An introduction to the Monte Carlo method and its applications

Vlad Bally

2.04.2003

1 Introduction

We give a short introduction to the Monte Carlo method and its applications. In the first section we explain the motivations of the Monte Carlo method, we present the random number generators and finally we discuss the error which is made in the Monte Carlo method.

In the second section we present the Brownian motion and the link with the heat equation. This appears as one of the most important fields in which the Monte Carlo method is used. In fact it is well known that the Monte Carlo method represents an alternative to analytical methods (finite differences, finite elements) for solving numerically Partial Differential Equations in high dimension. The starting point of such a procedure is the representation of the solution of a partial differential equation as an expectation related to a diffusion process (the Feynman-Kac formula). Of course this subject is much to involved and requires a good knowledge of the stochastic calculus so we are not able to treat it in a complete way here. But the simple case of the Brownian motion in relation with the heat equation already represents a good introduction to the subject and we are able to present it in an elementary frame. We also give in this section two other algorithms (except for the Monte Carlo method) in order to solve the heat equation: the finite differences algorithm and the tree algorithm. So one has a view on the different ways in which the numerical problem may be solved by analysts or probabilists.

In the third section we give an elementary description of the quantities that one wants to compute in mathematical finance, essentially the European option prices. We keep to an extremely superficial level and almost no proves are given. This is just to mention one field of applications of the Monte Carlo method which is blowing up in this moment.

Finally in the forth and last section we come back to the Monte Carlo method itself and discuss some reduction of variance technics: control variables and importance sampling.

2 General presentation of the Monte Carlo method

2.1 The problem

We consider the following simple problem: compute

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

A first deterministic method is to split $[0,1]$ in small intervals and then to approximate the integral by Riemann sums. Say that one takes $x_k = \frac{k}{n}, k = 0,...,n-1$ and one denotes $I_k = [x_k, x_{k+1})$. Then

$$\begin{aligned}\int_{[0,1]^d} f(x)dx &\sim \sum_{k_1=0}^{n-1} ... \sum_{k_d=0}^{n-1} f(x_{k_1},...,x_{k_n}) m(I_{k_1} \times ... \times I_{k_n}) \\ &= \sum_{k_1=0}^{n-1} ... \sum_{k_d=0}^{n-1} f(x_{k_1},...,x_{k_n}) \frac{1}{n^d} =: S_n(f).\end{aligned}$$

If one looks to the error one obtains

$$\left| \int_{[0,1]^d} f(x)dx - S_n(f) \right| \sim \frac{[f]_1}{n}$$

where $[f]_1$ is the Lipschiz constant of f, that is $|f(x) - f(y)| \leq [f]_1 |x-y|$.

There are two things which has to be noted:

$\Diamond$ If f is Lipschiz continuous the evaluation of the error is simple - but if f is just a measurable function then the error may be very bad. Let us take the toy example $f = 1_{[0,1]^d - A}$ where A is the set of the points in $[0,1]$ which have rational coordinates. Then $\int_{[0,1]^d} f(x)dx = 1$ and $S_n(f) = 0$ for every n (because $(x_{k_1},...,x_{k_n})$ has rational coordinates). So we have even not convergence.

Conclusion: if the function f is not regular there are some troubles with deterministic methods.

$\Diamond$ Suppose that we work with a Lipschiz continuous function (so there is no problem to evaluate the error) and we want to get an error of order $\varepsilon = 0.01$. Then $\varepsilon = \frac{[f]_1}{n}$ gives $n = 100$ (say that $[f]_1 = 1$). Suppose also that we are in large dimension, say $d = 30$. Then which is the number of points that we have to use? Answer: $n^d = 100^{30} = 10^{60}$ and this seems rather huge.... Let us put the problem otherwise. One allows himself to use $n = 10^6$ points - say that this comes from the power of his computer. Then one asks about the precision that he may obtain. If one wants to cover the interval $[0,1]$ with these points than the precision is 10^{-6} and this is fine. Now, if one wants to cover the square $[0,1]^2$ then one obtains a precision $\sqrt{10^{-6}} = 10^{-6/2} = 10^{-3}$ essentially because one has to cover each of the boards and there are two boards. If one goes on in R^3 then one obtains $10^{-6/3} = 10^{-2}$ and if one goes further to dimension $d = 30$ one obtains the error $10^{-6/30} = 10^{-1/5} = \frac{1}{1.58} = 0.63$. And this is not famous.

Conclusion: in large dimension deterministic methods get in trouble because the number of points needed in order to cover the unit hypercube growth exponentially with the dimension.

We go on and present the probabilistic approach, that is the Monte Carlo method. Suppose that on some probability space (Ω, F, P) we have a sequence of random variables $X_n, n = 1, ...$ which take values in $[0,1]^d$, and are independent and identically distributed. In order to feet in our previous example we suppose that they are uniformly distributed, that is $P(X_n \in [a,b]) = b - a$. So the law is given by the Lebsgue measure and we have $E(f(X_n)) = \int_{[0,1]^d} f(x)dx$. Then the **strong Law of Large Numbers** says that

$$\int_{[0,1]^d} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} f(X_n) \quad a.s.$$

if $N \to \infty$.

So, at soon as we have the sequence $X_n, n = 1, ..., N$ the above formula produces an approximation of our integral. And this is the **Monte Carlo method**.

In the previous example we have considered the uniform law because the integral that we have to compute is with respect to the Lebesgue'ss measure. But in the general formulation we consider some arbitrary probability measure μ on R^d and we want to compute $\int_{R^d} f(x)d\mu(x)$. Then we take the sequence X_n to have the common law μ and we have

$$E(f(X_1)) = \int_{R^d} f(x)d\mu(x) \sim \frac{1}{N}\sum_{n=1}^{N} f(X_n) \quad a.s.$$

For example, if μ is the standard Gaussian law on R then we obtain

$$\int_{R^d} f(x)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \sim \frac{1}{N}\sum_{n=1}^{N} f(X_n) \quad a.s.$$

with X_n a sequence of standard Gaussian random variables.

Let us compare this approach with the deterministic one. Take the toy example $f = 1_{R^d - A}$. Since the law of X_n is the Lebsgue measure, one has $P(X_n \in A) = 0$ and so $f(X_n) = 1$ almost surely. Then $\frac{1}{N}\sum_{n=1}^{N} f(X_n) = 1$ almost surely and our approximation is perfect. In other words X_n avoids to go in the "bad" region A just because its law do not "see" this region. There is a coherence which does not appear if one chose the discretization grid in an analytical, objective, way. So one advantage of the Monte Carlo method is that it works well for functions which are not regular - just measurable functions. Let us go on to the convergence speed. We will prove in the following section that $\int_{R^d} f(x)d\mu(x) - \frac{1}{N}\sum_{n=1}^{N} f(X_n) \sim \frac{C}{\sqrt{N}}$ in some sense which will be precized there. And this is true in any dimension, for example for $d = 30$. So the speed of convergence is **dimension free**. This is maybe the main advantage

of the Monte Carlo method. It is a common fact accepted by everyone that in small dimension (typically $d = 1, 2$ and maybe 3) one may manage to produce deterministic algorithms which are faster then the Monte Carlo method. But as the dimension increases, the Monte Carlo method remains the only alternative - because the analytical methods become to heavy, need to many points... think for example to dimension 30.

Now there are several natural questions related to the Monte Carlo method:

1. How to produce in practice a sample $X_1, ..., X_N$ of independent random variables of a given law μ? Here comes on the random number generators.

2. Which is the speed of convergence? How to measure the speed of convergence? Here comes on the Central Limit Theorem, the Berry Essen theorem, confidence intervals.

3. How to accelerate the speed of convergence in order to improve the algorithms? Here comes on the reduction of variance technics as control variables, importance sampelig, anthitetic variables.

4. Which are the typical examples in which the Monte Carlo method is used? There will be some examples and maybe stochastic calculus and Stochastic Differential Equations. But anyway something is clear from the beginning: the quality which we want to compute has to written as an expectation $Ef(X)$!

2.2 Random numbers generator

The problem we address in this subsection is to produce a sequence of real numbers $x_1, ..., x_N$ such that $x_1 = X_1(\omega), ..., x_N = X_N(\omega)$ where $X_1, ..., X_N$ is a sample of a random variable X of given law μ. Say for a moment that X has a Bernoulli law that is $P(X = 1) = P(X = -1) = \frac{1}{2}$. Then one would toast a coin N times and note the results - and this would be the sequence $x_n, n = 1, ..., N$. But think for a moment that $N = 10^6$. Then this takes some time and does not seem realistic from a practical point of view. Then, what people do is that they produce the numbers $x_1, ..., x_N$ using a mathematical formula - which of course is deterministic and consequently the result is not random. This is why we say that the numbers $x_1, ..., x_N$ are not random numbers but "pseudo random numbers". They behave in some statistical sense as the authentic random numbers obtained by toasting a coin, but they are not these real random numbers. But recall that these numbers will be used in approximation algorithms, and, as we will see in the sequel, the errors produced in this algorithms are expressed in probabilistic terms. So the fact that we do not use real random numbers but only pseudo random numbers produces one more 'statistical error'. The uncomfortable fact about this is that the systhematic study of this type of error and of its contribution to the global error seems to be very difficult to achieve - but this is mathematics! So practicians call for a "pseudo random number generator" which is included in the software of their computer, this generator generates (using of course a deterministic procedure) a sequence $x_1, ..., x_N$ of numbers and we **believe** that these numbers are real random numbers, that is the result of toasting N times a coin. And we forget about the possible error because we have more serious problems to deal with!

But there is a second problem that we have to face. There exists an infinity of possible probability laws and the user of the Monte Carlo method would need to produce random numbers coming from a law or another. So does the pseudo number generator contain an infinity of sub-routines each of them corresponding to a different law? Of course not. So the idea is the following: one is able to produce random numbers corresponding to the uniform law on $[0,1]$. This is the only law for which we produce random numbers. Then, using some probabilistic methods one succeed to transform the numbers $x_1, ..., x_N$ which corespond to the uniform law in some other numbers $y_1, ..., y_N$ which correspond to another law μ, say the standard Gaussian law. This is done essentially for the more important laws which appear in the probability theory: Gaussian, Poisson... And this is done by the number generator itself, automatically. We may ask to the random generator to produce Gaussian random numbers for example and we get what we want - it is the computer which produces the uniform random numbers first and then achieves the transformation itself and finally gives us directly Gaussian random numbers. But of course there exists only a finite number of transformations which may be included in a software so there is only a finite number of laws for which we may obtain directly the result. What about other laws which appear naturally in concrete problems? Here comes on a completely different approach. In fact, when we deal with a concrete problem a first step consists in giving a mathematical modelisation for this problem and the model that we produce has to be compatible with an "algorithmic approach". This means that this model has to be described by a finite number of operations - if not there is no chance to implement an algorithm in order to solve the problem. So our initial problem may be "infinite dimensional" or may describe some "continuous time evolution". But the algorithm that we produce in order to solve numerically this problem has to be finite. So an approximation procedure is always necessary in order to pass from our initial problem to a finite model which may be implemented. And one part of this approximation procedure consists in the approximation of the general laws μ coming on in the initial problem by another law $\overline{\mu}$ which itself is an aggregate of "standard laws" for which random numbers are viable. So the chain is the following:

Produce uniformly distributed random numbers

$\longrightarrow$ Produce random numbers for standard laws (Gauss, Poisson...)

$\longrightarrow$ Produce random numbers for $\overline{\mu}$which is an agragate of standard laws

$\longrightarrow$ Approximate $\mu$by $\overline{\mu}$.

The first two steps are achieved automatically and the computer does it. The last two steps represent the job of the guy who modelize and implement the algorithm. And this is specific to each particular problem at hand.

We give now an idea about the two first steps. How to obtain uniformly distributed random numbers? One employes the following congruential method. Four numbers U_0, a, b, m are given. U_0 represents the starting point, a and b define an affine transformation and m is the order of the congruence. So we

define recursively

$$U_{n+1} = aU_n + b \pmod{m}.$$

Then ,in order to obtain numbers in $[0,1]$ one takes $x_n = \frac{1}{m}U_n$. Note that once U_0, a, b, m are given, the procedure is completely deterministic and this is why we get "pseudo" random numbers. Note also that the sequence U_n is periodic, that is: there exists a number N_0 such that $U_n = U_{N_0+n}, n = 1, ..., N_0$. This gives the following problem: if one needs a very large sample - N is huge - then one would exhaust the number generator. This means that if N is larger then N_0 then the random numbers begin to behave as a periodic sequence and this is completely different from the behavior of a really random sample from the uniform law. So the problem of the adequacy of the pseudo random numbers becomes serious. In practice people take care to use a number generator which is sufficiently powerful in order to treat their problem - the length of the sample is significantly smaller then the period of the number generator. Typically $N = 10^3$ up to 10^6 and the period of the generator may be $2^{16} \sim 65000$ or, for powerful computers, $2^{32} \sim 10^9$. There exists also some technics which use a combination of congruencies as the one presented above and which permit to obtain a period of order 2^{61}.

Let us now present some standard transformations which produce random numbers for other laws. We begin with the **standard normal law** which is the law on R of density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

We use the following identity of laws result (Box-Muller' method): If U and V are two independent random variables which have uniform law on $[0,1]$ then

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

are independent random variables with normal law.

Once we know this we proceed as follows. Suppose that we need a sample of size N. We call a number generator which produces a sequences $\alpha_n, n = 1..., 2N$ of numbers which correspond to a sample of size $2N$ of independent random variables of uniform law. Then we put $u_n = \alpha_n, n = 1, ..., N$ and $v_n = \alpha_{N+n}, n = 1, ..., N$. So the sequences $U_n = u_n, n = 1, ..., N$ and $V_n = v_n, n = 1, ..., N$ are independent each other and inside each sequence, the random variables are independent also. We conclude that the sequence $(U_n, V_n) = (u_n, v_n), n = 1, ..., N$ represent a sample of the random variables (U, V) which are independent and uniformly distributed. We define now

$$\begin{aligned} x_n &= \sqrt{-2\log u_n}\sin(2\pi v_n) \\ y_n &= \sqrt{-2\log u_n}\cos(2\pi v_n) \end{aligned}$$

and obtain in this way a sample for the random variable (X, Y) with X and Y independent Gaussian random variables. In particular $x_n, n = 1, ..., N$ are the random numbers corresponding to the Gaussian random variable X.

The same idea produces random numbers corresponding to a $d-$dimensional Gaussian vector $(X_1, ..., X_n)$ (so $X_i, i = 1, ..., N$ are independent and standard normal distributed). The components of such a vector are uncorelated, that is $E(X_iX_j) = E(X_i)E(X_j) = 0$. If we want to produce random numbers for a Gaussian vectors with non null correlations (that is $E(X_iX_j) = c^{ij}$) then we have to use some diagonalization of matrix procedure so there are some formulas from linear algebra comming on. Suppose that we want to simulate a Gaussian random variable X which has mean a and variance σ. Then it suffices to note that, if H is a standard normal distributed random variable (mean zero and variance 1) then $X = a + \sigma H$ has the law we need. So we produce $h_n, n = 1, ..., N$ using the above procedure and then take $x_n = a + \sigma h_n$. What is important to note here is that our problem concerns just the law of the random variable at hand and so the manipulations involving law identities are welcomed.

We have seen that everything is OK as long as we work with Gaussian random variables, but the method presented in that case is very particular. We present now another method, based on the **repartition function** and which is much more general. Given a random variable X the repartition function is defined by

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

We assume that the law of X has a continuous strictly positive density p_X so that

$$F(x) = \int_{-\infty}^{x} p_X(y)dy$$

and this is a strictly increasing function taking values in $(0, 1)$. So this function is invertible. We denote by F^{-1} its inverse. Then we take U a random variable which has an uniform law on $(0, 1)$ and define $H = F^{-1}(U)$. One has

$$P(H \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = \int_0^{F(x)} 1dy = F(x).$$

Since H has the same repartition function as X their laws coincide. So, if we know the repartition function of X, we proceed as follows. We produce the sequence of random numbers $u_n, n = 1, ..., N$ using the generator of uniform distributed random numbers and then we put $x_n = F^{-1}(u_n)$.

Let us give an example: the **exponential law** which is given by the density function

$$p_X(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$$

and its repartition function is $F(x) = 1 - e^{\lambda x}$. It is easy to see that $F^{-1}(y) = -\frac{1}{\lambda}\log(1 - y)$ and so, if U is uniformly distributed then $X = -\frac{1}{\lambda}\log(1 - U)$ has an exponential law of parameter λ.

Note that in the repartition function method one is supposed to know the repartition function of the random variable at hand and to be able to invert it. If one wants to use this method in order to simulate a Gaussian law for example, then a problem appears because there is no explicit formula for the repartition function of this law. So one may not use directly this method but has first to

approximate the repartition function by some polynomials and then to put the machinery to work. This si why people prefer the simpler Box-Muller method.

Consider now a **discrete law**, that is a law on $\{a_1, ..., a_n\}$ of discrete density $\{p_1, ..., p_n\}$. This does not enter in the above frame because this law is not absolutely continuous with respect to the Lebesgue's measure, but the construction is analogous. We denote $F_0 = 0$ and $F_k = p_1 + ... + p_k, k = 1, ..., n$ and, having at hand an uniform distributed random variable U we define

$$X = \sum_{k=0}^{n-1} a_k 1_{(F_k, F_{k+1}]}(U).$$

Then X takes values in $\{a_1, ..., a_n\}$ and

$$P(X = a_k) = P(U \in (F_k, F_{k+1}]) = F_{k+1} - F_k = p_k.$$

So we have our random variable. Then we produce the sequence of random numbers $u_n, n = 1, ..., N$ and we put $x_n = \sum_{k=0}^{n-1} a_k 1_{(F_k, F_{k+1}]}(u_n)$.

Bibliography: A very complete discussion about random number generation is to be found in [$Dev85$]. See also [$Nie95$] for a survey paper. Web site devoted to Monte Carlo simulation: http://random.mat.sbg.ac.at/links/.

2.3 Error analysis

The aim of this section is to evaluate the error

$$\varepsilon_N = E(X) - \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Now on we forget the fact that in practice one works with pseudo random numbers and we assume that $X_i = X_i(\omega)$ are really random variables on some probability space (Ω, F, P) and this random variables are independent and have the same law as X. We also assume that $E|X|^2$ is finite. In fact we will not be very careful about the precise hypothesis that are needed but will give the main ideas and avoid technicalities.

The specificity of our problem is that the error $\varepsilon_N = \varepsilon_N(\omega)$ is not a constant (as in deterministic algorithms) but a random variable - because $X_n = X_n(\omega), n = 1, ..., N$ are random variables. So the meaning of the requirement "the error has to be small" has to be understood in a probabilistic sense and this sense is not clear and has to be carefully precised. In fact they are several ways to say that the error is small. The first important fact which represents the foundation of the Monte Carlo method is the **Strong Law of Large numbers** which guarantees that $\varepsilon_N(\omega) \to 0$ for almost every ω. We recall that the weak law of large numbers says that we have convergence in probability, or maybe in square mean, that is

$$E|\varepsilon_N|^2 = E\left|E(X) - \frac{1}{N} \sum_{i=1}^{N} X_i\right|^2 \to 0.$$

But this is not sufficient for the foundation of the Monte Carlo method because what we have is just a sequence of numbers $x_n = X_n(\omega), n = 1, ..., N$ which represent a particular fixed realization of the sequence of random variables $X_n, n = 1, ..., N$. So we have not access to expectations but just to specific realizations. In other words, we have a fixed ω and we hope that this is a good one, that is, for this precise ω the convergence $\varepsilon_N(\omega) \to 0$ holds true. The **strong** law of large numbers says that the probability to be wrong is null.

So the strong law of large numbers guarantees that the error vanishes when $N \to \infty$. The speed of convergence in the law of large numbers is given by the **Central Limit Theorem** which asserts that

$$(CLT) \quad \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (X_n - E(X)) \to Z.$$

The convergence is in law and Z is a Gaussian random variable with zero mean and variance $\sigma^2 = E\left|X - EX\right|^2$. This may be interpreted as

$$(CLT') \quad E(X) \approx \frac{1}{N} \sum_{n=1}^{N} X_n + \frac{Z}{\sqrt{N}}$$

so that

$$(CLT'') \quad \varepsilon_N = \frac{Z}{\sqrt{N}}.$$

This is a rather loosely equality because (CLT) does not give an equality for a fixed value of N, but only an asymptotic result, as $N \to \infty$. This is why we employ the notation "$\approx$ " and not "$=$ " in (CLT'). But in order to be more precise one has to employ a more sophisticated result, the Berry Essen theorem -we discuss this further on. For the moment we note that practiciens consider that (CLT'') gives a sufficient accurate idea about the error and this is the basis of all the error analysis which is usually done.

Let us focus on the expression of the error given in (CLT'') and ask about its significance. Note first that Z is a random variable and this random variable may take any real value. Think for example that N is very large, say $N = 10^6$, but we have bad lack and $Z = 10^7$. Then $\varepsilon_N = 10$ and this is not a small error. But

$$P(Z \geq 10^7) = \int_{10^7}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2}) dx \leq \exp(-\frac{10^{2\times 7}}{\sigma^2})$$

and this number is astronomically small. So the probability to have such a bad lack is almost null. But not null. So the Central Limit Theorem does not provide a deterministic evaluation but is just a way of saying that except for a really bad lack (which occurs with a small probability) the error is of order $Z/\sqrt{N}$ and Z has reasonable size. This is all it says, but people feel already safe with this. So this is not a real problem in practice.

Let us look once again to this error. We have

$$E\left|\varepsilon_N\right|^2 = \frac{E\left|Z\right|^2}{N} = \frac{\sigma^2}{N}.$$

Take once again $N = 10^6$ but think that $\sigma^2 = 10^6$ also. Then $E|\varepsilon_N|^2 = 1$ which means that $\varepsilon_N \sim 1$ and this error is not small. And this is a real problem which represents one of the main difficulties that people working in concrete implementation of the Monte Carlo method have to face. And they have to do a lot of tricks in order to arrange things in such a way that the variance of the random variables they work with remains reasonably small, and if they do not succeed then the whole algorithm may fail. We will discuss this problem further on when dealing with the so called "reduction of variance methods".

Let us see another way of reading the information given in the Central Limit Theorem, by means of **confidence intervals.** Take some $a < b$. Then using (CLT') one gets

$$\begin{aligned}(CI) \quad P(a \ &\leq \ \frac{1}{N}\sum_{n=1}^{N} X_n - E(X) \leq b) = P(a \leq \varepsilon_N \leq b) \\ &\approx \ P(a \leq \frac{Z}{\sqrt{N}} \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a\sqrt{N}}^{b\sqrt{N}} \exp(-\frac{x^2}{2\sigma^2})dx.\end{aligned}$$

We look to $[a, b]$ as to a 'confidence interval' in which we want that the error lies. For example one decides to trust the result of the Monte Carlo algorithm if the approximative value $\frac{1}{N}\sum_{n=1}^{N} X_n$ lies in an interval of radius 0.01 around the true value $E(X)$. But, since the result produced by the Monte Carlo method is a random variable, this will not happen for sure, and one has to accept a risk. For example one accepts a risk of 0.001. This means that one asks that

$$P(\frac{1}{N}\sum_{n=1}^{N} X_n \notin [E(X) - 0.01, E(X) + 0.01]) \leq 0.001.$$

Then one asks the question: how large we have to take N in order to achieve this? In view of (CI) this amounts to

$$\begin{aligned}&\frac{2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{-0.01\sqrt{N}} \exp(-\frac{x^2}{2\sigma^2})dx = \\ &\frac{2}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{0.01\sqrt{N}}{\sigma^2}} \exp(-\frac{x^2}{2})dx = \Phi(-\frac{0.01\sqrt{N}}{\sigma^2}) \leq 0.001.\end{aligned}$$

Here Φ is the repartition function of the standard normal distribution

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

which is tabulated. So we look in the tables and find η such that $2\Phi(-\eta) \leq 0.001$ and then we take N such that $\frac{0.01\sqrt{N}}{\sigma^2} = \eta$ that is $N = 100\eta^2\sigma^4$. Once again it appears that if σ is large then we have to take a very large N. This is the way of computing confidence intervals.

Note that we still have $\approx$ in (CI) which means that this represents a loosely evaluation. We present now the theorem which gives the speed of convergence in the Central Limit Theorem and so permits to obtain rigorous equalities for the confidence intervals (even if practicians does not really bother obout this). In order to simply the notation we suppose that $E(X) = 0$ (if this is not true then we replace X by $Y = X - E(X)$).

Theorem 1 *(**Berry-Essen**) Suppose that* $E|X|^3 < \infty$ *and* $Ef(X) = 0$. *Then*

$$(B-E) \quad \delta_N := \sup_{x \in R} \left| P(\frac{X_1 + ... + X_N}{\sqrt{\sigma N}} \leq x) - \Phi(x) \right| \leq \frac{CE|X|^3}{\sigma \sqrt{\sigma N}}$$

and $0.398 \leq C \leq 0.8$.

The CLT says that the sequence $\frac{X_1 + ... + X_N}{\sqrt{\sigma N}}$ converge in law to the standard normal distribution and this is equivalent to the fact that the repartition functions converge to Φ. The above theorem gives a precise evaluation of the speed of convergence. Let us now see how to use this information for evaluations concerning confidence intervals. So we come back to (CI) and write it (recall that we suppose that $E(X) = 0$)

$$\begin{aligned} P(\frac{1}{N}\sum_{n=1}^{N} X_n \quad &\in \quad [a,b]) = P(\frac{1}{\sqrt{\sigma N}}\sum_{n=1}^{N} X_n \in [\frac{a\sqrt{N}}{\sqrt{\sigma}}, \frac{b\sqrt{N}}{\sqrt{\sigma}}]) \\ &= \quad P(\frac{\sum_{n=1}^{N} X_n}{\sqrt{\sigma N}} \leq \frac{b\sqrt{N}}{\sqrt{\sigma}}) - P(\frac{\sum_{n=1}^{N} X_n}{\sqrt{\sigma N}} \leq \frac{a\sqrt{N}}{\sqrt{\sigma}}) \\ &= \quad \Phi(\frac{b\sqrt{N}}{\sqrt{\sigma}}) - \Phi(\frac{a\sqrt{N}}{\sqrt{\sigma}}) + R_N \end{aligned}$$

where R_N is a remainder which satisfies

$$|R_N| \leq 2\delta_N = \frac{2CE|X|^3}{\sigma\sqrt{\sigma N}}.$$

The new point is that this evaluation holds true for every fixed N and not only in a loosely asymptotic sense. What is also important to note is that the evaluation in $(B-E)$ is uniform with respect to x. In our case we have used this evaluation for the values $x = \frac{b\sqrt{N}}{\sqrt{\sigma}}$ and $x = \frac{a\sqrt{N}}{\sqrt{\sigma}}$. And these two points "move" as $N \to \infty$. So, if the evaluation was not uniform ("the same for all $x's$") then we would have bad surprises for large N. Just to illustrate what a bad surprise we would have, suppose for a moment that the evaluation is not uniform. For example we replace $(B-E)$ by the weaker inequality

$$\left| P(\frac{X_1 + ... + X_N}{\sqrt{\sigma N}} \leq x) - \Phi(x) \right| \leq \frac{CE|X|^3}{\sigma\sqrt{\sigma N}} \times |x|, \quad x \in R.$$

Then we use this inequality for our two points and we obtain

$$|R_N| \leq \frac{CE\left|X\right|^3}{\sigma\sqrt{\sigma N}}. \times \frac{|b|\sqrt{N}}{\sqrt{\sigma}} + \frac{CE\left|X\right|^3}{\sigma\sqrt{\sigma N}}. \times \frac{|a|\sqrt{N}}{\sqrt{\sigma}} = \frac{CE\left|X\right|^3}{\sigma^2}(|a|+|b|).$$

And this inequality does even not say that $\lim_N R_N = 0$! So the fact that $(B-E)$ is an uniform evaluation is a key point.

We conclude this section with a discussion on the **empirical variation**. The key constant which appears in all the error evaluations presented before is $\sigma^2 = E\left|X - E(X)\right|^2$. It is clear that this constant influences on the theoretical bounds of the error, but what practice proves is that, bond any theoretical preoccupations, a big variance of X has a concrete influence on the algorithm itself. So having an idea about the size of σ is not a theoretical preoccupation only but a crucial point for practical implementation as well. And generally a theoretical accurate evaluation of σ is difficult to obtain in concrete examples. So all the theory turns around a quantity that one does not really know! But one may construct an estimator for σ :

$$\sigma_N = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \overline{X}_N)^2 \quad with \quad \overline{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

So the algorithm itself permits to obtain in the same time an estimation of the variance and this estimator σ_N may be used in order to measure the error by means of the CLT or confidence intervals. Of course, from a theoretical point of view there are some troubles because one does not know how precise this estimator is - in order to bound the error concerning σ_N we have to use some evaluations CLT type and there σ appears again, so we come back to the starting point. But the important think is that σ_N already gives a "good" idea about the size of σ and from a practical point of view this is sufficient. And this is considered as a very good point for the Monte Carlo method.

3 Brownian motion, heat equation and mathematical finance problems

3.1 The Brownian motion

I begin with some terminology. We are given a probability space (Ω, F, P) and a family of random variables $X_t : \Omega \to R, t \in [0,\infty)$. This is a **stochastic process**. You have to think to the evolution in time (this is why t appears) of some random phenomenons. Long time ago (19'th century) a biologist (called Brown - this is just a coincidence) observed the evolution of a particle of pollen on the surface of very quiet water. And the particle evaluates in a completely unreasonable way - now we would say, in a "random" way. In our notation X_t is the position of the particle at time t. Let us give the key properties of such a process. We say that the process is **continuous** if for almost every $\omega \in \Omega, t \to$

$X_t(\omega)$ is a continuous function. We say that the process has **independent increments** if for every $0 = t_0 < t_1 < ... < t_n$ the random variables $X_{t_1} - X_{t_0}, ..., X_{t_n} - X_{t_{n-1}}$ are independent, this means that for $A_1, ..., A_n \subset R$ which are measurable,

$$P(X_{t_1} - X_{t_0} \in A_1, ..., X_{t_n} - X_{t_{n-1}} \in A_n) = \\ P(X_{t_1} - X_{t_0} \in A_1) \times ... \times P(X_{t_n} - X_{t_{n-1}} \in A_n).$$

From an intuitive point of view $X_{t_i} - X_{t_{i-1}}$ says how the process moved from $X_{t_{i-1}}$ to X_{t_i}, so describes the dynamics of the process between t_{i-1} and t_i. And this is independent of all he has done before. We say that the process is **homogenous** if $X_{t+h} - X_t$ has the same law for every t, in particular the same law as $X_h - X_0$. So he is doing all the time the "same type" of things - but not the same think - because $X_{t+h} - X_t$ has the same law as $X_h - X_0$ but they are not equal (they are even independent, if the process is with independent increments). The dynamics of the pollen particle from the 19'th century had all these properties and this is why people called it Brownian motion. In fact the definition of the Brownian motion is the following:

A process $(X_t)_{t \geq 0}$ is called **standard Brownian motion** if it is continuous, $X_0 = 0$ and for every $0 = t_0 < t_1 < ... < t_n$ the random vector $(X_{t_1} - X_{t_0}, ..., X_{t_n} - X_{t_{n-1}})$ is Gaussian with zero mean and covariance

$$\begin{aligned} E(X_{t_i} - X_{t_{i-1}})E(X_{t_j} - X_{t_{j-1}}) &= 0 \quad if \quad i \neq j \\ &= t_i - t_{i-1} \quad if \quad i = j. \end{aligned}$$

Note that, according to the above definition a Brownian motion has all the above properties: it has independent increments because uncorelated Gaussian random variables are independent, and it is homogenous because the Gaussian random variables $X_{t+h} - X_t$ and $X_h - X_0$ have both zero mean and the same variance h, so they have the same law.

One may wonder how Brown has succeed to find the Brownian motion just looking to a pollen particle moving on a quiet water surface... I may admit that he observed that the small movements of the particle seem independent each another and I can also admit that he observed that the particle does similar things all the time. This means that he observed that the dynamics of the particle corresponds to a continuous process with homogenous independent increments - but how can he see appearing the density $\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ which is specific to the normal distribution, just looking a particle in quiet water? This seems a miracle - but the explanation of this miracle is to be found in the Central Limit Theorem. In order to precise things I will give a precise statement:

Proposition 1 *Suppose that $X_t, t \geq 0$, is a continuous stochastic process which is homogeneous and has independent increments. Suppose also that $E|X_t|^2 < \infty, \forall t > 0$. Then there exists some constants x_0, σ, b such that*

$$X_t = x_0 + \sigma B_t + bt$$

where $B_t, t \geq 0$ is a standard Brownian motion.

Remark 1 *A process X_t of the previous for is called a generalized Brownian motion in contrast with the standard Brownian motion for which $x_0 = b = 0$ and $\sigma = 1$.*

Before giving the proof let us note that x_0 is the "initial value" because $B_0 = 0$ and so, for $t = 0$ we have $X_0 = x_0 + \sigma \times 0 + b \times 0$. Note also that $EX_t = x_0 + bt$. This is because B_t is a centered Gaussian variable so that $EB_t = 0$. If you thing that B_t describes the "disorder" - the movement of the pollen particle - then you may see the dynamics of X in the following way: There is a "principal movement" which is described by the deterministic curve $t \to x_0 + bt$. And this movement is perturbed by a "noise" B_t and σ is a coefficient which describes the intensity of this noise: for a large σ we have much noise and for small σ we have few noise. In mathematical finance σ is called the volatility. The deterministic movement $x_0 + bt$ represents the mean of X_t.

Sketch of the proof. We will not go in all the technicalities which are needed in order to give a rigorous proof but we just give the idea which explains the "miracle". We assume that except the above hypothesis we also know that $EX_t = 0$ which amounts to take $b = x_0 = 0$. Then the assertion is that there exists σ such that $X_t = \sigma B_t$ where B is a Brownian motion. The striking fact is that this means that X_t has a Gaussian law. This is what we have to prove now. In order to simplify even more the notation we will do it for $t = 1$.

We take some $n \in N$ and cut the time interval $[0, 1]$ in n equal intervals of length $\frac{1}{n}$. Then we write

$$X(1) = \sum_{k=1}^{n} (X(\frac{k}{n}) - X(\frac{k-1}{n})).$$

Here and in the sequel we denote $X(t)$ for X_t. If we put $\Delta_k = X(\frac{k}{n}) - X(\frac{k-1}{n}), k = 1, ..., n$ then these are independent identically distributed random variables - this is what our hypothesis say. We also have $E\Delta_k = E\Delta_1 = EX(\frac{1}{n}) = 0$ because of our supplementary assumption. Note also that, since they are identically distributed, they have the same variance. Let $a_n^2 = E\left|X(\frac{1}{n})\right|^2$ be the common variance. Let us see which is the relation between a_n and the variance of $X(1) = \sum_{k=1}^{n} \Delta_k$. We have

$$E\left|X(1)\right|^2 = E\left|\sum_{k=1}^{n} \Delta_k\right|^2 = \sum_{k=1}^{n}\sum_{p=1}^{n} E(\Delta_k \Delta_p).$$

Note that for $k \neq p$, Δ_k and Δ_p are independent random variables of zero mean so that $E(\Delta_k \Delta_p) = E(\Delta_k)E(\Delta_p) = 0$. We conclude that

$$E\left|X(1)\right|^2 = \sum_{k=1}^{n} E(\Delta_k^2) = na_n^2.$$

Denote now $\sigma^2 = E\left|X(1)\right|^2$ - we know from our hypothesis that this quantity

is finite. Then we have $\sigma^2 = na_n^2$ so that

$$a_n = \frac{\sigma^2}{n}.$$

We are now ready to employ the Central Limit Theorem. Define

$$U_k = \frac{\sqrt{n}}{\sigma}\Delta_k.$$

Then $U_k, k = 1, ..., n$ are independent identically distributed random variables with zero mean and variance one. We also have

$$\frac{1}{\sigma}X(1) = \frac{1}{\sigma}\sum_{k=1}^{n}\Delta_k = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}U_k.$$

The Central Limit Theorem says that $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}U_k$ converges in law to a random variable which is normal distributed and has variance one (note that for different values of n the laws of the $U_k, k = 1, ..., n$ are different so the standard CLT does not work directly - one has to use a slight variant which is the so called CLT for "triangles" - but such a theorem exists, and so let us admit the fact that the convergence to the normal law takes place). Then we pass to the limit in the above equality. The term in the right hand side converges to a standard normal distributed random variable and the term in the left hand side remains all the time the same. So we conclude that $\frac{1}{\sigma}X(1)$ has a standard normal distribution and so $X(1)$ has a normal distribution of mean zero and variance σ. □

This is the proof. And so the miracle comes from the fact that **matter the laws** of U_k, the renormalzed sum $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}U_k$ always converge to the normal law - and this is why the density $\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$ appears. This is a central fact which is crucial for probabilistic modelisation. Because Brown may observe some qalitative properties of movement of the pollen particle but he is not able to do any computation. But after using the CLT we know the precise law of the increments of our stochastic process and this is a quantitative information which permits to start computations!

Before going further I have to stop and to mention that from a mathematical point of view things are not so simple. For example a mathematical will ask if the Brownian motion exists. Even the sense in which such a question has to be asked correctly represents a big theoretical effort, and we live this out. We keep ourselves to the level of intuition - Brown looking to the pollen particle - it is there, so it exists.

We present now a discrete version of the Brownian motion - **the Random Walk**. This is an original man who toasts a coin in order to decide if he goes to the right or to the left. We take the time grid $t_k = kh$ where $h > 0$. At each moment t_k the man steps to the right or to the left with a step of size $\delta > 0$. In order to decide in which direction he will go he toasts a coin and goes with probability $\frac{1}{2}$ to the left and with probability $\frac{1}{2}$ to the right. We denote by X_k the position of the man at time t_k. Then what we are saying is that the

sequence of random variables $X_{k+1} - X_k$ are independent and $P(X_{k+1} - X_k = \delta) = P(X_{k+1} - X_k = -\delta) = \frac{1}{2}$. This is a random walk. Now we would like to pass to the limit with $h \to 0$ and $\delta \to 0$ and see what happens. It turns out that we have to be careful and take the time step h and the space step δ in a certain given echilibrium - if not everything blows up or vanishes. And the right equilibrium is given once again by the Central Limit Theorem. We have to take $h = \frac{1}{n}$ and $\delta = \frac{1}{\sqrt{n}}$ so that $\delta = \sqrt{h}$ - the space step is the square-root of the time space. Suppose that this is true and let us see where the man will be at time t. Up to this moment he has done $n_t = [nt]$ steps (because $\frac{[nt]}{n} \leq t < \frac{[nt]+1}{n}$) and so he will be in $X_{n_t} = \sum_{k=1}^{[nt]} (X_k - X_{k-1})$ (we suppose that one starts from the origin, that is $X_0 = 0$). Let us denote $U_k = \sqrt{n}(X_k - X_{k-1})$ so that

$$X_{n_t} = \frac{U_1 + ...U_{[nt]}}{\sqrt{n}} = \sqrt{\frac{[nt]}{n}} \frac{U_1 + ...U_{[nt]}}{\sqrt{[nt]}}.$$

We have $P(U_k = +1) = P(U_k = -1) = \frac{1}{2}$ so that

$$EU_k = 0 \quad and \quad EU_k^2 = 1.$$

The CLT says that $\frac{U_1 + ...U_{[nt]}}{\sqrt{[nt]}}$ converges in law to a standard normal distribution. On the other hand $\sqrt{\frac{[nt]}{n}} \to \sqrt{t}$ and so $X_{n_t} \to B_t$ in law, where B_t is a centred Gaussian random variable of variance t - that is the Brownian motion. So we have

Proposition 2 *If the time step and the space step of a random walk are in the relation $\delta_n = \sqrt{h_n}$ then the random walk converges in law to the Brownian motion as $n \to \infty$.*

Think once again to Brown's pollen particle. In fact the motion of this particle is more like a random walk then like a Brownian motion: at infinitesimal time intervals the particle moves to the left or to the right, according to infinitesimal movements of the water. But the particle has no coin - so this is not exactly a random walk. Anyway, from a "statistical point of view", since the time step is infinitesimal and the space step is also infinitesimal, the Brownian motion is a good approximation of the dynamics of the particle. Another think that one remarks looking to the pollen particle is that it moves all the time but does not go far away! This is because it does not go in a given direction but once to the left, two times to the right and so on... The same happens with the random walk. Think for example that $n = 100$. Then the time step is $h_n = \frac{1}{100}$ and the space step is $\delta_n = \frac{1}{10}$. This means that the space step is huge with respect to the time step - so one moves very much in a short time interval - but it does not go to far. This is just disorder. As a consequence one has the following striking fact about the Brownian path: $t \to B_t$ is continuous but is never differentiable. This means that it is a very rough line - on can even not make a precise drawing of it!

3.2 The heat equation

The heat equation is the following.

$$(H) \quad \begin{aligned} \frac{\partial u}{\partial t}(t,x) &= \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2}(t,x), \quad t>0, x\in R, \\ u(0,x) &= f(x) \end{aligned}$$

where $\sigma > 0$ is a given parameter. The function f is called the initial condition. One looks for a function $u : [0,\infty) \times R \to R$ which is one time differentiable with respect to the time variable t and two times differentiable with respect to the space variable x and which verifies the equation given above. The physical interpretation of this equation is the following. We have a (infinite) bar and x is a point on this bar. At time $t = 0$ we have a quantity $f(x)$ of heat in the point x. So f represents the initial distribution of temperature in the bar. As time goes this distribution changes (very warm points become less warm and very cool points become wormer). Then $u(t,x), x \in R$, represents the distribution of the temperature at time t. And this distribution evaluates all the time. The heat equation describes this evolution.

In order to understand better what is going on we consider a discrete variant of this equation (as the random walk was a discrete variant of the Brownian motion). So instead of continous time we consider just a discrete time grid $t_k = kh, k \in N$ and instead of a continuous space variable we consider a discrete grid $x_p = p\delta, p \in Z$. And we look to the function $\overline{u}(t_k, x_p)$ which represents the quantity of heat at time t_k in the point x_p of the bar. Let us write down the discrete version of the equation (H). In order to do it we have to replace the derivatives which appear there with some "discrete versions". First of all

$$\frac{\partial u}{\partial t}(t_k,x_p) \sim \frac{\overline{u}(t_k+h,x_p)-\overline{u}(t_k,x_p)}{h} = \frac{\overline{u}(t_{k+1},x_p)-\overline{u}(t_k,x_p)}{h}.$$

Moreover the second order derivative is approximated by

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2}(t_k,x_p) &\sim \frac{\frac{\partial \overline{u}}{\partial x}(t_k,x_p+\delta)-\frac{\partial \overline{u}}{\partial x}(t_k,x_p)}{\delta} \\ &\sim \frac{\frac{\overline{u}(t_k,x_p+\delta)-\overline{u}(t_k,x_p)}{\delta} - \frac{\overline{u}(t_k,x_p)-\overline{u}(t_k,x_p-\delta)}{\delta}}{\delta} \\ &= \frac{\overline{u}(t_k,x_p+\delta)+\overline{u}(t_k,x_p-\delta)-2\overline{u}(t_k,x_p)}{\delta^2}. \end{aligned}$$

So we replace the equation (H) by

$$(\overline{H}) \quad \begin{aligned} \frac{\overline{u}(t_{k+1},x_p)-\overline{u}(t_k,x_p)}{h} &= \frac{\sigma^2}{2}\frac{\overline{u}(t_k,x_p+\delta)+\overline{u}(t_k,x_p-\delta)-2\overline{u}(t_k,x_p)}{\delta^2}, \\ \overline{u}(0,x_p) &= f(x_p). \end{aligned}$$

If we solve the above equation we obtain

$$\begin{aligned} \overline{u}(t_{k+1},x_p) &= \overline{u}(t_k,x_p) + \frac{\sigma^2 h}{2\delta^2}(\overline{u}(t_k,x_p+\delta)+\overline{u}(t_k,x_p-\delta)-2\overline{u}(t_k,x_p)) \\ \overline{u}(0,x_p) &= f(x_p). \end{aligned}$$

This is a recursive formula which gives us the dynamics of the heat potential. Note that this produces also an algorithm which permits to solve the equation. In fact we know the solution at the time level $t_0 = 0$ from the initial condition. Then using the equation we may compute

$$\overline{u}(t_1, x_p) = \overline{u}(t_0, x_p) + \frac{h}{2\delta^2}(\overline{u}(t_0, x_p + \delta) + \overline{u}(t_0, x_p - \delta) - 2\overline{u}(t_0, x_p)).$$

Now we know $\overline{u}(t_1, x_p), x_p \in Z$ and we may use $(\overline{H})$ in order to compute $\overline{u}(t_2, x_p), x_p \in Z$, and so on. So we have here an approximation scheme. This is the so called "finite differences" scheme because the derivatives are computed by a finite difference approximation. It is the simpler analytical approximation method. More sophisticated methods using "finite elements" are used in practice.

We go now further to the probabilistic interpretation of the solution of the heat equation. We consider a standard Brownian motion B_t and we define

$$u(t, x) = E(f(x + \sigma B_t)).$$

Theorem 2 *$u(t, x) = E(f(x+\sigma B_t))$ is the unique solution of the heat equation.*

Proof. Since B_t is Gaussian we have

$$E(f(x + \sigma B_t)) = \int f(y) p_t(x, y) dy$$

with

$$p_t(x, y) = \frac{1}{\sigma\sqrt{2\pi t}} \exp(-\frac{|x - y|^2}{2\sigma^2 t}).$$

We compute the derivatives now

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x, y) &= \frac{1}{\sigma\sqrt{2\pi t}} \exp(-\frac{|x - y|^2}{2\sigma^2 t})(-\frac{1}{2t} + \frac{\partial}{\partial t}(-\frac{|x - y|^2}{2\sigma^2 t}) \\ &= p_t(x, y)(-\frac{1}{2t} + \frac{|x - y|^2}{2\sigma^2 t^2}) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial x} p_t(x, y) &= p_t(x, y)\frac{x - y}{\sigma^2 t}, \\ \frac{\partial^2}{\partial x^2} p_t(x, y) &= p_t(x, y)(\frac{1}{\sigma^2 t} + \frac{(x - y)^2}{\sigma^2 t^2}). \end{aligned}$$

So we obtain

$$\frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2} p_t(x, y) = \frac{\partial}{\partial t} p_t(x, y)$$

that is the heat equation. □

We have here the simplest form of a very deep result: the Feynman Kac formula. This formula says that the solution of a Partial Differential Equation

(in short PDE) may be represented as the expectation of the initial value function computed on the trajectories of a diffusion process. In our very particular case we are in dimension one so we have no "partial derivatives" but just usual derivatives. We have also the simplest example of diffusion process, that is the Brownian motion. But the idea is already there: one things that heat is moving in the bar as the pollen particle moves in water. So there is some dynamics of an underlying stochastic process which give rise to the deterministic PDE. And the solution of the deterministic PDE is nothing else then the expectation (average, behavior in "mean value") of the underlying process. So we have two ways of telling the story, in two different languages: the deterministic language - and then we have $PDE's$. And the probabilistic language and then we have "diffusion processes". And we pass from the probabilistic point of view to the deterministic one just by taking expectations.

It is worth to mention that this is also the starting point of the Monte Carlo method for computing solutions of $PDE's$. Because, as told in the beginning, the basic fact which permits to start an Monte Carlo algorithm is that the quantity which is to be computed is an expectation. And the F-K formula produces this representation.

3.3 Algorithms

We already have two algorithms for computing the solution $u(t,x)$.

♦ **Monte Carlo**. We take a sequence $\Delta^1, ..., \Delta^N$ of standard independent Gaussian random variables. We put $B_t^i := \sigma\sqrt{t}\Delta^i$. Then B_t^i will be independent random variables which have the same law as the Brownian motion underlying our model. So the Monte Carlo method consists in computing

$$u(t,x) = E(f(x+B_t)) \sim \frac{1}{N}\sum_{n=1}^{N} f(x+\sqrt{\sigma t}\Delta^n).$$

♦ **Finite differences**. We employ the formula $(\overline{H})$. What people are doing usually is to take $\delta = h = \frac{1}{n}$. Then one computes recursively

$$\begin{aligned} \text{(Step 0)} \quad \overline{u}(0,x_p) &= f(x_p) \\ \text{(Step k)} \quad \overline{u}(t_{k+1},x_p) &= \overline{u}(t_k,x_p) + \frac{\sigma^2 n}{2}(\overline{u}(t_k,x_p+\frac{1}{n}) + \\ &\quad \overline{u}(t_k,x_p-\frac{1}{n}) - 2\overline{u}(t_k,x_p)) \end{aligned}$$

where $x_p = \frac{p}{n}$ and $t_k = \frac{k}{n}$.

It is clear that here we face the "boundary problem". At the step k we have to know the values of $\overline{u}(t_k,x_p)$ for infinitely many values of x_p, and this is not possible. In order to solve this difficulty one takes a "very large" interval $[x-M, x+M]$ and assumes that the values of $\overline{u}(t_k,x_p)$ for x_p outside of this interval does not really meter. So one defines $\overline{u}(t_k,x_p) = 0$ if $x_p \notin [-M,M]$ - and so one comes back to a finite dimensional problem (a finite number of

points). Now the question is: how to determine M? One way of doing it is to employ the probabilistic representation of the solution. One writes

$$u(t,x) = E(f(x+B_t)) =$$
$$E(f(x+B_t)1_{[x-M,x+M]}(x+B_t)) + E(f(x+B_t)1_{[x-M,x+M]^c}(x+B_t)).$$

Then one ignores the second term (there where we are far from the starting point x). Say that we admit an error $\frac{1}{1000}$. Then we want to have

$$E(f(x+B_t)1_{[x-M,x+M]^c}(x+B_t)) \leq \frac{1}{1000}.$$

Suppose that $|f(y)| \leq K$ for every y. We write

$$\begin{aligned} & E(f(x+B_t)1_{[x-M,x+M]^c}(x+B_t)) \\ \leq \quad & KP(|B_t| \geq M) = 2K\int_{-\infty}^{-M} \frac{1}{\sigma\sqrt{2\pi t}} \exp(-\frac{|y|^2}{2\sigma^2 t})dy \\ = \quad & 2K\int_{-\infty}^{-M/\sigma^2 t} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})dy = 2K\Phi(-\frac{M}{\sigma^2 t}) \leq \frac{1}{1000}. \end{aligned}$$

where Φ is the repartition function of the standard normal distribution -which is tabulated. So we look in the tables for the value of η for which $\Phi(-\eta) = 1/2000K$ and take $M = \sigma t\eta$.

Another point which is to be stressed is the following. We want to compute the solution u of the heat equation in **one point**, precisely in (t,x). If we perform the analytical algorithm (finite differences for example) we are **obliged** to compute, in the intermediary steps, the approximative values of the solution in the points of a whole time-space grid (t_k, x_p). This may be considered as an advantage because we obtain, as a free by-product of the algorithm, the whole surface of the solution. And this advantage is crucial in some "non-linear problems" (we do not discuss here such problems). But in the linear problem at hand this is "optional" . If we are able to do it, this is great, but we are not obliged to do it if we want just the value in (t,x). And this is what the Monte Carlo method is doing - computing directly the solution in (t,x) without having to compute this solution on a whole time-space grid. If we are in dimension one ($x \in R$) then this advantage is not essential - and anyway in dimension one everyone will use a deterministic algorithm as finite differences which works much better then the Monte Carlo method. But if $x \in R^{30}$ then the space grid will be in dimension 30 and we have seen that in such a grid we need a huge amounts of points. Consequently the task of computing the solution on the whole grid becomes out of rich - this is the frame in which one has to give up the deterministic method and use Monte Carlo. And here the fact that we may compute directly the solution, without having to use the grid approximation, becomes crucial.

♦ **The tree method**. This method is closed to the finite differences method. We think that we have a random walk in the time grid: at time t_k we are in x_p, we

toast a coin and we go with probability $\frac{1}{2}$ in x_{p-1} and with probability $\frac{1}{2}$ in x_{p+1}. This may be interpreted as a tree: at time $t_0 = 0$ we have just one point $x = x_0$, at time $t_1 = h$ we have two points (lives) $x_{-1} = x_0 - \delta$ and $x_1 = x_0 + \delta$, at time t_2 we have the points $x_{-2} = x_{-1} - \delta, x_{-1}, x_0 = x_{-1} + \delta = x_1 - \delta, x_1, x_2 = x_1 + \delta$ and so on. This is the tree. Now we recall that the condition which is necessary in order that the random walk converges to the Brownian motion (and so does not blow up and does not vanish) is that time and space are in a good relation, namely $\delta = \sqrt{h}$. So if we take $h = \frac{1}{n}$ we have to take $\delta = \frac{\sigma}{\sqrt{n}}$. This is different from the finite differences method in which we take $\delta = h = \frac{1}{n}$. We denote now by $\overline{X}(t_k)$ the random walk. Since the random walk approximates the Brownian motion we have

$$u(t,x) = E(f(x + B_t)) \sim E(f(\overline{X}(t_n))$$

where $t_n = \frac{[nt]}{n}$. We compute now the above quantity using the so called **Dynamical Programing Principle.** This principle amounts, in probabilistic language, to take conditional expectations. Denote

$$v_1(x) = E(f(\overline{X}(t_n) \mid \overline{X}(t_{n-1}) = x).$$

Then

$$E(f(\overline{X}(t_n))) = E(E(f(\overline{X}(t_n) \mid \overline{X}(t_{n-1}))) = E(v_1(\overline{X}(t_{n-1}))).$$

This means that we have transformed an expectation concerning $\overline{X}(t_n)$ into an expectation concerning $\overline{X}(t_{n-1})$. So we make a step down in the tree. The price is that we have to replace the function $v_0 = f$ by the function $v_1(x) = E(v_0(\overline{X}(t_n) \mid \overline{X}(t_{n-1}) = x)$. We will do this n times. So we define recursively

$$(A_{k+1}) \quad v_{k+1}(x) = E(v_k(\overline{X}(t_{n-k}) \mid \overline{X}(t_{n-(k+1)}) = x)$$

and write

$$\begin{aligned} E(f(\overline{X}(t_n))) &= E(v_0(\overline{X}(t_n))) = E(E(v_0(\overline{X}(t_n) \mid \overline{X}(t_{n-1}))) = E(v_1(\overline{X}(t_{n-1}))) \\ &= E(E(v_1(\overline{X}(t_{n-1}) \mid \overline{X}(t_{n-2}))) = E(v_2(\overline{X}(t_{n-2}))) \\ &= ... \\ &= E(v_n(\overline{X}(t_0))) = v_n(x_0). \end{aligned}$$

So, if we are able to compute $v_n(x_0)$ we have in fact computed $E(f(\overline{X}(t_n)))$. This is done recursively: at step 1 we know the initial condition $v_0(x_p) = f(x_p), p = -n, ..., n$ and we "compute" $v_1(x_p), p = -(n-1), ..., n-1$ according to (A_1). Then we know $v_1(x_p), p = -(n-1), ..., n-1$ and we compute $v_1(x_p), p = -(n-2), ..., n-2$ according to (A_2), and so on.

Now the question is how to perform numerically (A_{k+1}). Note that $\overline{X}(t_{n-k}) \in \{x_{-k}, ..., x_k\}$ and note also that if $\overline{X}(t_{n-k}) = x_p$ then we have $\overline{X}(t_{n-(k+1)}) = x_{p-1}$ or x_{p+1} (one arrives in x_p coming from x_{p-1} or x_{p+1}). So we have

$$\begin{aligned} (A_{k+1}) \quad v_{k+1}(x_p) &= E(v_k(\overline{X}(t_{n-k}) \mid \overline{X}(t_{n-(k+1)}) = x_p) \\ &= \frac{1}{2}(v_k(x_{p+1}) + v_k(x_{p-1})). \end{aligned}$$

This is the tree algorithm. As the finite difference method, it is a "grid method" and works very well in low dimension, but is useless in high dimension.

The tree method presented above in connection with the random walk is a good frame in which one may see what a "diffusion process" represents. Recall that the pollen particle which is in position x_p at time t_n will move at the right or at the left with probability $\frac{1}{2}$ and then, at time t_{n+1} will be in the position x_{p-1} or x_{p+1}. This means that $\overline{X}(t_{n+1}) = \overline{X}(t_n) + \Delta_k$ where Δ_k is a random variable which is independent of the past - in particular of $\overline{X}(t_n)$ and takes values $\pm\delta$ with probability $\frac{1}{2}$. We insist on the fact that the movement described by Δ_k is independent of the starting point x_p. But one may think that this is not true and in fact the point in which the process is located at time t_n has an influence on the movement - and in fact this is generally the case. For example there are some regions in which water is more quite and other regions in which the water moves more - and consequently the impulse given to the pollen particle will be stronger. As a consequence the amplitude of the movement will be more important. In order to describe this type of phenomenons one has to consider a coefficient which multiplies Δ_k and which represents the amplitude of the motion which is specific to the point x_p. So our dynamics are now described by the more general equation

$$\overline{X}(t_{n+1}) = \overline{X}(t_n) + \sigma(\overline{X}(t_n))\Delta_k .$$

From an analytical point of view this amounts to replace the heat equation with constant coefficients by an equation which state dependent coefficients:

$$(H) \quad \begin{array}{rcl} \dfrac{\partial u}{\partial t}(t,x) & = & \dfrac{\sigma^2(x)}{2}\dfrac{\partial^2 u}{\partial x^2}(t,x), \quad t>0, x\in R, \\ u(0,x) & = & f(x). \end{array}$$

From a probabilistic point of view this amounts to replace the Brownian motion by a more complicated process which is called diffusion process. Such processes appear as solutions of so called Stochastic Differential Equations. We discuss this in a following section.

Another point which is to be discussed is the "multi-dimensional" case. In fact the pollen particle moves in R^2 and not in R^1 and so the position is described by $x = (x_1, x_2)$. But the random walk and the Brownian motion have been presented as one dimensional motions. In fact, in the two dimensional case on has to assume that each component moves as an independent Brownian motion. So instead of one Brownian motion B_t we have two **independent** Brownian motions B_t^1 and B_t^2 and each of the variables x_1 and x_2 moves according to one of these Brownian motions. Since for each fixed t, B_t^1 and B_t^2 are two independent Gaussian random variables, the couple $B_t = (B_t^1, B_t^2) \in R^2$ is a two dimensional Gaussian variable so all we said before extends to the multi-dimensional case. In all generality an $d-$dimensional Brownian motion is a process $B_t = (B_t^1, ...B_t^d)$ which takes values in R^d, is continuous homogenous and has independent increments of gaussian law. The components $B_t^1, ...B_t^d$ are supposed to be independent. So the density of the law of B_t is given by the

$d-$dimensional Gaussian density: Now if we consider a general process $X_t = (X_t^1, ..., X_t^d)$ with $X_t^i = x_i + \sigma_{ii} B_t^i + b^i t$ then the components $X^1, ..., X^d$ are independent and the density of the law of X_t is given by

$$\begin{aligned} E(f(X_t)) &= \int_{R^d} f(y) p_t(x,y) dx \quad with \\ p_t(x,y) &= \frac{1}{(2\pi t)^{d/2} \Pi_{i=1}^d \sigma_{ii}} \exp(-\sum_{i=1}^d \frac{(x_i - y_i - b^i t)^2}{2\sigma_{ii}^2 t}) \\ &= \prod_{i=1}^d \frac{1}{\sigma_{ii}\sqrt{2\pi t}} \exp(-\frac{(x_i - y_i - b^i t)^2}{2\sigma_{ii}^2 t}). \end{aligned}$$

We have denoted σ_{ii} instead of the more natural notation σ_i because general one considers a process of the form

$$X_t^i = x_i + \sum_{j=1}^d \sigma_{ij} B_t^j + b^i t, \quad i = 1, ..., d.$$

In this case X_t is still a Gaussian random variable but the components are no more independent. We have a Gaussian vector of covariance matrix $\sigma = (\sigma_{ij})_{i,j=1,...,d}$ that is $E(X_t^i X_t^j - E(X_t^i)E(X_t^j)) = \sigma_{ij}$.

Let us come back (just for notational simplicity) to independent components and also assume for simplicity that $b^i = 0$. Put $X_t(x)$ instead of X_t in order to notice that the starting point is x and denote $u(t,x) = E(f(X_t(x))$. The same computations based on the density $p_t(x,y)$ as in the one dimensional case show that u solves the $d-$dimensional heat equation

$$\begin{aligned} (H) \quad \frac{\partial u}{\partial t}(t,x) &= \sum_{i=1}^d \frac{\sigma_{ii}^2}{2} \frac{\partial^2 u}{\partial x_i^2}(t,x), \quad t > 0, x \in R, \\ u(0,x) &= f(x). \end{aligned}$$

Notice that partial derivatives come on here - and this is why such an equation is called a Partial Derivatives Equation (PDE).

If one goes a step further and make the computations in the more general frame described above (the covariance matrix σ appears and the coefficients $b^i, i = 1, ..., d$ are no more null) then one finds that $u(t,x) = E(f(X_t(x))$ solves the PDE

$$\begin{aligned} (H) \quad \frac{\partial u}{\partial t}(t,x) &= Lu(t,x), \quad t > 0, x \in R, \\ u(0,x) &= f(x) \end{aligned}$$

where L is the second order differential operator

$$Lu(t,x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^d b^i \frac{\partial u}{\partial x_i}(t,x) \quad with$$

$$a_{ij} = \sum_{k=1}^{d} \sigma_{ik}\sigma_{jk}.$$

Computations are a little bit more complicated but of exactly the same nature as in the one dimensional case. What is to be stressed is that the Brownian motion $\sigma_{ij}B_t^j$ gives rise to second order derivatives in the PDE while the time $b^i t$ gives rise to first order derivatives. This is because path of the Brownian motion $t \to B_t^j$ is much more irregular then the standard "time" $t \to b^i t$ - but a rigorous development of this idea supposes a much more involved theory (stochastic calculus).

Finally, in order to obtain the completely general form of the heat equation one has to assume that σ and b depend on the space variable x and the time variable t so that

$$Lu(t,x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,x)\frac{\partial^2 u}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^{d} b^i(t,x)\frac{\partial u}{\partial x_i}(t,x).$$

In fact this is the really interesting situation from the point of view of numerical calculus - in the case of constant coefficients we already have the explicit expression of the density $p_t(x,y)$ and this is sufficient in order to compute explicit solutions of the heat equation. But if the coefficients are not constant explicit solutions are not known and this is why "numerical solutions" have to be produced. Note also that in the one dimensional case numerical methods as the finite differences of the tree method are very simple and so one does not employ the Monte Carlo method, but if we consider higher dimensions then we have the grid problems mentioned in the first section and this is why we have to employ the Monte Carlo method.

3.4 Diffusion Processes

We come back to our initial example of the pollen particle moving on the water surface. We recall that the dynamics of the particle were produced by the small movements of the water - but what was specific there was that these movements were the same does not matter the position of the particle - in other words we were in an homogenous frame. We consider now the more general situation when the frame is no more homogeneous: there are some regions in which the water is very still and then the intensity of its action on the particle is very small and there are some other regions in which the this intensity is more important. It is difficult to model this phenomenons directly in continuous time-space, so we will model it by means of the random walk first. We recall the construction of the random walk. We fix n and take the time step $h = \frac{1}{n}$ and the space step $\delta = \frac{1}{\sqrt{n}}$. Then we define recursively

$$X^n(t_{k+1}) = X^n(t_k) + \frac{1}{\sqrt{n}}U_k$$

where $t_k = \frac{k}{n}$ and $U_k, k = 0, 1, ...$ are independent random variables such that $P(U_k = 1) = P(U_k = -1) = \frac{1}{2}$. We interpret this as follows: at time t_k the particle is in the position $X^n(t_k)$ and then moves to the left or to the right ($U_k = \pm 1$) with probability $\frac{1}{2}$ and with a step $\delta = \frac{1}{\sqrt{n}}$. The "intensity" of the action on the particle is here 1. If we want to consider another intensity say σ we have to put $\frac{\sigma}{\sqrt{n}}U_k$ and if we want to include some "drift" action we have to add $\frac{b}{n}$. So we obtain

$$X^n(t_{k+1}) = X^n(t_k) + \frac{\sigma}{\sqrt{n}}U_k + \frac{b}{n}.$$

This random walk will no more converge to the standard Brownian motion but to the generalized Brownian motion $X(t) = X(0) + \sigma B_t + bt$. The arguments are the same (the CLT). But we are still in an space-homogenous situation because the intensity of the action on the particle is described by σ and b and these are constants which does not depend on the position of the particle. If we want to include the non-homogeneous frame we have to consider instead of a constant σ a function $\sigma(x)$ which represents the intensity in the point x. Then our scheme becomes

$$X^n(t_{k+1}) = X^n(t_k) + \frac{1}{\sqrt{n}}\sigma(X^n(t_k))U_k + \frac{1}{n}b(X^n(t_k)).$$

This is the random walk which modelize the dynamics of the particle in an inhomogeneous frame. This is also known as the Euler Scheme based on the Bernoulli sample $U_k, k = 0, ...,$ And this is an approximation for a "diffusion process" which is a generalization of the generalized Brownian motion in the sense that instead of the constants σ and b which appear there we will have some functions - but this generalization has to be done carefully and so in fact, for the moment we have the approximation but do not have the limit in noway.

I will try now to explain how diffusion processes are constructed (but I do not go up to the end). First of all I will replace the Bernoulli distributed random variables U_k by some standard normal distributed, independent, random variables V_k so that the new Euler Scheme is now

$$X^n(t_{k+1}) = X^n(t_k) + \frac{1}{\sqrt{n}}\sigma(X^n(t_k))V_k + \frac{1}{n}b(X^n(t_k)).$$

This is known as the Euler scheme based on Normal distributed random variables. I may legitimate such a new choice by the model hypothesis itself - the water pushes the particle with a force which is normal distributed instead of Bernoulli - who may know exactly what water is doing....Another more mathematical motivation would be that the limit is the same does not matter the law of U_k - and this comes from the fact that the limit is the same in the Central Limit Theorem does not matter the law of the random variables which are involved. Anyway, now on I take V_k. I consider now a Brownian motion B_t and I replace $\frac{1}{\sqrt{n}}V_k$ by $B(t_{k+1}) - B(t_k)$. Note that these two random variables

have the same law - they are normal distributed with mean zero and variance $\frac{1}{\sqrt{n}}$. Note also that $X^n(t_k)$ is a function of $V_0, ..., V_{k-1}$ - this is true just by the recursive construction given above. So there exists a complicated function f_k such that $X(t_k) = f_k(V_0, ..., V_{k-1})$. Now if I change V_k by another random variable which has the same law, the law of the functional remains the same. So the law of $X(t_k)$ constructed with $\frac{1}{\sqrt{n}}V_k$ or with $B(t_{k+1}) - B(t_k)$ is the same. We conclude that as long as we discuss convergence in law (and this is the case here) we may define

$$X^n(t_{k+1}) = X^n(t_k) + \sigma(X^n(t_k))(B(t_{k+1}) - B(t_k)) + \frac{1}{n}b(X^n(t_k))$$

where B_t is a given Brownian motion.

This is the first step of our reasoning. In a second stage we use the recurrence formula and write

$$\begin{aligned} X^n(t_{k+1}) &= X^n(t_0) + \sum_{i=0}^{k} \sigma(X^n(t_i))(B(t_{i+1}) - B(t_i)) + \frac{1}{n}\sum_{i=0}^{k} b(X^n(t_i)) \\ &= x_0 + \sum_{i=0}^{k} \sigma(X^n(t_i))(B(t_{i+1}) - B(t_i)) + \sum_{i=0}^{k} b(X^n(t_i))(t_{k+1} - t_k). \end{aligned}$$

This is known as the Euler Scheme bsed on the Brownian increaments. The advantage of this way of writing things is that Riemann sums appear. It is clear that $\sum_{i=0}^{k} b(X^n(t_i))(t_{k+1} - t_k)$ is a Riemann sum for $\int_0^t b(X^n(s))ds$. For the first sum things are much less clear. If $t \to B_t$ is differentiable (but it is not!) one would have $B(t_{i+1}) - B(t_i) = \int_{t_i}^{t_{i+1}} B'(s)ds$ and then

$$\sum_{i=0}^{k} \sigma(X^n(t_i))(B(t_{i+1}) - B(t_i)) \sim \int_0^t \sigma(X_s^n)B_s' ds.$$

But this is not rigorous. And this also shows that there will be ruther difficult to give a sense to such an integral. But one may define a new type of integral, called "stochastic integral" and then

$$\sum_{i=0}^{k} \sigma(X^n(t_i))(B(t_{i+1}) - B(t_i)) \sim \int_0^t \sigma(X_s^n)dB_s$$

where in the right hand side you have the notation for a stochastic integral. If you accept this you will expect that X_t^n converges to the solution of

$$X(t) = x_0 + \int_0^t \sigma(X(s))dB_s + \int_0^t b(X(s))ds$$

where the stochastic integral appears. Note that the Euler scheme was an "algorithm" in the sense that one computes recursively $X^n(t_k)$. But now we

have an equation because X appears in both sides of the equality - this is a so called Stochastic Differential Equation. So it is not a priory clear that a solution of this equation exists and if yes, that it is unique. One may prove that if the coefficients σ and b are Lipschiz continuous then there exists a unique solution of this equation and such a solution is called Diffusion Process. One is also able to prove that the two first Euler schemes (based on the Bernoulli random variables U_k or on the normal distributed random variables V_k) converge in law to the diffusion process. If we put $B(t_{i+1}) - B(t_i)$ in the Euler Scheme then the convergence is even stronger (in L^p for example). There is a huge amount of work on this topic and you will find it in any book of stochastic calculus under the title "approximation of diffusion process". It is motivated by numerical applications as we will see in a moment.

The link between diffusion processes and Partial Differential equations is the same as in the case of the generalized Brownian motion - but the proof is more involved now and employes stochastic calculus. Nevertheless the following result, known as the Faynmann Kac formula, is true. Define

$$u(t,x) = E(f(X(t,x))$$

where $X(t,x)$ is the diffusion process which starts from x - that is the solution of the above Stochastic Differential Equation with $x_0 = x$. Then u is the unique solution of the second order differential equation

$$\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x) \quad u(0,x) = f(x).$$

Recall that in the case of the Brownian motion $X(t,x) = x + B_t$ and we know the density of the law of this random variable. This is why we proved very easily the statement: we just checked that the Gaussian density verifies the hear equation. But now the density of the law of $X(t,x)$ is no more known (and may even not exist) and this is why the previous argument breaks down. This is also why it is necessary to use numerical methods in order to compute the solution u - because we have no explicit formula for it.

Now we are really in business with the Monte Carlo method. Let us describe how this method works in this frame. We want to compute $u(t,x)$ for some fixed t and x. We proceed as follows:

Step 1. We represent the solution by means of the diffusion process, that is $u(t,x) = E(f(X(t,x))$.

Step 2. We have now a quantity which is represented as an expectation so we may use the Monte Carlo method in order to compute this quantity. In order to do it we have to simulate a sample of the random variable at hand, that is of $f(X(t,x))$. In order to do it we have to know the law of this random variable - but we do not know it. In fact, if we knew the law, than we would also know the mean of the random variable - this appears as rather paradoxal. But the answer is the following. We approximate the diffusion process by the Euler Scheme (one of the variants presented above) and so we have

$$u(t,x) = E(f(X(t,x))) = E(f(X^n(t,x))) + \varepsilon_n$$

where ε_n is an error that I will discuss in a moment. Now we will compute $E(f(X^n(t,x)))$ instead of $E(f(X(t,x)))$. Since we are able to simulate $f(X^n(t,x))$ the paradox is now solved.

Step 3. We simulate $X^n(t_k,x), k=1,...,n$ in the following way. We want to produce M independent copies of $X^n(1,x)$, that we denote by $X^{n,1}(1,x),....,X^{n,M}(1,x)$. We call a random number generator and we produce a long sequence of independent random variables $V_p, p=1,....$ of standard normal law (or Bernoulli if you want). Then we re index these random variables as $(V_0^1,...,V_n^1),....,(V_0^M,...,V_n^M)$. We will use the random variables $(V_0^i,...,V_n^i)$ in order to construct $X^{n,i}(1,x)$. This is done by the recursive algorithm

$$\begin{aligned} X^{n,i}(0,x) &= x \\ X^{n,i}(t_{k+1}) &= X^{,in}(t_k) + \frac{1}{\sqrt{n}}\sigma(X^{n,i}(t_k))V_k^i + \frac{1}{n}b(X^{n,i}(t_k)). \end{aligned}$$

Now we have our sample $X^{n,1}(1,x),....,X^{n,M}(1,x)$ and we may compute

$$E(f(X^n(t,x))) = \frac{1}{M}\sum_{i=1}^{M} f(X^{n,i}(1,x)) + \delta_n$$

where δ_n is the error in the Monte Carlo method.

This is the algorithm. Let us make several comments..

First of all: we replace the "unknown" random variable $X(t,x)$ by some "known" random variable $X^n(t,x)$. This random variables is known because it appears as a complex but explicit functional of a finite number of standard normal distributed random variables - and we are able to simulate such random variables. So we have the following chain: 1. We produce uniform distributed random variables by the congruential method. 2. We extend the procedure by some clever transformation and we produce Gaussian random variables. 3. We make some complicated puzzles with Gaussian random variables (I mean the Euler Scheme) and we produce sufficiently accurate approximations of diffusion processes. This is the chain every time that we want to use the Monte Carlo method: we have to approximate the random variable at hand (which we can not simulate because we do not know the law, because if we know the law we also know the mean value) by some function of a finite number of random variables which we may simulate.

What about errors? The error δ_n is controlled by the Central Limit Theorem and we will have all the "reduction of variance" problems we mentioned in the firs section.. The approximation error ε_n may be of order $\frac{1}{\sqrt{n}}$ in very bad situations but under reasonable hypothesis one may see that it is of order $\frac{1}{n}$.

Finally we discuss the following strange thing. Our initial modelization was done using the random walk and so our initial model was directly the Euler scheme. Afterwards we pretended that we may introduce the diffusion processes by passing to the limit with $n \to \infty$, and so we replaced a discrete time model by a continuous time model. Finally we stressed that we may compute nothing directly for the continuous time model and so we have to take some discrete

time approximation (the Euler scheme) in order to perform our computations. And this introduced an error ε_n. But why not using directly the initial Euler Scheme which we have used for the modelisation? Why are we obliged to pass through the continuous time model? Just to make pleasure to mathematicians? This would be a very convenient reason but there is a deeper one. Think once again to the pollen particle and to water pushing it. What do you think: water employes Bernoulli random variables or standard normal distributed random variables in order to push the pollen particle? In other words, which is the structure of the action which is going on? We do not know. The only think we know is that the successive actions at times $t_k, k = 0, ..., n$ are independent and identically distributed. Of course this is a model hypothesis but this hypothesis is common sense while saying that water employes Bernoulli random variables is non-sense. And the striking thing is that doesn'tt matter the specific structure (law) of the infinitesimal actions, the limit is has always the same law - this is the specificity of the Central Limit Theorem which gives a normal law does not matter the law of the random variables at hand (in our frame the limit law is the law of the diffusion process which, if the coefficients are not constants, is not normal. But the phenomenons is the same). We conclude that in our problem there is not only one approximation but we have two approximations: We have a physical phenomenons which is in discrete time and which involves a multitude of infinitesimal actions.. We approximate this discrete phenomenons by a continuous time phenomenons which is the diffusion process. The diffusion process does not represent the real phenomenons but just an ideal model for it - and this ideal model represents somehow an approximation of the very complex real physical phenomenons.. The reason of being of this approximation is that we do not know the specific structure of the infinitesimal actions which produce the real phenomenons and then we have to use the Central Limit Theorem in order to "forget" this unknown specific structure. Once we have the ideal model - in continuous time - we produce a second approximation by the Euler scheme. But this time we know the law of the infinitesimal actions V_k because it is us who decide on which law we will use in our approximation. So the second Euler scheme is known and is different from the initial one which was unknown.

3.5 What does people compute in mathematical finance problems?

An introduction to mathematical finance requires an important investments. First of all one needs a consistent background concerning stochastic processes, stochastic calculus and partial differential equations. On the other hand the understanding of the financial problems themselves take a little bit time. So a serious introduction in this topic is out of rich in our frame. But on the other hand mathematical finance provides interesting and concrete problems in which the Monte Carlo method is intensively used by practicients. So it is very interesting for us to have an idea about what these people want to compute, even if the theory which lids to these problems remains beond our control. This is why I will tell you **a story** rather then trying the impossible task of honestly

covering the subject.

A market model is described by two types of assets. First of all a riskless asset S_t^0 which represent the quantity of money you have in the bank. It is supposed that the bank account has a deterministic rate of return which is given by a constant and known interest rate $r > 0$. This means that if at time $t = 0$ you have a quantity $S_0^0 = s_0$ of money in the bank account then at time t your bank account is $S_t^0 = s_0 e^{rt}$. In fact things are more complicated and there exists a whole theory of the interest rates but we keep here in the simplest possible situation (and this is general the frame in which people discuss the "option market models"). Except the riskless asset there is a second asset which is subject to a random evolution and so to risk - for example a given stock which is quoted in the stock exchange market. Let S_t be the price in the market of the stock at time t. Generally one considers a finite number of stocks and not a single one, but we keep for the moment in the one dimensional setting so S_t is a real strictly positive number. This is subject to random fluctuations so in fact we have a random process, and at each fixed time t, a random variable $S_t(\omega)$. The question is now how to modelize the evolution of this process, which is the natural law that we have to expect for S_t? Saying that S_t is just random is too poor and does not permit to make computations related to the price evolution. Let us see which are the natural model hypothesis that we may assume. First of all we assume that the evolution of the price is **continuous,** so $t \to S_t$ is supposed to be a continuous function. The second hypothesis is that if we are at time t and look to the variation of the price $S_{t+h} - S_t$ then this variation is independent of the past evolution of the price, that is of all $S_s, 0 \leq s \leq t$. This hypothesis may be subject to some objections: for example one would expect that, if the price is very high then it has a tendency to go down, or someone else would say that if in the previous month the prices went up, then they will continue to go up - something as an inertia of the market. Such economic considerations are beond my competency but one may see in any mathematical finance book that the independence hypothesis is universally accepted. This is somehow a minimal hypothesis in order to construct a model which is sufficiently simple and robust - and consequently permits to "compute things". But there is a more subtle think to be noted. In order to understand this let us think to a stock which has the price $S_t = 100$ euros at time t and $S_{t+h} = 110$ euros. Then the increment is $S_{t+h} - S_t = 10$ euros. Think now that at time t the price was $S_t = 10$ euros and at time $t + h$ the price is $S_{t+h} = 20$ euros. Then the increment is still $S_{t+h} - S_t = 10$ euros but in the first case we have a variation of 10% and in the second case the price is double! And it is more credible that the price moves with 10% for a short period, say $h =$ one month, then with 100%. And so it appears that the size of the increment $S_{t+h} - S_t$ depends on the size of the price S_t - and then it is not independent. We conclude that there is a problem concerning the independence hypothesis. But the answer is already in the example: it is not significant to look to the increment of the price itself, but to the "relative increment" that is to $(S_{t+h} - S_t)/S_t$. This is the quantity which is significant for the performance of the stock, and not the absolute value of the increment. In the first case 1 euro gives 1.1 euros and

in the second case 1 euro gives 2 euros, so the performance in not the same. This lids us to the correct hypothesis: for each $t_0 < t_1 < ...t_n$ the relative increments $(S_{t_{i+1}} - S_{t_i})/S_{t_i}, i = 1, ..., n-1$ are independent. So we assume that $t \to S_t$ is a process with **independent relative increaments**. Some simple considerations show that this is the same thing as $t \to X_t := \log S_t$ is a process with independent increments.. Finally we assume homogeneity and of course we do it in terms of relative increments also. This amounts to say that given $h > 0, (S_{t+h} - S_t)/S_t$ have the same law for every $t \geq 0$. Passing to the logarithm this is equivalent with the fact that $X_{t+h} - X_t$ have the same law for every t. So our hypothesis is

$$(H) \quad \text{The process } X_t = \ln S_t \text{is a continuous process which is homogenuos and has independent increaments.}$$

We add to this the more technical and cautionable hypothesis that $EX_t^2 < \infty$ - this just means that we do not expect that prices are too high - and we accept it.

Now if we make these "qualitative" hypothesis, the theorem given in a previous section guarantees that there exists a standard Brownian motion B_t and some constants x_0, σ, b such that $X_t = x_0 + \sigma B_t + bt$. And consequently

$$(B - S.Model) \quad S_t = e^{X_t} = e^{x_0 + \sigma B_t + bt} = s_0 e^{\sigma B_t + bt} \quad with \quad s_0 = e^{x_0}.$$

This is the celebrated **Black-Scholes model**. The miracle here is that we started with some "random process" S_t on which nothing was known and, after making some rather natural calitative hypothesis we find a model in which everything is computable. Anyway note that there are still two parameters which are not known: the coefficient σ which is called volatility in mathematical finance and the coefficient b. For reasons that I can not explain here b is not important - it just disappears in the computations which are done in mathematical finance. At the contrary one needs to know the volatility σ and there is no theoretical way to determine it. In some way one would say that "it is the market who determine the volatility" and this means that people trading in the market have a certain dealing about the fact that the price would move very fast - and the σ will be large - or will be rather stable, and then σ is rather small. But this dealing is difficult to determine and measure and it is not the dealing of one person but of huge and unknown aggregate of people. So there is no way to have a clear idea about what σ has to be. But people have to know σ if they want to compute something. There is a huge variety of more or less empirical methods for getting an idea about what a reasonable value of σ may be - and all these methods entered in the so called "calibration theory". But in our frame we assume that σ is known and so the law of S_t is completely determined.

Let us come back to history. The first who observed a phenomenons which was of the "same nature" as the Brownian motion was Brown. But he did no computations related to it. The first who is known to have used the Brownian motion in a physical model and who has done computations with it was Einstein

in a paper from 1905 - but I do not know more on this paper. But before Einstein, around 1900 it was Bachelier who used the Brownian motion exactly in the frame here - to give some market model. Unfortunately his work did not have much echo. Maybe because the market activity was not sufficiently developed at that time but for sure because he has done an error - he tried to model the stock price itself as a Brownian motion and not the logarithm of the price. And of course this was not satisfactory - in particular the price may be negative with non-null probability if it is a Brownian motion. It is only in the years $70's$ that Black and Scholes, independently of the work of Bachelier, introduced the model presented here. Meantime, between the years 30 and 70 a huge amount of work has been done concerning the Brownian motion and a very consistent theory concerning Markov processes, diffusion processes and stochastic calculus appeared. So the theoretical basis needed in mathematical finance already existed at that time. Important contributions to the foundation of this theory has been done by Winner in the years 30 and this is why the Brownian motion is also called the Winner process.

Let us make a step further and come to the notion of "**options**" and more precisely "European options". It is somehow an insurance contract. Let us give an example in order to understand what it is. You have an Italian company which produces cars and these cars will be sold in one year. But the price of a car may go down and the company wants to insure itself against a very law price. Say that cars are sold now a day at 10500 euros a price and if the price will be less then 10000 euros in one year then the company has serious troubles. So they want an insurance against such an event - this means that they want that some bank pays what is lost if the price goes beond 10000. In this case the stock is the price of the car. So $S_0 = 10500$ today and evolutes according a stochastic process S_t. The company wants to sell the car in one year, say $T = 1$ year. This is the "time horizont or the maturity. In one year the price of the car will be S_T and this is a random variable and no one knows today this future price. So the company wants that the bank pays to him the amount $10000 - S_T$ if this sum is positive - which means that the price of a car is lower then 10000. In the case that $S_T > 10000$ then there is no problem and nothing happens. This is called a put option, $K := 10000$ is called the strike of the option and $(K - S_T)_+$ - which represents the amount of money that the bank pays to the company at time T - is called the payoff or contingent claim. Of course the company has to pay a certain sum to the bank for this insurance contract and this is the price of the option. In order to understand why such a contract is called an option we have to put it other way. The contract says that the company has the right to sell to the bank a car at time T, at price $K = 10000$. If the price S_T is lower then 10000 the bank is obliged by this contract to buy the car at this price and this amounts to payment of $10000 - S_T$. If the price is larger then 10000, say 10020 then the company will not sell the car to the bank but in the regular market. So the company is not obliged to sell - this is an "option". If $S_T < 10000$ the company exercises the option and if not, she does not exercise. Note that the company has the right to exercise (or not) his option at the time T exactly - this restriction is specific to "European options".

If the company has the right to exercise at any time $t \leq T$ then such an option is called an "American option". In the specific case we present here this is an European put option and if one may exercise at any moment before T this is an American put option. This gives the right to sell. If the option gives the right to buy, then we have a call option. For example the company is working in euros but in six month they have to make a deal in dollars. So they need to buy one million dollars in six month and they want to insure against a high price of the dollar. Now the stock is just the dollar. Nowadays one dollar is 0.97 euros and one wants to be sure not to have to pay more then 1 euro for one dollar. So now the strike is $K = 1$ and the maturity is $T =$ six month. The payoff is $(S_T - K)_+$ where S_T is the price of a dollar in euros in six month. Once again this is an option: in six month, if the dollar is 1.01 the company exercises the option and buys the dollars to the bank for 1 dollar for 1 euro. If the price is 0.98 they buy the dollar on the market - and so do not exercise the option. This operations amounts to the fact that the bank pays to the company the sum $(S_T - 1)_+$ at time T. Of course there is a price for this option - a certain price that the company pays at time $t = 0$. And the first problem is to find the price which will be convenient both for the company and for the bank.

Now the question is the following: how to determine the price of an European option, and in order to be more specific, of an European call option. So we have to pay a deterministic sum P_0 at time $t = 0$ in order to obtain the stochastic sum $(S_T - K)_+$ at time T. Which is the right P_0 to be payed?

I will not answer to this question - it seems to me that this is too involved and needs some more detailed discussions concerning the so called arbitrage theory in mathematical finance. So I just give the formula which is used and live for a serious mathematical finance course the motivation of this formula. One has

$$P_0 = e^{-rT} E^*((S_T - K)_+).$$

Note that in this formula we consider an expectation with respect to the probability P^* (this is the sense of E^*). But who is P^*? It is the so called "risk neutral probability" which plays a central part in mathematical finance. So from the beginning we face the following problem: the price of an European option is computed as an expectation (and this is great because we may use the Monte Carlo method) but this expectation is not taken with respect to the "natural" probability measure but with respect to another probability.... and this seems bad, because we loose all the advantage we have won by the Black-Scholes model: in this model we know, up to the constants σ and b, the law of S_T but if we change of probability, then this variable changes of law and so once again we know nothing. Here comes on the arbitrage theory which not only explains why the price is computed as an expectation under the "risk neutral probability" but also gives the law (dynamics) of the price under this new probability. It is proved that

$$(B - S.Model^*) \quad S_t = s_0 e^{\sigma B_t^* - \frac{\sigma^2}{2} t} \quad with \quad s_0 = e^{x_0}$$

where B^* is a Brownian motion under P^*. So the price formula becomes

$$P_0 = e^{-rT}E^*((s_0 e^{\sigma B_T^* - \frac{\sigma^2}{2}T} - K)_+) = e^{-rT}\int_{-\infty}^{\infty}(s_0 e^{\sigma y - \frac{\sigma^2}{2}T} - K)_+ \frac{1}{\sqrt{2\pi T}} e^{-\frac{y^2}{2T}} dy.$$

And so the price is completely known (note that b does not appear in this formula - it depends just on σ). Some elementary computations (exercise) permit to express this integral in terms of the repartition function of the standard normal distribution:

$$\begin{aligned} (B - S.Formula) \quad P_0 &= s_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) \quad where \\ d_1 &= \frac{\log\frac{s_0}{K} + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \quad and \quad d_2 = d_1 - \sigma\sqrt{T}. \end{aligned}$$

This is the celebrated Black-Scholes formula for the call option price. An analogues formula holds true for the put option price. The success of this formula is due to its simplicity - once we know the price of the stock at time $t = 0$, we may compute the price of the option just using a table for the normal distribution repartition function. But then what the Monte Carlo may still do here? In the case of a call option on a single stock, nothing - the formula gives everything. But in more complicated problems "closed formulas" are not available and then one has to use approximation methods. Suppose for example that the payoff is not $\phi(x) = (x - K)_+$ but a general function ϕ. Then the pricing formula is the same

$$P_0 = e^{-rT}E^*(\phi(s_0 e^{\sigma B_T^* - \frac{\sigma^2}{2}T})) = e^{-rT}\int_{-\infty}^{\infty}\phi(s_0 e^{\sigma y - \frac{\sigma^2}{2}T})\frac{1}{\sqrt{2\pi T}} e^{-\frac{y^2}{2T}} dy$$

but now we have no more closed formula in order to compute this quantity. And we use the Monte Carlo method.

We finish this section with the link between the pricing formula and $DPE's$. You can see that the price P_0 is a function of the initial value of the price $x = s_0$ and of the maturity T. So one may write

$$\begin{aligned} u(T, s_0) = P_0 = e^{-rT}E^*(\phi(s_0 e^{\sigma B_T^* - \frac{\sigma^2}{2}T})) = \\ e^{-rT}\int_{-\infty}^{\infty}\phi(s_0 e^{\sigma y - \frac{\sigma^2}{2}T})\frac{1}{\sqrt{2\pi T}} e^{-\frac{y^2}{2T}} dy. \end{aligned}$$

Some standard computations (change of variable + the same computations as in the case of the heat equation) show that u solves the $PDED$

$$\frac{\partial u}{\partial t}(t, x) = \frac{\sigma^2 x^2}{2}\frac{\partial^2 u}{\partial x^2}(t, x) \quad u(0, x) = \phi(x).$$

So in order to compute the price we have to solve a PDE and this may be done by the finite differences method or by a tree method as well. In practice, as long as people work in low dimension (one or two) they employ this type of methods, but if we are in higher dimension we have to employ the Monte Carlo method. Here the dimension is the dimension of the stock - up to now we are in dimension one because we have a single stock, but if we work on the CAC 40 then we will have 40 stocks which are involved in the formula.

4 References

[Dev85] L. Devroye. Non uniform random variate generation. Springer Verlag 1985.

[Nie95] H. Niederreiter. New developments in uniform pseudorandom number and vector generation. In Monte Carlo and Quasi-Monte Carlo methods in scientific computing, volume 106, Lecture Notes in Statistics, pages 87-120, Heidelberg, New York, 1995, Springer Verlag.