Introduction to Monte-Carlo Methods

Bernard Lapeyre
Halmstad, January 2007

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

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

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

where $f(\cdot)$ is a bounded real valued function. Represent I as $\mathbf{E}(f(U))$, 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

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

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

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

Remark 1.2 The random variable X_1 needs to be integrable. Therefore the Strong Law of Large Numbers does not apply when X_1 is Cauchy distributed, that is when its density is $\frac{1}{\pi(1+x^2)}$.

Convergence rate We now seek estimates on the error

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

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

Theorem 1.3 (Central Limit Theorem) *Let $(X_i, i \geq 1)$ be a sequence of independent identically distributed random variables such that $\mathbf{E}(X_1^2) < +\infty$.Let σ^2 denote the variance of X_1, that is*

$$\sigma^2 = \mathbf{E}(X_1^2) - \mathbf{E}(X-1)^2 = \mathbf{E}\left((X_1 - \mathbf{E}(X_1))^2\right).$$

Then :

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

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

Remark 1.4 From this theorem it follows that for all $c_1 < c_2$

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

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

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

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

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

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

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

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

A standard estimator for the variance is given by

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

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

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

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

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

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

So, with very little additional computations, (we only have to compute $\bar{\sigma}_N$ on a sample already drawn) we can give an reasonable estimate of the error done by approximating $\mathbf{E}(X)$ 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.

In some situation the Central limit theorem can be improved. The Berry–Esseen Theorem gives a speed of convergence of the Central Limit Theorem.

Theorem 1.5 (Berry–Esseen) *Let $(X_i)_{i\geq 1}$ be a sequence of independent and identically distributed random variables with zero mean. Denote by σ the common standard deviation. Suppose that*

$$\mathbf{E}(|X_1|^3) < +\infty.$$

Define F_N

$$F_N(x) = \mathbf{P}\left(\frac{\sum_{j=1}^N X_j}{\sigma\sqrt{N}} \leq x\right),$$

and ϕ,the distribution function of a Gaussian law with zero mean and unit variance, as

$$\phi(x) = \int_{-\infty}^{x} e^{-u^2/2}\frac{du}{\sqrt{2\pi}}.$$

Then it holds that

$$\begin{aligned}\epsilon_N &= \sup_{x\in\mathbf{R}} |F_N(x) - \phi(x)| \\ &\leq \frac{C\mathbf{E}|X_1|^3}{\sigma^3\sqrt{N}}.\end{aligned}$$

Furthermore, $0.398 \leq C \leq 0.8$.

For a proof of the preceding theorem, see Shiryayev [**?**]. This theorem can be extended when $\mathbf{E}(|X_1|^{2+\delta}) < +\infty$, for a $\delta > 0$. In this case, it is known as the Bikelis theorem.

Theorem 1.6 (Bikelis) *Let $(X_n, n \geq 1)$ be a sequence of independent real valued random variables, which are identically distributed. Suppose that $\mathbf{E}(X_1) = 0$ and that there exists $0 < \delta \leq 1$ such that $\mathbf{E}(|X_1|^{2+\delta}) < +\infty$. Set*

$$\sigma^2 = \mathbf{E}(X_1^2).$$

Denote by Φ the distribution function of a Gaussian law with zero mean and unit variance.

There exists a universal constant A, independent of N and of the sequence $(X_n, n \geq 1)$ such that, for all x,

$$|F_N(x) - \Phi(x)| \leq \frac{A\mathbf{E}(|X_1|^{2+\delta})}{N^{\delta/2}\sigma^{2+\delta}(1+|x|)^{2+\delta}}.$$

There is no precise estimate of the constant A, but one knows that

$$\frac{1}{\sqrt{2\pi}} \leq A < 1.$$

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., [**?**]) 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.

3 Simulation of the uniform law

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

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

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

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

4 Simulation of some common laws of finance

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

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

Proposition 4.1 *Let X be a real random variable with strictly positive and continuous density $p_X(x)$. Let F be its distribution function defined by*

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

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

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

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

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

Remark 4.2 Simulation of an exponential law
This result can be extended to a general case (that is to say a law which does not admit a density, or with density not necessarily strictly positive). In this case we have to define the inverse F^{-1} of the increasing function F by

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

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

Remark 4.3 Simulation of asset models with jumps uses random variables with exponential laws. The preceding proposition applies to the simulation of an exponential law of parameter $\lambda > 0$, whose density is given by

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

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

Remark 4.4 This method can also be used to sample Gaussian random variables. Of course neither the distribution function nor its inverse are exactly known but some rather good polynomial approximations can be found, for instance, in [**?**]. 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, $F(x) = \mathbf{P}(X \leq x)$ and F^{-1} be its inverse. The law of Y defined by

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

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

Gaussian Law The Gaussian law with mean 0 and variance 1 on $\mathbf{R}$ 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

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

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

Proposition 4.5 *Let U and V be two independent random variables which are uniformly distributed on $[0, 1]$. Let X and Y be defined by*

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

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

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

Simulation of a Gaussian vector To simulate a Gaussian vector

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

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

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

$$AA^* = C,$$

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

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

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

then i increasing from 2 to d,

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

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

Now, if we assume that $G = (G^1, \ldots, G^d)$ 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 $(G^1, \ldots, G^d)$ of *independent* Gaussian variables as explained above. Then return the vector $X = AG$.

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

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

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

Bibliographic remark A very complete discussion on the simulation of non uniform random variables can be found in [**?**], results and discussion on the construction of pseudo-random sequences in Knuth [**?**]. [**?**],[**?**] and [**?**] are reference books on simulation methods. See also the survey paper by Niederreiter [**?**] and the references therein, in particular these which concern nonlinear random number generators.

5 Exercises and problems

Exercise 5.1 Let X and Y be two independent Gaussian random variables with mean 0 and variance 1.

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

2. Using the previous result, deduce proposition **??**.

Exercise 5.2 Let λ and K be two real positive numbers such that $\lambda < K$ and X_m be the random variable

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

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

Problem 5.3 Let Z be a random variable given by

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

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

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

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

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

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

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

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

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

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

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

   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 $\mathbf{1}_{\left\{\lambda_1 e^{\beta_1 X_1} + \lambda_2 e^{\lambda_2 X_2} > t\right\}}$. Propose a new Monte-Carlo method to compute p.

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

   Deduce a variance reduction method computing p.

References

[AS70] M. Abramovitz and A. Stegun. *Handbook of Mathematical Functions*. Dover, 9th edition, 1970.

[Dev85] L. Devroye. *Non Uniform Random Variate Generation*. Springer-Verlag, 1985.

[HH79] J. Hammersley and D. Handscomb. *Monte Carlo Methods*. Chapman and Hall, London, 1979.

[Knu81] D.E. Knuth. *The Art of Computer programming, Seminumerical Algorithms*, volume 2. Addison-Wesley, 1981.

[L'E90] P. L'Ecuyer. Random numbers for simulation. *Communications of the ACM*, 33(10), Octobre 1990.

[Nie95] H. Niederreiter. New developments in uniform pseudorandom number and vector generation. Dans P.J.-S. Shiue and H. Niederreiter, editeurs, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 de *Lecture Notes in Statistics*, pages 87–120, Heidelberg, New York, 1995. Springer-Verlag.

[Rip87] B.D. Ripley. *Stochastic Simulation*. Wiley, 1987.

[Rub81] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley Series in Probabilities and Mathematical Statistics, 1981.

[Shi84] A.N. Shiryayev. *Probability*. Springer-Verlag, 1984.