathscr{F}_n\right) = \frac{\langle M\rangle_{n+1} - \langle M\rangle_n}{1+a(\langle M\rangle_{n+1})}.$$

Hence, with the monotonicity of a,

$$\langle N\rangle_n = \sum_{k=1}^{n} \frac{\langle M\rangle_k - \langle M\rangle_{k-1}}{1+a(\langle M\rangle_k)} \leq \sum_{k=1}^{n} \int_{\langle M\rangle_{k-1}}^{\langle M\rangle_k} \frac{dt}{1+a(t)} \leq \int_0^{+\infty} \frac{dt}{1+a(t)} < +\infty.$$

So $\langle N\rangle_\infty < +\infty$, a.s., and, using first part of this theorem, N_n converge a.s. to N_∞. The monotonicity of a combined with $\int_0^{+\infty} \frac{dt}{1+a(t)} < +\infty$ implies that $\lim_{t\to+\infty} a(t) = +\infty$ so that on $\{\langle M\rangle_\infty = +\infty\}$, $\lim_{k\to+\infty} a(\langle M\rangle_k) = +\infty$.
Using Kronecker's lemma with $\varepsilon_k = M_k - M_{k-1}$ and $A_k = \sqrt{1+a(\langle M\rangle_k)}$, we conclude that, on $\{\langle M\rangle_\infty = +\infty\}$, $\lim_{n\to+\infty} \frac{M_n - M_0}{\sqrt{1+a(\langle M\rangle_n)}} = 0 = \lim_{n\to+\infty} \frac{M_n}{\sqrt{a(\langle M\rangle_n)}}$. The first case is obtained for the choice $a(t) = t^2$. □

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, $\mathbb{E}\left((S_{n+1} - S_n)^2|\mathscr{F}_n\right) = \mathbb{E}((X'_{n+1})^2|\mathscr{F}_n) = \text{Var}\,(X)$, $\langle S\rangle_n = n\text{Var}\,(X)$. So $\langle S\rangle_\infty = \infty$ and using the previous theorem we get $\lim_{n\to+\infty} S_n/n = 0$, which gives the strong law of large numbers.

Moreover using $a(t) = t^{1+\varepsilon}$, with $\varepsilon > 0$, we get

$$\lim_{n\to+\infty} \frac{1}{n^{\varepsilon/2}}\sqrt{n}\left\{\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X)\right\} = 0, a.s..$$

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 (resp. a locally square integrable martingale), if there exists a sequence of stopping times $(\tau_p, p \geq 0)$ such that*

- *τ_p is non-decreasing in p and, a.s., goes to ∞ when p goes to ∞.*
- *$(M_{n\wedge\tau_p}, n \geq 0)$ is a martingale (resp. a martingale bounded in L^2) for each $p \geq 0$.*

Such a sequence is called a localizing sequence of stopping times for the local martingale (resp. locally square integrable martingale) $(M_n, n \geq 0)$.

Proposition 2.3.2. *If $(M_n, n \geq 0)$ is a locally square integrable martingale, there exists a unique $\mathscr{F}_{n-1}$-adapted non-decreasing 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. This process is $(\langle M\rangle_n = \sum_{k=1}^n \mathbb{E}\left[(M_k - M_{k-1})^2|\mathscr{F}_{k-1}\right], n \geq 0)$. If σ is a stopping time such that $(M_n^\sigma = M_{n\wedge\sigma}, n \geq 0)$ is a square integrable martingale, then $\langle M^\sigma\rangle_n = \langle M\rangle_{n\wedge\sigma}$.*

Proof. The process $\langle M\rangle_n = \sum_{k=1}^n \mathbb{E}\left[(M_k - M_{k-1})^2|\mathscr{F}_{k-1}\right]$ is predictable and non-decreasing (and may take the value $+\infty$ with positive probability). For a stopping time σ such that $(M_n^\sigma = M_{n\wedge\sigma}, n \geq 0)$ is a square integrable martingale, using that $\{\sigma \geq k\} \in \mathscr{F}_{k-1}$ and $\mathbf{1}_{\{\sigma\geq k\}}(M_k - M_{k-1})^2 = (M_k^\sigma - M_{k-1}^\sigma)^2$, we have

$$\begin{aligned}\langle M\rangle_{n\wedge\sigma} &= \sum_{k=1}^n \mathbf{1}_{\{\sigma\geq k\}}\mathbb{E}\left[(M_k - M_{k-1})^2|\mathscr{F}_{k-1}\right] = \sum_{k=1}^n \mathbb{E}\left[\mathbf{1}_{\{\sigma\geq k\}}(M_k - M_{k-1})^2|\mathscr{F}_{k-1}\right]\\ &= \sum_{k=1}^n \mathbb{E}\left[(M_k^\sigma - M_{k-1}^\sigma)^2|\mathscr{F}_{k-1}\right] = \langle M^\sigma\rangle_n.\end{aligned}$$

Let now $(\tau_p, p \geq 0)$ be a non-decreasing sequence of stopping times going to $+\infty$ with p such that for each $p \geq 0$, $(M_n^{\tau_p} = M_{n\wedge\tau_p}, n \geq 0)$ is a martingale bounded in L_2. We have

$$\mathbb{P}(\langle M\rangle_n < +\infty) \geq \mathbb{P}(\langle M\rangle_n = \langle M^{\tau_p}\rangle_n) \geq \mathbb{P}(\tau_p \geq n) \overset{p\to\infty}{\longrightarrow} 1.$$

Moreover $(M_{n\wedge\tau_p}^2 - \langle M\rangle_{n\wedge\tau_p} = (M_n^{\tau_p})^2 - \langle M^{\tau_p}\rangle_n, n \geq 0)$ is a martingale for each $p \geq 0$, so that $(M_n^2 - \langle M\rangle_n, n \geq 0)$ is a local martingale.

To check the uniqueness of the bracket of $(M_n, n \geq 0)$, we suppose that $(A_n, n \geq 0)$ is a non-decreasing predictable process such that $A_0 = 0$ and $(M_n^2 - A_n, n \geq 0)$ is a local martingale. Let $(\sigma_p, p \geq 0)$ be a non-decreasing sequence of stopping times going to $+\infty$ with p such that $(M_{n\wedge\sigma_p}^2 - A_{n\wedge\sigma_p}, n \geq 0)$ is a martingale. By uniqueness of the bracket of the square integrable martingale $(M_n^{\sigma_p\wedge\tau_p} = M_{n\wedge\sigma_p\wedge\tau_p}, n \geq 0)$ and since $(M_{n\wedge\sigma_p\wedge\tau_p}^2 - A_{n\wedge\sigma_p\wedge\tau_p}, n \geq 0)$ is a martingale, $A_{n\wedge\sigma_p\wedge\tau_p} = \langle M^{\sigma_p\wedge\tau_p}\rangle_n = \langle M\rangle_{n\wedge\sigma_p\wedge\tau_p}$ and uniqueness follows since $\sigma_p \wedge \tau_p$ goes to $+\infty$ with p. □

The strong law of large number remains true and unchanged for martingales locally in L^2.

Theorem 2.3.5. *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 non-negative, non-decreasing 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. Let $(\sigma_q, q \geq 0)$ be a localizing sequence of stopping times for the locally square integrable martingale $(M_n, n \geq 0)$ and $\tau_p = \inf\{n \geq 0, \langle M \rangle_{n+1} > p\}$. One has

$$\mathbb{E}(M^2_{n\wedge\tau_p} 1_{\{\sigma_q \geq n\wedge\tau_p\}}) \leq \mathbb{E}(M^2_{n\wedge\tau_p\wedge\sigma_q}) \leq \mathbb{E}(M_0^2) + \mathbb{E}(\langle M \rangle_{n\wedge\tau_p\wedge\sigma_q}) \leq \mathbb{E}(M_0^2) + p.$$

Since, by monotone convergence, $\mathbb{E}(M^2_{n\wedge\tau_p} 1_{\{\sigma_q \geq n\wedge\tau_p\}})$ converges to $\mathbb{E}(M^2_{n\wedge\tau_p})$ as $q \to \infty$, one concludes that $\mathbb{E}(M^2_{n\wedge\tau_p}) \leq \mathbb{E}(M_0^2) + p$ so that $(M_{n\wedge\tau_p}, n \geq 0)$ is a martingale bounded in L^2. Then we check the first assertion like in the proof of Theorem 2.3.3 dedicated to the square integrable case. Note that $(\tau_p, p \geq 0)$ is a localizing sequence and any locally square integrable martingale admits a localizing sequence such that the corresponding stopped martingales are bounded in L^2.
The second assertion is proved along the same lines as in Theorem 2.3.3 by replacing the first assertion of this theorem by the current one, once we check that $\left(N_n = \sum_{k=1}^n \frac{M_k - M_{k-1}}{\sqrt{1+a(\langle M \rangle_k)}}, n \geq 0\right)$ is a locally square integrable martingale. This follows from the estimation

$$\begin{aligned}
\mathbb{E}(N^2_{n\wedge\sigma_q}) &= \mathbb{E}\left(\left(\sum_{k=1}^n \frac{M_{k\wedge\sigma_q} - M_{(k-1)\wedge\sigma_q}}{\sqrt{1+a(\langle M \rangle_k)}}\right)^2\right) \\
&= \sum_{k=1}^n \mathbb{E}\left(\frac{(M_{k\wedge\sigma_q} - M_{(k-1)\wedge\sigma_q})^2}{1+a(\langle M \rangle_k)}\right) \\
&+ 2\sum_{1\leq\ell\leq k\leq n} \mathbb{E}\left(\frac{M_{\ell\wedge\sigma_q} - M_{(\ell-1)\wedge\sigma_q}}{\sqrt{(1+a(\langle M \rangle_\ell))(1+a(\langle M \rangle_k))}} \mathbb{E}\left(M_{k\wedge\sigma_q} - M_{(k-1)\wedge\sigma_q} | \mathscr{F}_{k-1}\right)\right) \\
&= \sum_{k=1}^n \mathbb{E}\left(\frac{(M_{k\wedge\sigma_q} - M_{(k-1)\wedge\sigma_q})^2}{1+a(\langle M \rangle_k)}\right) \leq \sum_{k=1}^n \mathbb{E}\left((M_{k\wedge\sigma_q} - M_{(k-1)\wedge\sigma_q})^2\right) \\
&= \sum_{k=1}^n \left(\mathbb{E}\left(M^2_{k\wedge\sigma_q}\right) - \mathbb{E}\left(M^2_{(k-1)\wedge\sigma_q}\right)\right) = \mathbb{E}(M^2_{n\wedge\sigma_q}) - \mathbb{E}(M_0^2) < +\infty.
\end{aligned}$$

□

Exercise 17. Let $(X_n, n \geq 1)$ be a sequence of independent real random variables following the law of X, such that $\mathbb{P}(X = \pm 1) = 1/2$ and by $(\lambda_n, n \geq 1)$ a sequence of random variables independant of the sequence $(X_n, n \geq 1)$. Denote by $\mathscr{F}_n = \sigma(X_1, \ldots, X_n, \lambda_0, \ldots, \lambda_{n-1})$. Define M_n by

$$M_n = \sum_{k=0}^{n-1} \lambda_k X_{k+1}.$$

1. Prove that M is an $\mathscr{F}_n$-martingale if and only if $\mathbb{E}(|\lambda_k|) < +\infty$, for all $k \geq 0$.

2. Prove that M is a L^2-martingale if and only if, for all $k \geq 0$ $\mathbb{E}\left(\lambda_k^2\right) < +\infty$.

3. Prove that M is bounded in L^2 if and only if $\sum_{k=0}^{+\infty} \mathbb{E}\left(\lambda_k^2\right) < +\infty$.

4. Prove that M is a locally L^2 martingale if and only if, for all $k \geq 0$ $|\lambda_k| < +\infty$.

5. Give an example of a martingale locally in L^2 which is not a martingale.

2.3.3 Central limit theorem for martingales

We begin by stating the central limit theorem for martingales.

Theorem 2.3.6. *Let $(M_n, n \geq 0)$ be a locally in L^2 martingale and $a(n)$ be a sequence of 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.* (2.4)

Lindeberg condition : *for all* $\varepsilon > 0$,

$$\lim_{n\to+\infty} \frac{1}{a(n)} \sum_{k=1}^{n} \mathbb{E}\left((\Delta M_k)^2 \mathbf{1}_{\left\{|\Delta M_k| \geq \varepsilon\sqrt{a(n)}\right\}} \middle| \mathscr{F}_{k-1} \right) = 0, \text{ in probability.} \quad (2.5)$$

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 2.3.3. 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 which ensures that no increment of M is too large to prevent the central limit behaviour.

Exercise 19 gives a proof in a simple case in which the role of the martingale hypothese is clearer than in the detailled proof which is given page 57. For a complete discussion on martingale convergence theorem, we refer to [Hall and Heyde(1980)].

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 $\langle M \rangle_n = n\text{Var}(X)$ (so obviously $\langle M \rangle_n / n$ converge to $\text{Var}(X) = \sigma^2$!). It remains to check the Lindeberg condition. We have

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(X_k^2 \mathbf{1}_{\{|X_k| \geq \varepsilon\sqrt{n}\}} \middle| \mathscr{F}_{k-1} \right) = \mathbb{E}\left(X^2 \mathbf{1}_{\{|X| \geq \varepsilon\sqrt{n}\}} \right).$$

where the right-hand side converges to 0 when n goes to ∞ by Lebesgue's theorem since X^2 is integrable.

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 2.3.7. *We consider a sequence $(U_n, n \geq 1)$ of i.i.d. $\mathbb{R}^p$-valued random vectors, a function $F : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ and set $\gamma_n = \frac{\alpha}{n+1}$ for $n \in \mathbb{N}$. We define $(X_n, n \in \mathbb{N})$ by*

$$X_{n+1} = X_n - \gamma_n F(X_n, U_{n+1}),\ X_0 = x \in \mathbb{R}.$$

We assume that

- *$\forall x \in \mathbb{R}$, $F(x,U_1)$ is integrable and $f(x) = \mathbb{E}(F(x,U_1))$ is C^2 on $\mathbb{R}$.*
- *$f(x^\star) = 0$ and $\langle f(x), x - x^\star \rangle > 0$, for $x \neq x^*$.*
- *$f'(x^\star) = c$, where $c > 0$.*
- *$\sigma^2(x) :=$ Var $(F(x,U_1))$ is continuous at $x^\star$ and*

$$\exists K < \infty,\ \forall x \in \mathbb{R},\ s^2(x) := \sigma^2(x) + f^2(x) \leq K\left(1 + |x - x^\star|^2\right).$$

- *It exists $\eta > 0$ such that, for all $x \in \mathbb{R}$*

$$v^{2+\eta}(x) := \mathbb{E}\left(|F(x,U_1) - f(x)|^{2+\eta}\right) < +\infty,$$

and $\sup_{n\geq 0} v^{2+\eta}(X_n) < +\infty$ a.s..

Then X_n converges almost surely to x^* and

- *if $c\alpha > 1/2$, $\sqrt{n}(X_n - x^\star)$ converges* in distribution *to a zero mean Gaussian random variable with variance $\alpha^2\sigma^2(x^\star)/(2c\alpha - 1)$*
- *if $c\alpha < 1/2$, $n^{c\alpha}(X_n - x^\star)$ converges* almost surely *to a finite random variable.*

Remark 2.3.4.

- For $F(x,u) = cx + u$ and $U \sim \mathcal{N}_1(0,\sigma^2)$, $\sigma^2(x) = \sigma^2$ so that $\alpha^2\sigma^2(x^\star)/(2c\alpha - 1) = \alpha^2\sigma^2/(2c\alpha - 1)$ and the asymptotic variance matches the one obtained by formal computations in Paragraph 2.3.1.
- It is easy to optimize over $\alpha \in]\frac{1}{2c}, +\infty[$ the asymptotic variance $g(\alpha) = \frac{\alpha^2\sigma^2(x^\star)}{2c\alpha-1}$. Indeed $\frac{(2c\alpha-1)^2}{\sigma^2(x^\star)} g'(\alpha) = 2\alpha(c\alpha - 1)$ so that the derivative is negative on $]\frac{1}{2c}, \frac{1}{c}[$ and positive on $]\frac{1}{c}, +\infty[$. The optimal choice is thus $g(1/c) = \sigma^2(x^\star)/c^2$.
- The same type of CLT can be obtained when $\gamma_n = \frac{\alpha}{(n+1)^\beta}$, with $1/2 < \beta < 1$. In this case it can be proved that, for every α, $(X_n - x^\star)/\sqrt{\gamma_n}$ converges in distribution to a Gaussian random variable, whatever the value of α. Since $1/\sqrt{\gamma_n} = \frac{(n+1)^{\beta/2}}{\sqrt{\alpha}}$, the order of convergence is increasing with β.

One has

$$X_{n+1} - x^\star = X_n - x^\star - \gamma_n F(X_n - x^\star + x^\star, U_{n+1}) \text{ for } n \in \mathbb{N}.$$

Hence replacing $((X_n, n \geq 0), F(x,u), f(x))$ by $((X_n - x^\star, n \geq 0), F(x + x^\star, u), f(x + x^\star))$, we may suppose that $x^\star = 0$. Our algorithm writes as

$$\begin{aligned} X_{n+1} &= X_n - \gamma_n F(X_n, U_{n+1}) \\ &= X_n(1 - c\gamma_n \mathbf{1}_{\{n \geq c\alpha\}}) + \gamma_n\left[f(X_n) - F(X_n, U_{n+1})\right] + \gamma_n\left[c\mathbf{1}_{\{n\geq c\alpha\}} X_n - f(X_n)\right]. \end{aligned}$$

Now, denote

- $\Delta N_{n+1} = f(X_n) - F(X_n, U_{n+1})$ (ΔN_{n+1} is a martingale increment for the filtration $\mathcal{F}_n = \sigma(U_1, \cdots, U_n)$),

- $R_n = c\mathbf{1}_{\{n\geq c\alpha\}}X_n - f(X_n)$ (since $f(x^\star)=0$, $f'(x^\star)=c$ and $x^\star=0$, for $n\geq c\alpha$, $R_n = f(x^\star)+f'(x^\star)(X_n-x^\star)-f(X_n)$ will be of order $(X_n-x^\star)^2$ as f is C^2),
- $\alpha_n = 1-c\gamma_n\mathbf{1}_{\{n\geq c\alpha\}}$ and $\beta_n=\prod_{k=0}^n \alpha_k$ with convention $\beta_{-1}=1$.

With these notations

$$\frac{X_{n+1}}{\beta_n} = \frac{X_n}{\beta_{n-1}} + \frac{\gamma_n}{\beta_n}\Delta N_{n+1} + \frac{\gamma_n}{\beta_n}R_n,$$

so

$$X_n = \beta_{n-1}\left(X_0 + M_n + \sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k\right) \text{ where } M_n = \sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}\Delta N_{k+1}. \tag{2.6}$$

The main point to obtain the asymptotic behavior of X_n is now to estimate the bracket of the martingale M_n. Clearly

$$\langle M\rangle_n = \sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_k),$$

and we already know from Theorem 2.2.1 that X_k converges to $x^\star=0$ as $k\to\infty$, so, by continuity of σ, $\lim_{k\to+\infty}\sigma^2(X_k)=\sigma^2(x^\star)$ a.s.. In what follows, we suppose that $\boxed{\sigma^2(x^\star)>0}$ so that we deduce from the next lemma that as $k\to\infty$, $\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_k)\sim\frac{\alpha^2\sigma^2(x^\star)}{C^2}k^{2(c\alpha-1)}$ a.s..

Lemma 2.3.5. *Let c,α be positive numbers . Assume that $\gamma_n=\frac{\alpha}{n+1}$.*
Define $\alpha_n = 1-c\gamma_n\mathbf{1}_{\{n\geq c\alpha\}}$, $\beta_n=\prod_{k=0}^n\alpha_k$, then there exists a 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 of Lemma 2.3.5. The function $h(y)=y+\ln(1-y)$ is such that $h'(y)=1-\frac{1}{1-y}$, $h''(y)=-\frac{1}{(1-y)^2}$ and $h(0)=h'(0)=0$. Hence for $x\in[0,1)$, $h(x)=\int_0^x(x-y)h''(y)dy$ and

$$0\geq h(x)\geq -\frac{1}{(1-x)^2}\int_0^x(x-y)dy = -\frac{x^2}{2(1-x)^2}.$$

So if $n\geq 2c\alpha$, $c\gamma_n\leq 1/2$ and

$$0\geq \ln(1-c\gamma_n)+c\gamma_n\geq -2c^2\gamma_n^2.$$

Hence $\sum_{k=0}^n(\ln(1-c\gamma_k\mathbf{1}_{\{k\geq c\alpha\}})+c\gamma_k\mathbf{1}_{\{k\geq c\alpha\}})$ converges as $n\to\infty$ and so does

$$\sum_{k=0}^n\ln(1-c\gamma_k\mathbf{1}_{\{k\geq c\alpha\}})+c\alpha\ln n = \sum_{k=0}^n(\ln(1-c\gamma_k\mathbf{1}_{\{k\geq c\alpha\}})+c\gamma_k\mathbf{1}_{\{k\geq c\alpha\}})+c\alpha\left(\ln n-\sum_{k=\lceil c\alpha\rceil}^n\frac{1}{k+1}\right)$$

since $\ln n-\sum_{k=\lceil c\alpha\rceil}^n\frac{1}{k+1}=\sum_{\ell=1}^{\lceil c\alpha\rceil}\frac{1}{\ell}-\left(\sum_{\ell=1}^{n+1}\frac{1}{\ell}-\ln n\right)$ converges to $\sum_{\ell=1}^{\lceil c\alpha\rceil}\frac{1}{\ell}$ minus the Euler constant as $n\to\infty$. By continuity of the exponential function, we deduce that

$$\lim_{n\to+\infty}\beta_n n^{\alpha c} = C>0.$$

and since $\lim_{n\to\infty}\gamma_n n=\alpha$, we deduce that $\lim_{n\to+\infty}(\gamma_n/\beta_n)n^{1-c\alpha}=\frac{\alpha}{C}$. □

The case when $2c\alpha > 1$. In this case we are interested in the convergence in distribution of $\sqrt{n}(X_n - x^\star)$ to a Gaussian random variable. The main step will be to apply the central limit theorem for the martingale M. So we have to get an asymptotic estimate for its bracket $\langle M\rangle_n$.

Since $\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_k) \sim \frac{\alpha^2\sigma^2(x^\star)}{C^2}k^{2(c\alpha-1)}$ with $2(c\alpha-1) > -1$, the series diverges and by a comparison with integrals, we deduce that

$$\langle M\rangle_n = \sum_{k=0}^{n-1}\frac{\gamma_k^2}{\beta_k^2}\sigma^2(X_k) \sim \frac{\alpha^2\sigma^2(x^\star)}{C^2}\int_0^n y^{2(c\alpha-1)}dy = \frac{\alpha^2\sigma^2(x^\star)}{C^2} \times \frac{n^{2c\alpha-1}}{2c\alpha-1} \text{ as } n\to\infty \text{ a.s..}$$

We have

$$\sqrt{n}\beta_{n-1}M_n = \frac{n^{c\alpha}\beta_{n-1}}{C} \times \frac{M_n}{\sqrt{a(n)}} \tag{2.7}$$

where $a(n) = n^{2c\alpha-1}/C^2$ and $\lim_{n\to\infty}\frac{n^{c\alpha}\beta_{n-1}}{C} = 1$ by Lemma 2.3.5. Moreover, $\lim_{n\to\infty}a(n) = +\infty$ since $2c\alpha - 1 > 0$ and

$$\lim_{n\to+\infty}\frac{\langle M\rangle_n}{a(n)} = \frac{\alpha^2\sigma^2(x^\star)}{2c\alpha-1}, \text{a.s..}$$

So the bracket condition needed to apply the central limit theorem to the martingale $(M_n, n\geq 0)$ is satisfied. If the Lindeberg condition is also satisfied, then $\frac{M_n}{\sqrt{a(n)}}$ converges in distribution to a centered Gaussian random variable with variance $\frac{\alpha^2\sigma^2(x^\star)}{2c\alpha-1}$. By (2.7), $\sqrt{n}\beta_{n-1}M_n$ also converges in distribution to this Gaussian random variable. Since, by (2.6),

$$\sqrt{n}X_n = \sqrt{n}\beta_{n-1}X_0 + \sqrt{n}\beta_{n-1}M_n + \sqrt{n}\beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k,$$

to conclude (using Slutsky's lemma) that $\sqrt{n}X_n$ converges in distribution to the same random variable, it is enough to prove that $\sqrt{n}\beta_{n-1}X_0$ converges to 0 in probability (the almost sure convergence to 0 even holds since $\sqrt{n}\beta_n \sim Cn^{\frac{1}{2}-c\alpha}$ as $n\to\infty$ with $\frac{1}{2} - c\alpha < 0$) and that $\sqrt{n}\beta_{n-1}\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k$ also converges to 0 in probability. This part is heavily technical and will not be proved here.

Let us check the Lindeberg condition. Since $\mathbf{1}_{\left\{|\Delta N_{k+1}|\geq\varepsilon\sqrt{a(n)}\right\}} \leq \frac{|\Delta N_{k+1}|^\eta}{(\varepsilon\sqrt{a(n)})^\eta}$, we have

$$\begin{aligned}
&\frac{1}{a(n)}\sum_{k=0}^{n-1}\mathbb{E}\left(|\Delta N_{k+1}|^2\mathbf{1}_{\left\{|\Delta N_{k+1}|\geq\varepsilon\sqrt{a(n)}\right\}}\middle|\mathscr{F}_k\right)\\
&\quad\leq \frac{1}{a(n)}\sum_{k=0}^{n-1}\frac{1}{(\varepsilon\sqrt{a(n)})^\eta}\mathbb{E}\left(|\Delta N_{k+1}|^{2+\eta}\middle|\mathscr{F}_k\right)\\
&\quad= \frac{1}{\varepsilon^\eta a(n)^{1+\eta/2}}\sum_{k=0}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta}\mathbb{E}\left(|F(X_k,U_{k+1}) - f(X_k)|^{2+\eta}\middle|\mathscr{F}_k\right)\\
&\quad\leq \frac{L(\omega)}{\varepsilon^\eta a(n)^{1+\eta/2}}\sum_{k=0}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta},
\end{aligned}$$

where $L(\omega) = \sup_{k\geq 0}\nu^{2+\eta}(X_k)$ (which is supposed to be a.s. finite by hypothesis). By Lemma 2.3.5, we have $\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta} \sim \left(\frac{\alpha}{C}\right)^{2+\eta}k^{(c\alpha-1)(2+\eta)}$ a.s. as $k\to\infty$. We want that $(c\alpha-1)(2+\eta) >$

-1 to apply the comparison with integrals to $\sum_{k=0}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta}$. When $c\alpha \geq 1$ this is satisfied. Otherwise this is equivalent to $\eta < \frac{1}{1-c\alpha} - 2$ where the right-hand side is positive since $c\alpha > \frac{1}{2}$ so that, up to possibly decreasing η (which makes the hypotheses weaker), we may suppose that the inequality is satisfied. Hence

$$\sum_{k=0}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta} \sim \frac{\alpha^{2+\eta}}{C^{2+\eta}((c\alpha-1)(2+\eta)+1)} n^{(c\alpha-1)(2+\eta)+1} \text{ as } n \to \infty.$$

With $a(n) = n^{2c\alpha-1}/C^2$, we get

$$\begin{aligned}\frac{1}{a(n)}\sum_{k=0}^{n-1}\mathbb{E}\left(|\Delta N_{k+1}|^2 \mathbf{1}_{\left\{|\Delta N_{k+1}|\geq \varepsilon\sqrt{a(n)}\right\}} \,\middle|\, \mathscr{F}_k\right) \\ &\leq \frac{L(\omega)}{\varepsilon^\eta} C^{2+\eta} n^{-(c\alpha-1/2)(2+\eta)} \sum_{k=0}^{n-1}\left(\frac{\gamma_k}{\beta_k}\right)^{2+\eta} \\ &\sim \frac{L(\omega)\alpha^{2+\eta}}{\varepsilon^\eta((c\alpha-1)(2+\eta)+1)} n^{-\eta/2}.\end{aligned}$$

And this ends the proof that the Lindeberg condition is fulfilled.

The case when $2c\alpha < 1$**.** Since, by (2.6), $n^{c\alpha}(X_n - x^\star) = n^{c\alpha}\beta_{n-1}\left(X_0 + M_n + \sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k\right)$, where, by Lemma 2.3.5, $n^{c\alpha}\beta_{n-1} \to C$ as $n \to \infty$, it is enough to check that M_n and $\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k$ converge a.s. as $n \to \infty$. For this, we first check that

$$\text{a.s.,}\ \langle M\rangle_n - \langle M\rangle_{n-1} = \frac{\gamma_{n-1}^2\sigma^2(X_{n-1})}{\beta_{n-1}^2} \sim \frac{\alpha^2\sigma^2(x^\star)}{C^2} n^{2c\alpha-2} \text{ as } n \to \infty.$$

Since $2c\alpha - 2 < -1$, $\langle M\rangle_\infty < \infty$ a.s. and, by the strong law of large numbers for martingales (see Theorem 2.3.5), M_n converges a.s. to M_∞. As $|R_k| \leq K|X_k|^2$, we have

$$\mathbb{E}\left(\left|\frac{\gamma_k}{\beta_k}R_k\right|\right) \leq K\frac{\gamma_k}{\beta_k}\mathbb{E}(|X_k|^2) \sim \frac{\alpha K}{C} k^{c\alpha-1}\mathbb{E}(|X_k|^2) \text{ as } k \to \infty.$$

So if we can prove that there exists $\beta > c\alpha$ such that $\sup_k k^\beta \mathbb{E}(|X_k|^2) < +\infty$, we will deduce the absolute convergence of $\sum_{k=0}^{n-1}\frac{\gamma_k}{\beta_k}R_k$. For a proof of this point we refer to [Duflo(1997)].

2.4 Exercises and problems

Exercise 18. We are interested in at the solution of

$$dX_t = -\gamma_t\left(cX_t dt + \sigma dW_t\right), X_0 = x. \tag{2.8}$$

where $\gamma_t = \frac{\alpha}{t+1}$, and $(W_t, t \geq 0)$ is a Brownian motion and c, σ and α are positive real numbers.

1. Prove that
$$X_t = \frac{x}{(t+1)^{c\alpha}} - \frac{\sigma\alpha}{(t+1)^{c\alpha}} \int_0^t \frac{1}{(s+1)^{1-c\alpha}} dW_s$$
and that X_t is a Gaussian random variable.

2. When $2c\alpha > 1$, prove that $\sqrt{(t+1)}X_t$ converges *in distribution* to a Gaussian random variable with mean 0 and variance $\frac{\sigma^2\alpha^2}{(2c\alpha-1)}$.

3. When $2c\alpha < 1$, prove that $(t+1)^{c\alpha}X_t$ converges in L^2 to some random variable distributed according to $\mathcal{N}(x, \frac{\sigma^2\alpha^2}{1-2c\alpha})$ (first check that the Cauchy criterion equivalent to convergence in the Hilbert space L^2 is satisfied).

Exercise 19. In this exercise, we prove the central limit theorem 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 $\Delta M_n = M_n - M_{n-1}$ and K is a constant. M is a square integrable martingale (why?) and, so, we can denote by $\langle M \rangle$ its bracket. Assume moreover that

$$\lim_{n\to+\infty} \frac{\langle M \rangle_n}{n} = \sigma^2, \text{a.s.} \tag{2.9}$$

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. We want to extend $\phi_j(z)$ to z a complex numbers. For this, we define the complex logarithm around 1 as, for $|z| \leq 1/2$
$$\log(1+z) = \sum_{k\geq 1} (-1)^{k+1} \frac{z^k}{k}. \tag{2.10}$$
We this definition, one can prove that $e^{\log(1+z)} = 1+z$ for $|z| \leq 1/2$, e denoting the complex exponential defined by $e^z = \sum_{k\geq 0} \frac{z^k}{k!}$.
Prove that, for u real, $\left|e^{iu\Delta M_j} - 1\right| \leq e^{|u|K} - 1$, then that
$$\left|\mathbb{E}\left(e^{iu\Delta M_j} \middle| \mathcal{F}_{j-1}\right) - 1\right| \leq e^{|u|K} - 1,$$

For $|u| \leq C_K = \frac{1}{K}\log(3/2)$, prove that we can define, using the definition (2.10)

$$\phi_j(iu) = \log \mathbb{E}\left(e^{iu\Delta M_j} \middle| \mathscr{F}_{j-1} \right),$$

and that we have $e^{\phi_j(iu)} = \mathbb{E}\left(e^{iu\Delta M_j} \middle| \mathscr{F}_{j-1}\right)$.

3. Prove that, for $|u| \leq C_K$,

$$\left(\exp\left\{ iuM_n - \sum_{j=1}^{n} \phi_j(iu) \right\}, n \geq 0 \right)$$

is a (complex) martingale.

4. Le u be a given real number, show that for a n large enough

$$\mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} - \sum_{j=1}^{n} \phi_j(iu/\sqrt{n}) \right) \right] = 1.$$

5. Prove that, for x a complex number such that $|x| \leq 1/2$

$$|e^x - 1 - x - x^2/2| \leq |x|^3 \text{ and } |\log(1+x) - x| \leq |x|^2.$$

6. Show that, for n large enough

$$\left| \mathbb{E}\left(e^{i\frac{u}{\sqrt{n}}\Delta M_j} \middle| \mathscr{F}_{j-1} \right) - 1 + \frac{u^2}{2n}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathscr{F}_{j-1} \right) \right| \leq \frac{u^3}{n^{3/2}} K^3,$$

and that, for a $c > 0$ (depending on u), for n large enough, for all $j \leq n$

$$\left| \phi_j\left(\frac{iu}{\sqrt{n}} \right) + \frac{u^2}{2n}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathscr{F}_{j-1} \right) \right| \leq \frac{c}{n^{3/2}},$$

and deduce, using (2.9), that, for a given u

$$\lim_{n\to+\infty} \sum_{j=1}^{n} \phi_j\left(\frac{iu}{\sqrt{n}} \right) = -\frac{\sigma^2 u^2}{2}, \text{a.s.}$$

7. Proves that

$$\lim_{n\to+\infty} \mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} - \sum_{j=1}^{n} \phi_j(iu/\sqrt{n}) \right) \right] - \mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} + \frac{\sigma^2 u^2}{2} \right) \right] = 0,$$

and deduce that $\lim_{n\to+\infty} \mathbb{E}\left[\exp\left(iu\frac{M_n}{\sqrt{n}} \right) \right] = \exp\left(\frac{\sigma^2 u^2}{2} \right)$. Conclude that $\frac{M_n}{\sqrt{n}}$ converge in distribution to a gaussian random variable.

8. Generalize the result when

$$\lim_{n\to+\infty} \frac{\langle M \rangle_n}{a(n)} = \sigma^2, \text{a.s.}$$

where $a(n)$ is a sequence of positive real numbers increasing to $+\infty$ with n.

Exercise 20. 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 $\beta_n = 1/\left(\prod_{k=1}^{n-1}(1 - \frac{c}{k})\right)$. 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 21. 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

$$\mathbb{E}(|X_n - X|) \leq \mathbb{E}(|\phi_K(X_n) - \phi_K(X)|) + 2\mathbb{E}(\hat{X}\mathbf{1}_{\{\hat{X}\geq K\}}).$$

3. Prove that $\lim_{K\to\infty} \mathbb{E}(\hat{X}\mathbf{1}_{\{\hat{X}\geq K\}}) = 0$.

4. Prove, for a given K, that

$$\mathbb{E}(|\phi_K(X_n) - \phi_K(X)|) \leq 2K\mathbb{P}(|X_n - X| \geq \varepsilon) + \varepsilon,$$

   and deduce the extended Lebesgue theorem.

Exercise 22. 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 21.

PROBLEM 1. 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 $\mathbb{E}(f(X))$ sous la forme

$$\mathbb{E}(f(X)) = \mathbb{E}(H(f;\lambda,U)), \tag{2.11}$$

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 :

   $$H(f;\lambda,U) = e^{-\lambda G - \frac{\lambda^2}{2}} f(G+\lambda), \text{ où } G = \phi^{-1}(U)$$

   permet de satisfaire la condition (2.11).

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 (2.11).

3. On considère une suite $(U^n, n \geq 1)$ de variables aléatoires indépendantes uniformément distribuées sur $[0,1]^d$. On note $\mathscr{F}_n = \sigma\left(U^k, 1 \leq k \leq n\right)$.

   Soit λ un vecteur fixé, comment peut on utiliser $H(f;\lambda,U)$ pour estimer $\mathbb{E}(f(X))$? 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 $\mathscr{F}_n$-mesurable (λ_0 est supposée constante). On suppose que

$$\begin{array}{l}\text{pour tout } \lambda \in \mathbb{R}^p, s^2(\lambda) = \text{Var}\,(H(f;\lambda,U)) < +\infty, \\ \text{et } s^2(\lambda) \text{ est une fonction continue de } \mathbb{R}^d \text{ dans } \mathbb{R}.\end{array} \tag{2.12}$$

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*[1], $(M_n, n \geq 0)$ est une $\mathscr{F}_n$-martingale dont le crochet $\langle M\rangle_n$ s'exprime sous la forme $\langle M\rangle_n = \sum_{i=0}^{n-1} s^2(\lambda_i)$.

[1] i.e., il existe K réel positif tel que pour tout λ et u, $|H(f;\lambda,u)| \leq K < +\infty$

5. Plus généralement montrer que, sous l'hypothèse (2.12), $(M_n, n \geq 0)$ 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
$$\lim_{n\to+\infty} \frac{\sum_{i=0}^{n-1} s^2(\lambda_i)}{n} = c, \tag{2.13}$$
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 (2.13) pour obtenir un théorème central limite dans la méthode précédente (ne pas chercher à la vérifier) ? Enoncer le résultat que l'on obtiendrait alors.

Appendix A

A proof of the central limit theorem for martingales

The proof given here is a slightly adapted version of [Major(2013)].

Proof. We will need an extension of Lebesgue theorem (also known as Lebesgue theorem) which says that if X_n converge in probability[1] 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 21 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

$$\tau_k = \inf\left\{ j \geq 0, \langle M^{(k)}\rangle_{j+1} > 2\sigma^2 \right\}, \tag{A.1}$$

The random variable τ_k is a stopping time with respect to the σ-algebras $\mathscr{F}_j$, $j \geq 0$, since the random variable $\langle M^{(k)}\rangle_{j+1}$ is $\mathscr{F}_j$ 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

$$\langle M^{[k]}\rangle_j = \langle M^{(k)}\rangle_{j\wedge\tau_k} \leq 2\sigma^2, \tag{A.2}$$

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

$$\Delta\langle M^{[k]}\rangle_j := \langle M^{[k]}\rangle_j - \langle M^{[k]}\rangle_{j-1} = \mathbb{E}\left((\Delta M_j^{[k]})^2 \,\middle|\, \mathscr{F}_{j-1}\right), \text{ where } \Delta M_j^{[k]} := M_j^{[k]} - M_{j-1}^{[k]}.$$

Note that we can rewrite the Lindeberg condition (2.5) as

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{(k)})^2 \mathbf{1}_{\left\{|\Delta M_j^{(k)}| \geq \varepsilon\right\}} \,\middle|\, \mathscr{F}_{j-1}\right) = 0, \text{ in probability.}$$

Moreover $|\Delta M_j^{[k]}| = \mathbf{1}_{\{j<\tau_k\}} |\Delta M_j^{(k)}| \leq |\Delta M_j^{(k)}|$, so

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}| \geq \varepsilon\right\}} \,\middle|\, \mathscr{F}_{j-1}\right) = 0, \text{ in probability.}$$

[1] which is a weaker asumption than the almost surely convergence usually assumed.

Now, as $\mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}|\geq\varepsilon\right\}} \middle| \mathscr{F}_{j-1}\right) \leq \mathbb{E}\left((\Delta M_j^{[k]})^2 \middle| \mathscr{F}_{j-1}\right) = \Delta\langle M^{[k]}\rangle_j$, 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 $M^{[k]}$

$$\lim_{n\to+\infty} \sum_{j=1}^{k} \mathbb{E}\left((\Delta M_j^{[k]})^2 \mathbf{1}_{\left\{|\Delta M_j^{[k]}|\geq\varepsilon\right\}}\right) = 0. \tag{A.3}$$

Now note that $\langle M^{[k]}\rangle_k = \frac{1}{a(k)}\langle M\rangle_{k\wedge\tau_k}$ and, using hyphothesis (2.4), that

$$\lim_{k\to+\infty} \mathbb{P}(\tau_k > k) = \lim_{k\to+\infty} \mathbb{P}\left(\langle M\rangle_k \leq 2\sigma^2 a(k)\right) = 1.$$

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 $M^{[k]}$

$$\lim_{k\to+\infty} \mathbb{E}\left(\langle M^{[k]}\rangle_k\right) = \sigma^2. \tag{A.4}$$

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 Slutsky's lemma, it remains to prove the convergence in probability of $\bar{S}_k$ to a gaussian random variable, i.e.

$$\lim_{k\to\infty} \mathbb{E}(e^{it\bar{S}_k}) = e^{-\sigma^2 t^2/2} \quad \text{for all real numbers } t. \tag{A.5}$$

And we will show that relation (A.5) follows from

$$\lim_{k\to\infty} \mathbb{E}(e^{itS_k + t^2 U_k/2}) = 1 \quad \text{for all real numbers } t, \tag{A.6}$$

where $U_k = \langle M^{[k]}\rangle_k$. Since U_k converge in probability to σ^2 as $k\to\infty$, and $0 \leq U_k \leq 2\sigma^2$ for all $k \geq 1$ because of (A.2), we have that $e^{t^2\sigma^2} \geq \left|e^{t^2 U_k/2} - e^{\sigma^2 t^2/2}\right|$ with the right-hand side converging in probability to 0 as $k\to\infty$. Hence, by the extended Lebesgue's theorem, $\lim_{k\to\infty}\mathbb{E}\left(\left|e^{t^2 U_k/2} - e^{\sigma^2 t^2/2}\right|\right) = 0$. Since

$$\left|\mathbb{E}\left(e^{itS_k + t^2 U_k/2} - e^{itS_k + \sigma^2 t^2/2}\right)\right| \leq \mathbb{E}\left(\left|e^{itS_k + t^2 U_k/2} - e^{itS_k + \sigma^2 t^2/2}\right|\right) = \mathbb{E}\left(\left|e^{t^2 U_k/2} - e^{\sigma^2 t^2/2}\right|\right),$$

we deduce that Hence $\lim_{k\to\infty}\mathbb{E}(e^{itS_k + t^2 U_k/2} - e^{itS_k + \sigma^2 t^2/2}) = 0$. Formula (A.5) follows from this statement if we can prove (A.6).

For this we first show that

$$\left|\mathbb{E}\left(e^{itS_k + t^2 U_k/2}\right) - 1\right| \leq e^{\sigma^2 t} \sum_{j=1}^{k} \mathbb{E}\left|e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}} \middle| \mathscr{F}_{j-1}\right) - 1\right|. \tag{A.7}$$

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

$$\begin{aligned}\mathbb{E}\left(e^{itS_k + t^2 U_k/2} - 1\right) &= \sum_{j=1}^{k} \mathbb{E}\left(e^{itS_{k,j} + t^2 U_{k,j}/2} - e^{itS_{k,j-1} + t^2 U_{k,j-1}/2}\right) \\ &= \sum_{j=1}^{k} \mathbb{E} e^{itS_{k,j-1} + t^2 U_{k,j-1}/2}\mathbb{E}\left[\left(e^{it\Delta M_j^{[k]} + t^2\Delta\langle M^{[k]}\rangle_j/2} - 1 \middle| \mathscr{F}_{j-1}\right)\right].\end{aligned}$$

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

$$\left|\mathbb{E}\left(e^{itS_k+t^2U_k/2}-1\right)\right| \le e^{\sigma^2 t}\sum_{j=1}^{k}\mathbb{E}\left|\mathbb{E}\left(e^{it\Delta M_j^{[k]}+t^2\Delta\langle M^{[k]}\rangle_j/2}-1\middle|\mathscr{F}_{j-1}\right)\right|,$$

and as $\mathbb{E}\left(e^{it\Delta M_j^{[k]}+t^2\Delta\langle M^{[k]}\rangle_j/2}-1\middle|\mathscr{F}_{j-1}\right) = e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\mathscr{F}_{j-1}\right)-1$, this implies the estimate (A.7).

To prove formula (A.6) with the help of inequality (A.7) we have to give an estimate for $\mathbb{E}\left|e^{t^2\Delta\langle M^{[k]}\rangle_j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\mathscr{F}_{j-1}\right)-1\right|$.

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 (A.2). We can estimate the expression

$$\eta_{k,j}^{(2)}=\mathbb{E}\left(e^{it\Delta M_j^{[k]}}-1+\frac{t^2(\Delta M_j^{[k]})^2}{2}\middle|\mathscr{F}_{j-1}\right)$$

in a similar way. To do this let us fix a small number $\varepsilon\in(0,1]$, 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 $\alpha(x)=t^2x^2\mathbf{1}_{\{|x|>\varepsilon\}}+\frac{\varepsilon}{6}|t|^3x^2\mathbf{1}_{\{|x|\le\varepsilon\}}$. Indeed, $|e^{itx}-1-itx|=\left|-\int_0^x(x-y)t^2e^{ity}dy\right|\le\frac{t^2x^2}{2}$ and $|e^{itx}-1-itx+\frac{t^2x^2}{2}|=\left|-i\int_0^x\frac{(x-y)^2}{2}t^3e^{ity}dy\right|\le\frac{|t|^3|x|^3}{6}$ and we get the estimate by bounding the expression $\left|e^{itx}-1-itx+\frac{t^2x^2}{2}\right|$ by t^2x^2 if $|x|>\varepsilon$ and by $\frac{|t|^3|x|^3}{6}\le\varepsilon\frac{|t|^3x^2}{6}$ if $|x|\le\varepsilon$. Using that $\mathbb{E}(\Delta M_j^{[k]}|\mathscr{F}_{j-1})=0$ and taking the conditional expectation in the last inequality with respect to $\mathscr{F}_{j-1}$ we get

$$\begin{aligned}|\eta_{k,j}^{(2)}|&\le\mathbb{E}\left(\left|e^{it\Delta M_j^{[k]}}-1-it\Delta M_j^{[k]}+\frac{t^2(\Delta M_j^{[k]})^2}{2}\right|\middle|\mathscr{F}_{j-1}\right)\\&\le\mathbb{E}\left(\alpha(\Delta M_j^{[k]})\middle|\mathscr{F}_{k,j}\right)\le t^2\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)+\frac{\varepsilon}{6}|t|^3\Delta\langle M^{[k]}\rangle_j.\end{aligned}$$

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

$$\begin{aligned}&\left|e^{t^2\Delta\langle M^{[k]}\rangle j/2}\mathbb{E}\left(e^{it\Delta M_j^{[k]}}\middle|\mathscr{F}_{j-1}\right)-1\right|\\&\quad=\left|\left(1+\frac{t^2\Delta\langle M^{[k]}\rangle_j}{2}+\eta_{k,j}^{(1)}\right)\left(1-\frac{t^2\Delta\langle M^{[k]}\rangle_j}{2}+\eta_{k,j}^{(2)}\right)-1\right|\\&\quad\le t^4(\Delta\langle M^{[k]}\rangle_j)^2+K_3(t)\left(|\eta_{k,j}^{(1)}|+|\eta_{k,j}^{(2)}|\right)\\&\quad\le K_4(t)\left((\Delta\langle M^{[k]}\rangle_j)^2+\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)+\varepsilon\Delta\langle M^{[k]}\rangle_j\right).\end{aligned}$$

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 (A.7) imply that

$$|\mathbb{E}e^{itS_k+t^2U_k/2}-1| \le K_5(t)\left(\sum_{j=1}^{k}\mathbb{E}\left((\Delta\langle M^{[k]}\rangle_j)^2\right)\right. \\ \left.+\sum_{j=1}^{k}\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\right)+\varepsilon\sum_{j=1}^{k}\mathbb{E}\{\Delta\langle M^{[k]}\rangle_j\}\right). \tag{A.8}$$

To estimate the first sum at the right-hand side of (A.8) let us make the following estimate:

$$\begin{aligned}
\mathbb{E}\left\{(\Delta\langle M^{[k]}\rangle_j)^2\right\} &= \mathbb{E}\left\{\left[\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)+\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\le\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)\right]^2\right\} \\
&\le 2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)^2\right\}+2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\le\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)^2\right\} \\
&\le 2\mathbb{E}\left\{\Delta\langle M^{[k]}\rangle_j\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)\right\}+2\varepsilon^2\mathbb{E}\left\{\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|\le\varepsilon\right\}}\middle|\mathscr{F}_{j-1}\right)\right\} \\
&\le 4\sigma^2\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\right)+2\varepsilon^2\mathbb{E}\left(\Delta\langle M^{[k]}\rangle_j\right).
\end{aligned}$$

Using this estimate and (A.8), we obtain

$$\left|\mathbb{E}e^{itS_k+t^2U_k/2}-1\right| \le K_6(t)\left(\sum_{j=1}^{k}\mathbb{E}\left((\Delta M_j^{[k]})^2\mathbf{1}_{\left\{|\Delta M_j^{[k]}|>\varepsilon\right\}}\right)+\varepsilon\sum_{j=1}^{k}\mathbb{E}\left(\Delta\langle M^{[k]}\rangle_j\right)\right),$$

and relations (A.3), (A.4) imply formula (A.6). Thus we have proved the central limit theorem. □

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.

[Cochran(1953)] William G. Cochran. *Sampling techniques*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953.

[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)Fox, Bratley, and Neiderreiter] 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.

[Glasserman et al.(1999)Glasserman, Heidelberger, and Shahabuddin] 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.

[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.

[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)Press, Teukolsky, Vetterling, and Flannery] 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)Radovic, Sobol', and Tichy] 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.

[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.

[Williams(1991)] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.