Monte Carlo Simulation for Jump processes. March 2021

Arturo Kohatsu Higa

Ritsumeikan University, Nojihigashi, Kusatsu, Shiga, Japan

Email address: khts00@fc.ritsumei.ac.jp

Abstract. These notes represent the graduate lectures delivered in March 2021 after the kind invitation of the organizers of the Beethoven grant titled "Sensitivity Analysis of Non-local Operators with Applications to Jump Processes".

1. Introduction

In these informal lecture notes, we will describe in a free format according to my interests a number of subjects that are linked by the common name of Monte Carlo methods, approximations and jumps. It follows the discussion through the talks.

We assume that the reader has basic knowledge of stochastic differential equations with jumps and some basics of simulation such as importance sampling and the acceptance and rejection method.

I have not tried to put the most general framework for each problem and avoided technicalities. I rather concentrated on getting certain ideas across in the simplest format possible. For general definitions and results please look at the references. The list of references should also be looked for when searching for extra references in applications.

I also have clearly taken the path of informality in exchange of serious proof format in order to cover a large amount of material in a short period of time. This may give a better view of the subject to a non-expert. On the other hand, it may leave the reader wondering how the actual proof is carried out. For this, look into the references.

Some of the references quoted at the end of the article are historical and some others are there for further study. I will try to comment on these throughout the text. A lot of references with some history about Multilevel Monte Carlo methods appears in the website http://people.maths.ox.ac.uk/~gilesm/mlmc_community.html

A short statement about modes of convergence. As we will deal with approximations, it is natural to think in which sense will these approximations converge. The two main modes of convergence used in simulation are L^p-norms and weak convergence. In this last sense, there are two types used for processes, one considers the limit of the approximation as processes (using e.g. any of the Skorohod topologies) and the other uses only marginal distributions. In more specialized research sometimes specific expectations are considered. In recent literature people also consider Wasserstein distances, total variation distances and with them coupling methodologies etc. All these types of convergence are interesting but in these lectures I will

concentrate on the convergence of marginals. Clearly, as one makes the norm stronger is more difficult to obtain significant results but also the simulation methodologies tend to be more robust.

In a general setting, one may say that the objective is to approximate a (potentially infinite dimensional) r.v. X using an approximation $\bar{X} \equiv \bar{X}^{\varepsilon}$ with approximation parameter $\varepsilon \to 0$. That is " $\lim_{\varepsilon \to 0}$ " $\bar{X}^{\varepsilon} = X$. We would like to measure the error of approximation in one of the following ways (they are not necessarily exclusive of each other)

$$\begin{aligned}
\mathbb{E}\left[\|X - \bar{X}\|^p\right]^{1/p} \leq & C_p \varepsilon(1), \quad \text{e.g. } \|X\| = \sup_{t \leq T} |X_t| \\
\left|\mathbb{E}\left[F(X) - F(\bar{X})\right]\right| \leq & C_F \varepsilon(2), \quad F \in \mathcal{C}_2 \\
\left|\mathbb{E}\left[f(X_T) - f(\bar{X}_T)\right]\right| \leq & C_f \varepsilon(3), \quad f \in \mathcal{C}_3 \\
\inf_{P_X = P^1, P_{\bar{X}} = \bar{P}^1} \mathbb{E}_P\left[\|X - \bar{X}\|^p\right]^{1/p} \leq & C_p \varepsilon(4).
\end{aligned}$$

Here $\varepsilon(i)$ are explicit functions of ε.

CHAPTER 1

Lecture 1: Unbiased Simulation methods for SDE in 1-dimension

1. Random Poisson measures

A (time-dependent) random Poisson measure process [1] is defined as follows:

DEFINITION 1. Let $(E, \mathcal{A}, \mu)$ be some measure space with finite measure μ. The (time-dependent) Poisson random measure process with intensity measure μ is a family of random variables $\{N(t, A)\}_{t\in[0,T], A\in\mathcal{A}}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

(1) $\forall A \in \mathcal{A}, \quad N(t, A)$ is a Poisson process with rate $\mu(A)$.
(2) If sets $A_1, A_2, \ldots, A_n \in \mathcal{A}$ are disjoint then the corresponding random processes from (1) are mutually independent.
(3) $\forall \omega \in \Omega,\ N(t, \cdot)$ is a measure on $(E, \mathcal{A})$.

In the classical simple Poisson process case one has that $E = \{1\}$ and $\mu(\{1\}) = 1$. In general, one has

$$\mathbb{P}(N(t, A) = k) = e^{-t\mu(A)} \frac{(t\mu(A))^k}{k!}.$$

EXERCISE 1. *Prove that the process $N(t, \cdot)$ is a Markov process. In fact, let $\mathcal{F}_t$, be the σ-field generated by $N(s, \cdot)$, $s \le t$. Prove that the generator given as*

$$\lim_{h\to 0} h^{-1}\mathbb{E}\left(\int f(y)N(t + h, dy) - \int f(y)N(t, dy)|\mathcal{F}_t\right) = \int f(y)\mu(dy).$$

Our goal is to show how to carry out the simulation of the points associated to the support of the random measure process $N(t, \cdot)$.

In order to carry out the simulation one may recall the following Binomial property of random Poisson measures: For $0 \le k \le n$,

$$\mathbb{P}\left(N(t, A) = k|N(t, E) = n\right) = \binom{n}{k} \frac{\mu(A)^k \mu(A^c)^{n-k}}{\mu(E)^n}.$$

[1] There are many other far more general definitions.

This means that once one knows how many points have appeared at time t in the space E their spatial distribution is determined by the measure $\mu(\cdot)/\mu(E)$.

Let us start with the simplest case: Suppose that $E = [0,1]^d$ and μ is the uniform measure. In this situation, one can propose

ALGORITHM 1. *(1) Iteratively, simulate the times $\{T_i < t, i \in \mathbb{N}\}$ for a simple Poisson process of parameter $\mu(E) = 1$.*
(2) For each time T_i simulate a point $U_i \in E$ with the uniform distribution.
(3) Then define $N(t, A) = \sum_{i=1}^{\infty} 1_{T_i \leq t} 1_{U_i \in A}$.

EXERCISE 2. *Prove that the above procedure satisfies the definition of Poisson random measure process.*

In many situations the measure μ can not be easily simulated although some kind of functional form is known. In that case, one uses the so-called "thining" algorithm which is an acceptance-rejection algorithm.

To introduce this, let us suppose that $E = \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}_+$ is an integrable function so that

$$\mu(A) = \int_A f(x)dx.$$

The methodology is to propose a function $g : \mathbb{R}^d \to \mathbb{R}_+$ which is a multivariate density function from which we can simulate. We denote by $G(A) := \int_A g(x)dx$ its associated measure.

We need to assume that there exists a constant M such that

$$\rho(x) := \frac{f(x)}{Mg(x)} \leq 1. \tag{1.1}$$

Then the proposed thining algorithm is as follows:

ALGORITHM 2. *(1) Iteratively, simulate the times $\{T_i < t, i \in \mathbb{N}\}$ for a simple Poisson process of parameter M.*
(2) For each time T_i simulate points U_i which follows the law given by the density function g.
(3) Throw a coin X_i with probability of heads $\mathbb{P}(X_i = H) = \rho(U_i)$ and accept the sample U_i if $X_i = H$. If this sample point is accepted continue with the space simulation on the next time point.
(4) Then define $N(t, A) = \sum_{i=1}^{\infty} 1_{T_i \leq t} 1_{U_i \in A} 1_{X_i = H}$.

Just to assure yourself that this methodology gives the correct law, we will compute

$$\begin{aligned}&\mathbb{P}(N(t,A)=1)\\ =&\sum_{n=1}^{\infty} e^{-Mt}\frac{(Mt)^n}{n!}n\int_{E^{n-1}\times A}\prod_{i=1}^{n-1}\left((1-\rho(x_i))g(x_i)1_A(x_i)+g(x_i)1_{A^c}(x_i)\right)\rho(x_n)g(x_n)dx_1...dx_n\\ =&e^{-t\mu(A)}\frac{(t\mu(A))^1}{1!}.\end{aligned}$$

This method is called "thinning" because one does not use all the simulated points in space. Clearly this simulation method has its drawbacks:

(1) If the inequality (1.1) that defines ρ uses a large value of M it means that one will be simulating many jump times for the Poisson process and many of them will be rejected. This implies a large computation time.
(2) The above issue can be somewhat resolved by dividing the space E into a number of disjoint subdomains (also called layers or strata). In each subdomain one can try to improve the inequality that defines ρ and then add the resulting processes in each subdomain. This is one of the reasons this methodology is called "thinning".
(3) Always under the assumption that $\mu(E)<\infty$, one can also consider an infinite number of layers by using a randomization procedure.

There are two other extensions which one has to keep in mind. The first is the time-inhomogeneous case.

EXERCISE 3. *Define the number of simulations required by the above method as a function of M and prove that there is a reduction of simulations when the function f is a density that can be simulated. That is, in the case $M=1$ and $g=f$.*

EXERCISE 4. *Design a thining algorithm in order to simulate a Poisson random measure process where the rate is given by $\mu([0,t]\times A)=\int_0^t\int_A f(s,x)dsdx$. In this sense, many authors do not consider the time dependent case as we have done here assuming that it is just a particular case of the general random Poisson measure random variable.*

Simulating spatial distributions is not always easy. For example:

EXERCISE 5. *Propose a simulation method in the case that one has the following ϵ-approximation of the α-stable measure ($\alpha\in(0,2)$) in dimension 2 which can be considered*

as the limit of the case

$$f_\epsilon(r,\theta) = \frac{1_{r>\epsilon}}{r^{1+\alpha}}\mu(d\theta)dr.$$

Here μ is a finite measure on $\theta \in (-\pi,\pi]$ and $r>0$.

Next consider a thinning algorithm to improve the previous simulation for a second value $\epsilon' < \epsilon$. Try to think how to do a simulation for a sequence $\epsilon_n \downarrow 0$.

This exercise is somewhat loose and it assumes some knowledge of stable processes in dimension 2. As this is a difficult computational task essentially due to the fact that the limit process has an infinite number of small jumps in any closed interval. For this reason, this model belongs to the class of "infinite activity" models.

Many problems can be reduced to the consideration to a functional of these small jumps such as is the case of the two dimensional α-stable process.

$$\int_{[0,t]\times(-\pi,\pi]\times\mathbb{R}_+} re^{i\theta}\left(N(ds,d\theta dr) - \frac{1_{r<1}}{r^{1+\alpha}}dsd\mu(\theta)dr\right).$$

We will come back to this subject later. Now we go back into the finite case.

One may also simulate the times first and then the space variables:

ALGORITHM 3. *(1) Iteratively simulate standard $[0,1]$ uniform random variables $\{U_i\}_{i\in\mathbb{N}}$.*

(2) Define iteratively with $T_0 = 0$,

$$\exp\left(-\int_{T_n}^{T_{n+1}}\int_E f(s,x)dsdx\right) = U_{n+1}.$$

(3) At each time T_i simulate the space variable according to the density

$$\left(\int_E f(T_i,y)dy\right)^{-1} f(T_i,x); \quad x\in E.$$

EXERCISE 6. *Prove that the two first steps of the above procedure gives the jump times of the associated Poisson random measure process. Adding the third step gives the space variable.*

1.1. Example 1: the Hawkes model. Returning to the case where the total mass is finite a class of models which has been attracting some attention are the self-excited type of models. One of the simplest models in this class is the so-called Hawkes model:

These type of models can be approached first with the solution X of the following equation

$$X_t = X_0 + N(\int_0^t \alpha(X_s)ds). \tag{1.2}$$

Here $X_0 > 0$ and N is a simple Poisson process (of parameter 1) and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$. Inductively, one can find that the solution of this equation is just another jump process. When one jump happens the clock of the simple Poisson process fastens and slows down depending of the value of $\alpha(X_s)$.

EXERCISE 7. *Propose a simulation method for X.*

The main theoretical interesting characteristic of the above model is that the "rate" at which the jumps happen is random and given by $\int_0^t \alpha(X_s)ds$.

For this reason, we need to introduce a different characterization of random Poisson measure processes using martingales.

DEFINITION 2. A point process $N : \Omega \times [0,T] \to \mathbb{N}$ is a stochastic process that increases in value at increasing random times T_i, $i \in \mathbb{N}$ in the following form

$$N_t = \sum_{i=1}^{\infty} 1_{T_i \leq t}.$$

We denote the filtration generated by N as $(\mathcal{F}_t)_{t\geq 0}$.

We say that the random measure $\mu : \Omega \times [0,T] \to \mathbb{R}_+$ which is predictable with respect to $(\mathcal{F}_t)_{t\geq 0}$ is the compensator of the point process $N : \Omega \times [0,T] \to \mathbb{N}$ if

$$\mathbb{E}\left[N_{t+s} - N_t | \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^{t+s} \mu(du) | \mathcal{F}_t\right].$$

This generalizes the most often seen definition of simple Poisson processes as a process which satisfies for $\ell \geq 1$

$$\mathbb{P}(N(t+h) = k + \ell | N(t) = k) = 1_{\ell=1}\mu h + o(h).$$

The advantage in this case is that now the compensator μ can be random which is the case of the Hawkes model.

EXERCISE 8. *Prove that in the model (1.2), one has for $s, t > 0$ and the filtration generated by X*

$$\mathbb{E}\left[X_{t+s} - X_t | \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^{t+s} \alpha(X_s)ds | \mathcal{F}_t\right].$$

The model (1.2) can be somewhat extended in the following format:

$$X_t = X_0 + \int_0^t \beta(X_s)ds + N(\int_0^t \alpha(X_s)ds).$$

This example shows one of the simplest examples of the so-called PDMP (piecewise deterministic Markov process). Its theoretical understanding is simple, one has ordinary differential equations to solve in between jumps. The frequency of the jumps is governed through the solution itself. The simulation may seem to be simple, but note that one has to compute solutions of ordinary differential equations between jumps but also needs to keep track on when a jump happens as it affects the clock of the jump process.

EXERCISE 9. *Let $\phi_t(x)$ denote the flow solution of the ode*

$$\phi_t(x) = x + \int_0^t \beta(\phi_s(x))ds.$$

Suppose that $\phi_t(x)$ can be computed in close form. Apply a variation of the simulation method in Exercise 6 in order to simulate X, using ϕ.

Hint: The first step is to define the time T_1 as the one that satisfies $\exp\left(-\int_0^{T_1} \alpha(\phi_s(x))ds\right) = U_1$.

This method is clearly a theoretical one as there are very few cases where $\phi_t(x)$ can be explicitly computed. In fact, there are many models where α or β are not even Lipschitz (e.g. in most biological models coefficients are polynomials).

Most of the proposed methods discretize the ordinary differential system and then combine them with some kind of thining algorithm. Very few methods have been developed for models where there is infinite activity as in the case of Exercise 5.

DEFINITION 3. A linear Hawkes model is a random Poisson jump process X such that it satisfies

$$X_t = X_0 + N\left(\int_0^t \left(\mu_0 + \int_0^u \alpha(u-s)dX_s\right)du\right).$$

Here $\alpha : [0,T] \to \mathbb{R}_+$ is a decreasing function such that $\|\alpha\|_{L^1[0,T]} < 1$. In the literature the model is mostly proposed at the level of intensities as X is a point process with compensator given by

$$\mu_t = \int_0^t \left(\mu_0 + \int_0^u \alpha(u-s)dX_s\right)du.$$

Many variants of the Hawkes model exist such as the non-linear version

$$\dot{\mu}_t = \psi\left(\int_0^t \alpha(t-s)dX_s\right)$$

and the multi-dimensional version of it.

The most common simulation method uses the thinning algorithm:

ALGORITHM 4. *(1) Iteratively, simulate the times* $\{T_i < t, i \in \mathbb{N}\}$ *using the rate* $\dot{\mu}_{T_{i-1}}$.

(2) As the rate is time dependent, we use the thining algorithm in each time interval $[T_{i-1}, T_i]$.

EXERCISE 10. *Give a detailed description of the above algorithm and prove that it satisfies the required conditions for the linear Hawkes model.*

Many models in neurosciences follow some of the features we have explained in this chapter. There are also some other features that are in important in that field and for this reason require special treatment such as multi-scaling (meaning that some components of the equation have large values while others are small) and the dissapeareance of chaos. This latter issue may appear to be philosophical but most deterministic models in neurosciences are unstable and lead to chaos type structures. On the other hand, models with "appropriate" noise should provide well defined models which can then be used for other purposes such as parameter estimation.

2. Unbiased simulation of Stochastic differential equations

The basic continuous diffusion process that we will discuss in this section is given in two formats one analytic and another stochastic.

For the sake of simplicity, I will mostly discuss the one dimensional case. When pertinent, I will note when the discussion is dimension dependent..

Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which supports a Brownian motion $B = (B_t)_{t\geq 0}$ adapted to the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$, we consider the unique solution of the following stochastic differential equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s. \tag{2.1}$$

Here b and σ can be considered for the moment smooth with bounded derivatives.

In the one-dimensional case there are a number of exact simulation algorithms which are an accumulation of various tricks in order to deal with the most general situations. We start from the simplest situation.

The first trick consists on using the Lamperti transform. That is, define

$$L(x) = \int_c^x \sigma(y)^{-1} dy.$$

Then the law of the solution can be written using Itô's formula and Girsanov transformation as

$$\mathbb{E}\left[F(X)\right] = \mathbb{E}\left[F(L^{-1}(Y)) \exp\left(\int_0^T \psi(Y_s) dY_s - \frac{1}{2}\int_0^T \psi^2(Y_s) ds\right)\right].$$

Here, F is a functional of continuous paths in the interval $[0,T]$. Furthermore, Y is a Brownian motion and

$$\psi(x) = b\sigma^{-1}(x) - \frac{1}{2}\sigma'(x).$$

Further, using "inverse" Itô formula one has for $\Psi' = \psi$

$$\int_0^T \psi(Y_s) dY_s = \Psi(Y_t) - \Psi(Y_0) - \frac{1}{2}\int_0^t \psi'(Y_s) ds.$$

Therefore, we arrive at the formula:

$$\mathbb{E}\left[F(X)\right] = \mathbb{E}\left[F(L^{-1}(Y)) \exp\left(\Psi(Y_T) - \Psi(Y_0) - \frac{1}{2}\int_0^T (\psi^2 + \psi')(Y_s) ds\right)\right].$$

Finally suppose that $\psi^2 + \psi' \in [0, M]$ for some $M > 0$. Then we can interpret the above expression using a point process N with random compensator given by $\frac{1}{2}(\psi^2 + \psi')(Y_s)$. Therefore

$$\mathbb{E}\left[F(X)\right] = \mathbb{E}\left[F(L^{-1}(Y)) \exp\left(\Psi(Y_T) - \Psi(Y_0)\right) 1_{N_T=0}\right] \tag{2.2}$$

The simulation in the case that $F(X) = X_T$ is performed as follows:

ALGORITHM 5. *(1) Simulate the jump times of a Poisson process with parameter M.*

(2) For each jump time S, simulate the Brownian motion $Y(S)$ and a $[0,1]$ uniform random variable, U. The result of the simulation is zero unless $U \leq \frac{1}{2M}(\psi^2 + \psi')(Y_S)$.

(3) Otherwise record the result and proceed until one overpases the time T (always with the condition in (2)) and simulate the last part of the Brownian motion.

EXERCISE 11. *Prove that the above algorithm produces the law stated in (2.2).*

REMARK 1.

Many extensions of this method are possible and we refer to the references for more information. For example, one may consider various other functionals F for which the laws under the Brownian case are known. Intermediate points between the jump times of the point process can be simulated using brownian bridges. We will consider the jump case below as this is one of the main directions of these informal notes.
On the other hand, it is not difficult to think of situations where L and/or Ψ are not explicit and require approximation.
The most restrictive aspect is the fact that it is essentially a one-dimensional methodology. An extension using rough path methods is discussed in [17] which is another variation of this method although the error is ϵ a.s.

We will now give the description in the jump case. Suppose we want to simulate paths of the process:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s + \int_0^t \int_E g(s, X_{s-}, z)N(X_{s-}, ds, dz).$$

Here N is a point process with compensator[2] $\mu(x, ds, dz) = f(z)\mu(s, x)dsdz$ which may depend on the state of the process. We will suppose that g is Lipschitz and that μ is bounded by M. Then one can show that a modification of the above ideas lead to a exact simulation method.

EXERCISE 12. *Perform the following steps in order to prove the above statement:*

(1) Use Itô's formula in order to find the stochastic differential equation satisfied by $Y = L(X)$.

(2) Use Girsanov's theorem to express the functional $\mathbb{E}[F(X)]$ *in function of the solution of a stochastic equation withour drift.*

(3) Importantly, consider the solution of a stochastic differential equation, say Z*, which is the same as* Y *except that the driving point process is now given by*

$$1\left(u \leq \frac{\mu(s, L^{-1}(x))}{M}\right)\bar{N}(ds, du, dz).$$

Here the compensating measure for $\bar{N}$ *is given by* $1_{[0,1]}(u)Mf(z)dudsdz$*. Prove that the law of* Z *and* Y *are the same.*

(4) Propose a Monte Carlo simulation method for $\mathbb{E}[F(Z)]$ *using the previous ideas.*

[2] with some abuse of notation.

A variation that has been attracting some attention is the so-called ϵ-strong algorithms. In essence it means that one can use instead of Monte Carlo simulation of Brownian motion as the initial brick in the construction any method which we know it converges a.s. to some Brownian motion. This will provide the so-called ϵ-strong procedure once we know that the error rate is almost surely controlled by $M\epsilon$ with an explicit constant M. This can be achieved in various ways. Two of them are by discretization of the space so that one can control the error of the process and the other is through rough path analysis which provides almost sure error bounds.

We will come back to some of the ideas explained here but one also understands that the main idea of the method is to reduce a complex stochastic differential equation to the simplicity of Brownian and Poisson components. From the theoretical point of view this idea maybe interesting as many theoretical properties may be obtained using these unbiased simulation structures.

3. The multi-dimensional case

If we consider the multi-dimensional case

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma_{\cdot i}(X_s)dB_s^i. \tag{3.1}$$

Here $\sigma_i, b : \mathbb{R}^d \to \mathbb{R}^d$, $i = 1, ..., k$ and B is a p-dimensional Brownian motion. Performing a Girsanov transform in order to modify the drift b into a drift h requires the obtain a solution $\bar{b}$ to the equation

$$\sigma(x)\bar{b}(x) = (h - b)(x).$$

Supposing that the above equation can be solved one can determine that h is such that the equation is transformed into

$$X_t = X_0 + \int_0^t \sigma_{\cdot i}(X_s) \circ dB_s^i.$$

In this situation, one may find an explicit solution to the above sde through the solution of the following muti-dimensional ode:

$$D_x F(\alpha, x) = \sigma(F(\alpha, x)).$$
$$F(\alpha, 0) = \alpha$$

Exercise 13. *Find the process Y so that $F(B_t, Y_t) = X_t$.*

Although the above may look promising, one needs to have the following condition on σ (the so-called Frobenius condition) to assure existence and uniqueness of F, i.e., for all $m \in \{1, \dots, k\}$

$$\sum_{l=1}^{d} \sigma_{lm} \frac{\partial \sigma_{ij}}{\partial x^l} = \sum_{l=1}^{d} \sigma_{lj} \frac{\partial \sigma_{im}}{\partial x^l}.$$

This condition is highly restrictive in multi-dimensional cases and most example do not satisfy it. One may think of approximations but even then the jump case seems difficult to be considered in such a framework.

This is a problem that will be discussed in later lectures and can also be dealt with using the following method.

3.1. The splitting method. Splitting methods are well known methods for approximation of solutions of partial differential equations. Their objective is to separate complex calculations in steps. In particular the most common examples is to separate the calculation of drift effects and diffusive effects.

To give the mathematical heuristics let $L = L_1 + L_2$ be the generator of a stochastic process that we want to approximate. Suppose that we want to approximate the semigroup P in the following way

$$P_t = e^{t(L_1+L_2)} = I + t(L_1 + L_2) + \frac{t^2}{2}(L_1 + L_2)^2 + ...$$

We will not discuss the issue concerning the stability properties which are technical properties. We will rather concentrate on how to build approximations to P. For this one may use the semigroups associated to L_1 and L_2 as follows

$$e^{tL_1} e^{tL_2} = I + t(L_1 + L_2) + \frac{t^2}{2}(L_1^2 + L_2^2 + 2L_1 L_2) + ...$$

Therefore this approximation has an error of order t^2. That is,

$$P_t - e^{tL_1} e^{tL_2} = \frac{t^2}{2}(L_2 L_1 - L_1 L_2) + ...$$

One idea to improve this approximation is to define the following approximation

$$P_t - \frac{1}{2}(e^{tL_1} e^{tL_2} + e^{tL_2} e^{tL_1}) = \frac{t^3}{3!} H_6 + ...$$

In fact, we leave the following approximation result to be checked by the reader (in the setting of sde's, this is called Fujiwara approximation of order 5)

$$\begin{aligned} P_t - \frac{1}{3}Q_t^{[1]} + \frac{4}{3}Q_t^{[2]} &= t^5 H_{10}. \\ Q_t^{[1]} &= \frac{1}{2}(e^{tL_1}e^{tL_2} + e^{tL_2}e^{tL_1}) \\ Q_t^{[2]} &= \frac{1}{2}(e^{\frac{t}{2}L_1}e^{\frac{t}{2}L_2}e^{\frac{t}{2}L_1}e^{\frac{t}{2}L_2} + e^{\frac{t}{2}L_2}e^{\frac{t}{2}L_1}e^{\frac{t}{2}L_2}e^{\frac{t}{2}L_1}). \end{aligned}$$

We remark that the use of the above approximation will lead to the following approximation result under sufficient conditions (here $\xi_1 = \frac{1}{3}$ and $\xi_2 = -\frac{4}{3}$)

$$\left| P_T f(x) - \sum_{i=1}^{2} \xi_i (Q_{T/n}^{[i]})^n f(x) \right| \leq \frac{C(T, f, x)}{n^4}.$$

We have up to now just defined approximations of the semigroup P_t using composition and combinations of the so-called "coordinate" semigroups e^{atL_1} and e^{atL_2} for appropiate values of a.

The next step consists of finding probabilistic interpretations of these semigroups or if needed, meaningful approximations. That is, in general we say that the stochastic process $X^i(x)$, $i = 1, 2$ are the coordinate processes associated to L_i, $i = 1, 2$ if

$$e^{tL_i} f(x) = \mathbb{E}[f(X_t^i(x)], \ i = 1, 2.$$

If such processes exist then we have a clear stochastic representation of the approximations using compositions and randomization. For example,

$$\frac{1}{2}(e^{tL_1}e^{tL_2} + e^{tL_2}e^{tL_1})f(x) = \mathbb{E}[f(SX_t^1(X_t^2(x)) + (1 - S)X_t^2(X_t^1(x)))].$$

Here S is a $1/2$-Bernoulli random variable which is independent of everything else. Clearly one may have that X^i is itself not the approximation of interest and that we may have to further approximate it. This may happen in the case of Lévy driven SDE's. In the case that a further approximation is needed we will instead use a bar on top of the approximating semigroup, generating operator etc.

Clearly the above decomposition has been done with just 2 operators L_1 and L_2 for the sake of clarity but one can do the same for any number of them. In fact, in what follows we

consider a decomposition using three operators as follows:

$$\begin{aligned} L_1 f(x) &= \int_{|y|>\epsilon} f(x+h(x)y) - f(x)\nu(dy) \\ L_2 f(x) &= \int_{|y|\leq\epsilon} \left(f(x+h(x)y) - f(x) - f'(x)h(x)y1(|y|\leq 1)\right)\nu(dy) \\ L_3 f(x) &= -f'(x)h(x)\int_{1\geq|y|>\epsilon} y1(|y|\leq 1)\nu(dy). \end{aligned}$$

The coordinate process associated with L_3 is the solution of the ODE

$$X_3(t) = x + \int_0^t h(X_3(s))ds \int_{1\geq|y|>\epsilon} y1(|y|\leq 1)\nu(dy).$$

We suppose that this can be solved explicitly or that it can be approximated up to high order.

The Asmussen-Rosiński approach in this setting is an approximation for L_2. In fact if we let

$$\bar{X}_2(t) = x + \int_0^t h(\bar{X}_2(s))dW_s\sigma_\epsilon,$$

where $\sigma_\epsilon^2 = \int_{|y|\leq\epsilon}|y|^2\nu(dy)$. In fact, the generator of this process is known and we have using Taylor's expansion that

$$\begin{aligned} & (L_2 - \bar{L}_2)f(x) \\ = & \int_{|y|\leq\epsilon} \left(f(x+h(x)y) - f(x) - f'(x)h(x)y1(|y|\leq 1)\right)\nu(dy) - \int_{|y|\leq\epsilon}\frac{1}{2}f''(x)h(x)^2y^2\nu(dy) \\ = & \int_{|y|\leq\epsilon}\int_0^1 \frac{\alpha^2}{2}f'''(x+\alpha h(x)y)d\alpha h(x)^3y^3\nu(dy). \end{aligned}$$

Therefore choosing ϵ small enough so that $\int_{|y|\leq\epsilon} y^3\nu(dy) \leq Ct$ will lead to a weak error of order 1 for this coordinate semigroup. For L_1 we consider the following approximation $\bar{X}_1 = x + h(x)Y$ where $Y = Z1(S=1)$ where S is a $p\equiv p(t)$-Bernoulli random variable and Z is a random variable with density function $g_\epsilon(y) = \lambda_\epsilon^{-1}g(y)1(|y|>\epsilon)$. Let us compute the semigroup expansion

$$\begin{aligned} \mathbb{E}[f(\bar{X}_1)] &= f(x)P(S=0) + \int f(x+h(x)y)g_\epsilon(y)dyP(S=1) \\ &= f(x) + \int_{|y|>\epsilon}(f(x+h(x)y) - f(x))g(y)dy\lambda_\epsilon^{-1}P(S=1). \end{aligned}$$

Therefore if $\lambda_\epsilon^{-1}P(S=1) = t$ we will have that

$$\mathbb{E}[f(\bar{X}_1)] - P_t^1 f(x) = t^2 H_4 + ...$$

Therefore putting the above three schemes together one may build a scheme of weak order one (assuming some additional regularity conditions) if the following conditions are met

$$\lambda_\epsilon^{-1} P(S=1) = t \text{ and } \int_{|y|\leq\epsilon} y^3 \nu(dy) \leq Ct^2.$$

The second condition fixes ϵ and the first will fix the probability of jump in the interval. Finally appropiate composition methods using the splitting method will finally give the approximation method.

REMARK 2. 1. Note that the proposed method gives a method with limited number of jumps (at most the number of interval partitions). This is different from other proposals when one simulates all jumps bigger than ϵ which on the average may lead to the same number of simulations but the level of variance is so big that may lead to long computation times.

2. The reasoning as to why the Asmussen-Rosiński approach works can be well understood from the above calculations. But it is also clear that this is not the only way of achieving this goal. In fact, using a moment matching method one may obtain simpler approaches with higher orders of convergence.

3. A further detailed analysis of the methodology makes you realize that in fact, increments for the approximating random variables do not need to have any particular laws but they have to respect some moment properties. This is called a moment method. This comment applies to many simulation methods.

4. Although we have analyzed only the weak convergence, strong convergence results can be achieved using embedding (or coupling) techniques. See [10] or [4].

5. Because this argument leaves a lot of error terms without any algebraic structure it is difficult to device a general methodology to increase the rate of convergence. This is done using the parametrix method that we will discuss later.

CHAPTER 2

Lecture 2: The Multi-level method

The first idea of this method is fairly simple and it is the use of the so-called telescopic formula. Suppose that $a = \lim_{n\to\infty} a_n$. Then one clearly has that

$$a = a_0 + \sum_{n=1}^{\infty}(a_n - a_{n-1}).$$

Now, one may assume that the calculation of a_n becomes increasingly complex with n as we are trying to estimate the quantity a. Therefore one expects that the approximation: $\sum_{n=1}^{N}(a_n - a_{n-1})$ has increasing complexity in each summand.

Therefore instead of putting all our efforts to estimate a_N with a high degree of accuracy, we can do it in reverse. That is, spend efforts in good estimation of a_n for low values of n and less effort in large values of n.

This can be understood with the notion of error and variance in the Monte Carlo setting. For this suppose that $a_n := \mathbb{E}[X_n]$ where X_n involves the amount of C_n calculations and that $|a_n - a_{n-1}| = |\mathbb{E}[X_n - X_{n-1}]| \leq s_n$.

Then we have to decide how many simulations to perform in each level. That is, we will simulate M_n times the difference $X_n - X_{n-1}$ so that our final estimate is

$$\bar{a}_N := \bar{X}_0(M_0) + \sum_{n=1}^{N} \overline{X_n - X_{n-1}}(M_n).$$

$$\bar{Y}(M) := \frac{1}{M}\sum_{i=1}^{M} Y^i.$$

Here Y^i indicate independent copies of the r.v. Y.

Therefore the total computational cost is $\sum_{n=0}^{N} M_n C_n$ and the variance of the estimate is $\sum_{n=0}^{N} M_n^{-1} s_n^2$. Various types of minimization procedure can be performed. For example, fixed the total computational cost, minimize the variance. Define a total performance indicator and minimize it.

The important point in the procedure is to achieve s_n to be small which implies that the construction of $X_n - X_{n-1}$ has to be done in such a way so that the variance of this difference

is small. his implies some coupling property in the construction or the study of the inner structure of this error process.

The simplest form of approximation for the solution of the continuous stochastic differential equation (2.1) is given by the Euler scheme defined on a closed interval $[0,T]$, using a time partition $t_k = \frac{kT}{n}$, $k = 1, ..., n$ and given for $t \in [t_k, t_{k+1}]$ as

$$\bar{X}_t = \bar{X}_{t_k} + b(\bar{X}_{t_k})(t - t_k) + \sigma(\bar{X}_{t_k})(B_t - B_{t_k}), \quad X_0 = X_0. \tag{0.1}$$

In the case that the equation has jumps then one has to also perform simulation of the increments of the Lévy process.

In contrast with the Brownian case, there are simulation methods for very explicit cases and in the general case one has to further use approximation procedures which will briefly explain here.

For example, in the (positive valued) stable case, one has that the characteristic function is given by

$$-\log\left(\mathbb{E}\left[e^{i\theta S}\right]\right) = c|\theta|^{\alpha}$$

An analytical decomposition the characteristic function gives the Chambers-Mallows-Stuck/Kanter probabilistic representation: $\alpha \in (0,1) \cup \{1\} \cup (1,2)$).

In fact, a "simple" random variate generator for S has been suggested by Kanter (1975), in the case $\alpha \in (0,1)$ who used an integral representation of Zolotarev (1966) (see Zolotarev (1986, p.74) for the argument which uses complex line integrals) which states that the distribution function of $S^{\alpha/(1-\alpha)}$ is given by

$$\mathbb{P}\left(S^{\alpha/(1-\alpha)} \leq x\right) = \frac{1}{\pi}\int_0^{\pi} e^{-\frac{A(u)}{x}} du$$

where A is Zolotarev's function:

$$A(u) := \left\{ \frac{(\sin(\alpha u))^{\alpha} \left(\sin((1-\alpha)u\right)^{1-\alpha}}{\sin u} \right\}^{\frac{1}{1-\alpha}}$$

Zolotarev's integral representation implies the following probabilistic representation

$$S := \left(\frac{A(U)}{E}\right)^{\frac{1-\alpha}{\alpha}}$$

where U is uniform on $[0,\pi]$ and E is exponential with mean one. This is Kanter's method. One can interpret $E^{\frac{1-\alpha}{\alpha}}$ as the "length" of the stable r.v. and $A(U)^{\frac{1-\alpha}{\alpha}}$ as its oscillation. Another variation of these formulas were provided and generalized by Chambers-Mallows-Stuck.

Note that the values of $A(U)$ near $U = \pi$ are explosive and a closer analysis indicates that $\mathbb{E}[S^p] < \infty$ if and only if $p \in (0, \alpha)$. This property is still valid for many other general stable r.v.'s.

Notably, I remark that there are a number of acceptance-rejection methods that have been proposed for tempered stable distributions. There are many other notable cases which have much simpler simulation methods such as Gamma, Cauchy, Inverse Gaussian etc.

Now, we return to the coupling issue.

For example, a simple result of this type is as follows (see [11]): Suppose that we propose approximations of the type

$$X^f_{j\Delta} = X^f_{(j-1)\Delta} + a(X^f_{(j-1)\Delta})\zeta^f_j, j = 0, ..., n_f = T/\Delta$$
$$X^c_{2j\Delta} = X^c_{(j-1)2\Delta} + a(X^f_{2(j-1)\Delta})\zeta^c_j,\ j = 0, ..., n^c = T/(2\Delta).$$

PROPOSITION 1. *Suppose that the coefficient function a is uniformly Lipschitz and has at most linear growth, that is,*

$$\left\|a(x) - a\left(x'\right)\right\| \leq L_a \left|x - x'\right|, \quad \|a(x)\|^2 \leq B_a^2 \left(1 + |x|^2\right)$$

for any $x, x' \in \mathbb{R}^d$ and some positive constants L_a and B_a, where for any matrix A, $\|A\|$ stands for the operator norm of the matrix A. Denote $\mathcal{R}_j := \zeta^c_j - \zeta^f_{2j-1} - \zeta^f_{2j}$ and suppose that $\mathcal{R}_j$ $j = 1, \ldots, n_c$, are zero mean i.i.d. random vectors. Moreover, assume that $\mathrm{E}\left[|X_0|^2\right] < \infty$, then the following estimate holds

$$\begin{aligned}\mathrm{E}\left[\max_{j=0,\ldots,n_c}\left|X^{\Delta_f}_{j\Delta} - X^{\Delta_c}_{j\Delta}\right|^2\right] \leq & c_1\left(n_f m^2_{f,2} + n^2_f m^2_{f,1} m_{f,2} + n_f \mathrm{E}\left[|\mathcal{R}_1|^2\right]\right) \\ & \times \exp\left[c_2\left(n_f m_{f,2} + n^2_f m^2_{f,1}\right)\right]\end{aligned}$$

for some constants $c_1 > 0, c_2 > 0$ depending on L_a and B_a. Here, $m_{f,2} = \mathbb{E}[|\zeta^f_j|^2]$ and $m_{f,1} := |\mathbb{E}[\zeta^f_j]|$.

In most continuous type cases, $\mathcal{R}_1 = 0$, $m_{f,2} = O(\Delta)$ and $m_{f,1} = O(\Delta)$. The proof of this proposition essentially follows carefully the proof of existence and uniqueness of strong solutions using a discrete time version of Gronwall's inequality.

A different problem appears when considering stochastic equations which are driven by Lévy processes with no moments like the stable case.

EXERCISE 14. *Try to propose a way to apply the above estimates in the stable case*

Before this, let us consider the case when the Lévy process can not be explicitly simulated. In this generality one has that the characteristic function of a multivariate Lévy process (without Brownian and drift component) with Lévy measure ν is given by

$$\mathbb{E}[e^{i\theta\cdot Z_t}] = \exp\left(t\int\left(e^{i\theta\cdot x} - 1 - 1_{|x|\leq 1} i\theta\cdot x\right)\nu(dx)\right).$$

Here ν satisfies that

$$\int (|x|^2 \wedge 1)\nu(dx) < \infty.$$

In many results we have given so far the condition that the compensating measure of a point process is finite is loosely equivalent to the fact that the Lévy process can be characterized as some variation of a compound Poisson process.

In the case that $\nu(\{x; |x| \leq 1\}) = \infty$, usually called the infinite activity case, the study can be characterized using the so-called Blumenthal-Getoor index which measures how many moments of Z are finite. Without going into much detail, I will just say that in the case $\nu(dx) = |x|^{1-\alpha}dx$ which corresponds to the α-stable case in one dimension explained before moments of order $p \in (0, \alpha)$ are finite while higher moments are infinite. This creates a problem in the application of Proposition 1.

A general approximation procedure was proposed previously as the Asmussen-Rosinński method [3]. The approximation consists in dividing the Lévy process in two parts. One contains the large jumps and the small jumps are approximated by a Gaussian random variable.

One way to explain why this approach works is through Itô's formula. Suppose that the approximation in one dimension is defined as $\bar{Z}_t = Z^\epsilon_t + a_\epsilon B_t$. Here $a_\epsilon > 0$ is a constant to be determined through the error analysis argument. Also Z^ϵ is the Lévy process defined through the Lévy measure $\nu(\cdot \cap \{x; |x| > \epsilon\})$. Consider the intermediate process $Y_t = x + Z_t + \bar{Z}_T - \bar{Z}_t$

and the smooth semigroup $\bar{P}_t f(x) = \mathbb{E}[f(x + \bar{X}_t)]$. Then

$$\begin{aligned}
&\mathbb{E}\left[f(Y_t) - f(Y_0)\right] = \mathbb{E}[\bar{P}_{T-t} f(x + Z_t)] - \bar{P}_T f(x) \\
=&\mathbb{E}\left[\int_0^t \int \bar{P}_{T-s} f(x + Z_s + y) - \bar{P}_{T-s} f(x + Z_s) - (\bar{P}_{T-s} f)'(x + Z_s) y 1_{(-1,1)}(y) \nu(dy)\right] \\
&- \mathbb{E}\left[\int_0^t \int_{[-\epsilon,\epsilon]^c} \bar{P}_{T-s} f(x + Z_s + y) - \bar{P}_{T-s} f(x + Z_s) + (\bar{P}_{T-s} f)'(x + Z_s) y 1_{(-1,1)}(y) \nu(dy)\right] \\
&- \mathbb{E}\left[\frac{a_\epsilon^2}{2} \int_0^t (\bar{P}_{T-s} f)''(x + Z_s) ds\right] \\
=&\mathbb{E}\left[\int_0^t \int_{[-\epsilon,\epsilon]} \bar{P}_{T-s} f(x + Z_s + y) - \bar{P}_{T-s} f(x + Z_s) + (\bar{P}_{T-s} f)'(x + Z_s) y 1_{(-1,1)}(y) \nu(dy)\right] \\
&- \mathbb{E}\left[\frac{a_\epsilon^2}{2} \int_0^t (\bar{P}_{T-s} f)''(x + Z_s) ds\right].
\end{aligned}$$

Using Taylor's formula one understands that to make the above difference small one needs to take

$$a_\epsilon^2 = \int_{[-\epsilon,\epsilon]} y^2 \nu(dy).$$

The the residue will be of order $\int_{[-\epsilon,\epsilon]} |y|^3 \nu(dy) \leq \epsilon a_\epsilon^2$.

Therefore if we consider $t = T$ in the above expression, we obtain for $f \in C_b^3$

$$|\mathbb{E}\left[f(x + Z_T) - f(x + \bar{Z}_T)\right]| \leq C_f \int_{[-\epsilon,\epsilon]} |y|^3 \nu(dy). \tag{0.2}$$

In order to consider the multilevel method one needs stronger results using coupling arguments. The next portion is taken from [30]. For the general case, see [64] and [21].

Consider two laws P, Q on $\mathbb{R}$ with finite variance. The Wasserstein distance $\mathcal{W}_2$ is defined by

$$\mathcal{W}_2^2(P, Q) = \inf\left\{\mathbb{E}\left[|X - Y|^2\right], \mathcal{L}(X) = P, \mathcal{L}(Y) = Q\right\}$$

With an abuse of notation, we also write $\mathcal{W}_2(X, Y) = \mathcal{W}_2(X, Q) = \mathcal{W}_2(P, Q)$ if $\mathcal{L}(X) = P$ and $\mathcal{L}(Y) = Q$. We recall the following result of Rio [Theorem 4.1]. Theorem 5. There is an universal constant C such that for any sequence of i.i.d. random variables $(Y_i)_{i \geq 1}$ with mean 0 and variance θ^2, for any $n \geq 1$,

$$\mathcal{W}_2^2\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i, \mathcal{N}\left(0, \theta^2\right)\right) \leq C \frac{\mathbb{E}\left[Y_1^4\right]}{n\theta^2}$$

Here $\mathcal{N}\left(0, \theta^2\right)$ is the Gaussian distribution with mean 0 and variance θ^2.

Corollary 6. Consider a pure jump centered Lévy process $(Z_t)_{t\geq 0}$ with Lévy measure ν. In other words $Z_t = \int_0^t \int_{\mathbb{R}_*} z\tilde{M}(ds,dz)$, where $\tilde{M}$ is a compensated Poisson measure with intensity $ds\nu(dz)$. There is an universal constant C such that

$$\forall t \geq 0, \quad \mathcal{W}_2^2\left(Z_t, \mathcal{N}\left(0, tm_2(\nu)\right)\right) \leq C\frac{m_4(\nu)}{m_2(\nu)}$$

In heuristic terms, the spirit of the above result resembles the basic argument to obtain (0.2) with more technicalities[1]. In fact, the fourth moment appears instead of the moment of order three and the variance in the denominator appears because one needs to use non-regular functions f and therefore the need for integration by parts formulas.

If this result is applied to the distance between Z and $\bar{Z}$ one obtains instead that the bound of the error is

$$\frac{\int_{[-\epsilon,\epsilon]} |y|^4 \nu(dy)}{\int_{[-\epsilon,\epsilon]} |y|^2 \nu(dy)} \leq \epsilon^2.$$

Still, how to apply the multi-level method for stable driven type sde's does not seem completely treated in the literature due to moment problems as explained before (for some results in some cases see [11], [24] and [23]). From now on, I start to describe results in cases where there are still many things to be developed.

Other models in this family that I would like to discuss later are the α-CIR and Boltzmann's models.

1. The multilevel method as a unbiased simulation method: The variance blow up problem

Because of the difficulty to quantify the discretization error, there is an interest in so-called unbiased simulation methods. These methods allow for automatization and parallelization which can increase the efficiency of the simulation.

We differentiate here between exact methods and unbiased methods. Exact meaning to sample a path, at a finite set of points, with the distribution of the sde, while unbiased means that we can, with out bias, estimate $\mathbb{E}\left[f(X_T)\right]$. As we have seen in Chapter 1 this is possible in one dimension using Girsanov's theorem. We will now discuss the situation in higher dimensions.

[1] One needs to apply some inequality related with Zolotarev distances which gives some weak duality form for $\mathcal{W}_2$

For the sake of the argument, let X_T^n denote an approximation of X_T using an Euler method with uniform time steps of length $2^{-n}T$, $n \geq 0$. The MLMC method, introduced in [33], is then to approximate $E[f(X_T)]$,

$$\mathbb{E}[f(X_T)] \approx \mathbb{E}\left[f(X_T^0)\right] + \sum_{n=1}^{\bar{n}} \mathbb{E}\left[f(X_T^n) - f(X_T^{n-1})\right],$$

up to some final level $\bar{n}$. At each level a certain number of Monte Carlo samples are generated and the sample average then approximates the expectation. Choosing the number of samples at each level in a good way will improve the convergence rate compared to the Euler method.

Now, letting $\bar{n} \to \infty$ we can write, assuming convergence,

$$\begin{aligned}
\mathbb{E}[f(X_T)] &= \mathbb{E}\left[f(X_T^0)\right] + \sum_{n=1}^{\infty} \mathbb{E}\left[f(X_T^n) - f(X_T^{n-1})\right] \\
&= \mathbb{E}\left[f(X_T^0)\right] + \sum_{n=1}^{\infty} p_n \frac{\mathbb{E}\left[f(X_T^n) - f(X_T^{n-1})\right]}{p_n} \\
&= \frac{1}{p_0}\mathbb{E}\left[1(N=0) f(X_T^0)\right] + \mathbb{E}\left[1(N \geq 1)\frac{f(X_T^N) - f(X_T^{N-1})}{p_N}\right], \qquad (1.1)
\end{aligned}$$

where N is a random variable with distribution $p_n > 0$, $n \geq 0$. The last expression above can be simulated unbiasedly using Monte Carlo methods. Now, we would like to discuss the second moment of the above proposed estimator. A straightforward calculation gives

$$\mathbb{E}\left[\left(\frac{f(X_T^N) - f(X_T^{N-1})}{p_N}\right)^2\right] \leq C \sum_{n=1}^{\infty} \frac{r_n}{p_n},$$

where $r_n := \mathbb{E}\left[\left(f(X_T^n) - f(X_T^{n-1})\right)^2\right]$. As discussed previously, the rate at which this error goes to to zero is well understood. In fact, under sufficient regularity hypotheses on b, σ and f one knows that $r_n = O(2^{-n})$. Then the variance of the method will be finite if we choose $p_n \sim 2^{-n}n^2$. On the other hand, note that the average number of random number generators used can be considered to be $\sum_{n=1}^{\infty} 2^n p_n = \infty$. Therefore the procedure is doomed to have infinite variance as long as ∞-level Monte Carlo method goes. The choice of p_n may be changed in order to make the average complexity finite but then the variance of the method will be infinite.

The interpretation of the method is clear. The method has no bias because it relies on an infinite order expansion. The level used in each simulation is determined by the value of the random variable N and the amount of simulation in each level is determined by the choice of

the probability distribution $p_n, n \geq 0$. Still, as it can not be applied in the ∞-dimensional case, one tries to approximate it by taking a certain number of levels. In fact, the previous calculation shows that taking too many levels in the MLMC may lead to the wrong result. One remedy is to use the Milstein scheme which improves on the order of strong convergence. However this assumes more regularity and is also difficult to use in multidimensional problems (for more on this, see [34]).

Furthermore, in the case that the coefficients σ, b are Hölder functions and f is not a regular function it is well known that the rate r_n degenerates quickly (see e.g. [41] and [5]). Therefore the applicability of the MLMC method as understood above is limited. For this reason, we will propose in the next chapter to use the parametrix method of approximation as one possible extension of the MLMC method in what follows. This method will allow a more profound analysis of the differences between approximations showing exactly where the variance explosion appears. This variance problem will then be solved by using a more accurate expansion.

The principle of simulation without bias just described can also be found in [59] and [60] which have already appeared in [51] which cites [36] as a source of this idea. As explained in Section 3 of [60] one can not apply $L^2(\Omega)$ criteria to this problem and even if a criteria in probability is applied as in Section 4 in that same paper then the computational complexity increases as the strong rate of convergence slows down.

Methods have also been derived for some special cases, e.g. [26] for the SABR model and [19] for the Heston model. These methodologies may be of interest in later chapters.

CHAPTER 3

Lecture 2A: An infinite level Monte Carlo method: The parametrix

We have seen in previous chapters that the main goal is to approximate $\mathbb{E}[f(X)]$ for various functionals f and for various stochastic processes X.

We have also observed that it is difficult to obtain an unbiased method for multidimensional processes X. In the methodology using Multi level Monte Carlo, we have seen a way of carrying out a randomization which allows the possibility of unbiased simulations.

In this chapter, I switch to one dimensional continuous diffusion processes with no drift for ease of explanation and more importantly because the problems with jumps in the case of finite compensator dealt as in chapter 1 have not been well developed as we will point out at the end of this lecture. We assume regularity of coefficients and uniform ellipticity.

Therefore, the goal in this chapter is to obtain formulas of the type:

$$\begin{aligned}\mathbb{E}[f(X_T)] =& \mathbb{E}\left[f(\bar{X}_T^{\pi}) Z_T\right] \\ Z_T =& \mathbf{1}_{N_T=0} + \mathbf{1}_{N_T>0} \prod_{i=1}^{N_T} \theta_{\zeta_{i+1}-\zeta_i}(\bar{X}_{\zeta_i}^{\pi}, \bar{X}_{\zeta_{i+1}}^{\pi}).\end{aligned}$$

In the above, we have used the following notation:

(1) $f \in \mathcal{B}_b(\mathbb{R};\mathbb{R})$

(2) $\bar{X}^{\pi_0}$: approximation process for X defined for any partition π_0 of $[0,T]$.

(3) N is a Poisson process with jump times $\{\tau_i\}_i$ independent of $\{\bar{X}^{\pi_0};\pi_0\}$

(4) In the above formula, we abuse the notation letting $\bar{X}_T^{\pi} := \bar{X}_T^{\pi_0}\big|_{\pi_0=\pi}$ where $\pi := \{\zeta_i := \tau_i \wedge T\}$.

The general parallel with the unbiased formulas in 1 is that Z_T replaces the change of measure due to Girsanov's theorem. As can be expected Z_T can not define a change of measure.

EXERCISE 15. *We may discuss in this exercise the change of measure in the case that there is no drift. Prove that the change of measure that changes the process $\bar{X}^\pi_T$ into X_T can not converge as $\|\pi\| \to 0$.*

In fact, consider π_n to be the uniform partition of size T/n and denote by $\bar{X}^n$ the respective Euler scheme. Prove that the Radon-Nykodim derivative associated to the change of measure $\mathbb{P}^n := \mathbb{P} \circ (\bar{X}^n_{t^m_1}, ..., \bar{X}^n_{t^m_m})\ -1$ to $\mathbb{P}^m := \mathbb{P} \circ (\bar{X}^m_{t^m_1}, ..., \bar{X}^m_{t^m_m})^{-1}$ for $m \geq n$ and $\pi^m = \{t^m_i = iT/m; i = 1, ..., m\}$ is given by

$$\frac{d\mathbb{P}^m}{d\mathbb{P}^n}(x_1, ..., x_m) := \prod_{j=1}^m \sqrt{\frac{\sigma^n_j}{\sigma^m_j}} \exp\left(-(a^n_j - a^m_j)\frac{(x_j - x_{j-1})^2}{2a^m_j a^n_j \Delta_m}\right).$$

Here we are using a simplified notation $f^n_j := f(\bar{X}^n_{t^m_j})$, $\Delta_m := T/m$ and $a = \sigma^2$.

One expects that

$$\sum_{j=1}^m (a^n_j - a^m_j)\frac{(\bar{X}^n_j - \bar{X}^n_{j-1})^2}{2a^m_j a^n_j} \to \int_0^T \frac{a(\bar{X}^n_{\eta(s)}) - a(\bar{X}^n_s)}{2a(\bar{X}^n_s)} ds.$$

Analyze the remaining terms and convince yourself that the above can not be expected to converge. Other indications can be obtained reading about Kakutani's theorem or Cameron-Martin theorem.

1. The Duhamel principle

Let L be a differential operator let $P_t = e^{-tL}$ its generating semigroup. Let $\bar{L}$ be an approximating operator and $\bar{P}$ its approximating semigroup. Then the following Taylor-like expansion can be obtained if $\bar{L}$ is "good enough"

$$\begin{aligned} P_t f - \bar{P}_t f &= \int_0^t \partial_s \left(P_s \bar{P}_{t-s} f\right) ds \\ &= \int_0^t \left(P_s (L - \bar{L}) \bar{P}_{t-s} f\right) ds \end{aligned}$$

This is a linear equation in P which can be solved by iteration. The first step is as follows:

(1.1)

$$P_t f - \bar{P}_t f = \int_0^t \left(\bar{P}_{s_1}(L - \bar{L})\bar{P}_{t-s_1} f\right) ds_1 + \int_0^t \int_0^{s_2} \left(P_{s_1}(L - \bar{L})\bar{P}_{s_2 - s_1}(L - \bar{L})\bar{P}_{t-s_2} f\right) ds_1 ds_2.$$

Denote by $S_s(x,y) := (L-\bar{L})\bar{P}_s\delta_y(x)$. Then we have

$$P_tf(x) = \sum_{m=0}^{\infty} I_m(t,x)(f) \tag{1.2}$$

$$I_0(t,x)(f) := \int f(y)\bar{p}_t(x,y)dy$$

$$I_1(t,x)(f) := \int\int_0^t \int \bar{p}_{s_1}(x,y_1)S_{t-s_1}(y_1,y_2)f(y_2)dy_1ds_1dy_2$$

$$I_{m+1}(t,x)(f) := \int_0^t I_m(s_{m+1},x)(F_{s_{m+1}})ds_{m+1}$$

$$F_s(y) := \int S_{t-s}(y,z)f(z)dz. \tag{1.3}$$

An explicit example:

If $\bar{P}$ denotes the semigroup generated by the Euler scheme, we have

$$\bar{P}_t^z f(x) = \int f(y)\bar{p}_t^z(x,y)dy$$

$$\bar{p}_t^z(x,y) := \frac{1}{\sqrt{2\pi\sigma(z)}}\exp\left(-\frac{(y-x)^2}{2a(z)t}\right)dy$$

$$\bar{L}^z f(y) = \frac{a(z)}{2}f''(y).$$

Note that one usually sets $z=x$ as the Euler scheme is described probabilistically as $\bar{X}_t = x+\sigma(x)W_t$ which has the density $\bar{p}_t^x(x,y)$. Note that in the above argument the expression

$$\begin{aligned}(L-\bar{L}^z)\bar{P}_{t-s_1}^z f(y_1) =& \frac{1}{2}(a(y_1)-a(z))\partial_x^2\int \bar{p}_{t-s_1}^z(x,y_2)f(y_2)dy_2\Big|_{x=y_1}\\ =& \frac{a(y_1)-a(z)}{2a(z)(t-s_1)}\int\left(\frac{(y_2-y_1)^2}{a(z)(t-s_1)}-1\right)\bar{p}_{t-s_1}^z(y_1,y_2)f(y_2)dy_2.\end{aligned} \tag{1.4}$$

The fact that the term $t-s_1$ appears in the denominator is problematic as we need to make sense of the first term in (1.1):

$$\int_0^t \left(\bar{P}_{s_1}(L-\bar{L})\bar{P}_{t-s_1}f\right)ds_1 \tag{1.5}$$

In fact, using the uniform ellipticity and the Liptschitz property of a, we obtain as a bound

$$\left|\left((L-\bar{L})\bar{P}_{t-s_1}^z f\right)(y_1)\right| \le C\|f\|_\infty \frac{|y_1-z|}{t-s_1}.$$

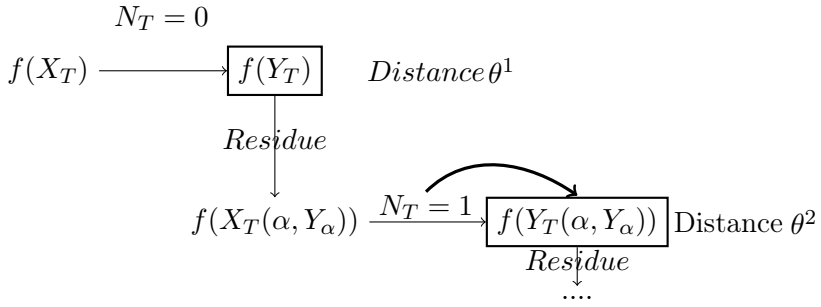
From here one obtains that the time integral in (1.5) is well defined if .

2. Probabilistic representation through Taylor

We can re-interpret the above developments as a random Taylor expansion

$$\mathbb{E}\left[f(X_T)\right] = \mathbb{E}\left[f(\bar{X}_{N_T+1}) \prod_{i=1}^{N_T+1} \theta^i\right].$$

Here Y is an approximation. In the which for some reason(s) maybe changed to $\bar{X}$.



Recall that the jump times of a Poisson process given the number of jumps follow the distribution of the order statistics. That is,

(1) The sum over the variable m in (1.2) indicates the number of jumps of the Poisson process N_t.

(2) The iterated time integrals in the simplex $0 \leq s_1 \leq ... \leq s_m \leq t$ can be interpreted probabilistically as the distribution of the m-jump times of the Poisson process conditioned on the fact that the Poisson process jumped m-times in the interval $[0, t]$.

(3) Each occurence of $\bar{p}$ in the formula for I_m corresponds to one iteration of the Euler scheme.

(4) The formula for θ corresponds to the evaluation in (1.4) for $z = x$ which equals to a multiple of the second order Hermite polynomial associated to $\bar{p}$.

From the above observations the probabilistic representation arises. In fact, one obtains that

$$\theta_s(y_1, y_2) = \frac{a(y_2) - a(y_1)}{2a(y_1)s}\left(\frac{(y_2 - y_1)^2}{a(y_1)s} - 1\right).$$

The problem with this representation is that its variance will be infinite (exercise).

One solution is to perform importance sampling with respect to the jump times. The most recent solution to this problem is to obtained a refined approximation using I_0 and I_1 and redo the iteration. This is done in two different ways in [2] and [20].

CHAPTER 4

Lecture 3A: Two problems

In this final chapter, I would like to describe two problems that I consider to be related.

1. The α-stable CIR model

The first problem is to consider the approximation to the following model called the α-CIR model:

$$X_t = x + \int_0^t a(b - X_s)ds + \sigma \int_0^t \sqrt{X_s} dB_s + \sigma_Z \int_0^t X_{s-}^{1/\alpha} dZ_s.$$

Here Z is an α-stable process with $\alpha \in (1, 2]$ and positive jumps. General considerations of existence and uniqueness can be found in [32] or [50]. This is a model that maybe used in problems related to interest rate models in mathematical finance (see [44]) but can also be considered as a limit for certain type of branching models (see [45]).

Building approximations to this problem can be obtained from the generator of X:

$$\begin{aligned} Lf(x) =& a(b-x)f'(x) + \frac{a}{2}xf''(x) + a_Z x \int (f(x+z) - f(x) - f'(x)z)\nu(dz) \\ \nu(dz) :=& C1(z > 0)z^{-(1+\alpha)}dz \end{aligned}$$

In fact the case $a_Z = 0$ corresponds to the so-called Cox-Ingersoll-Ross model for which there is an explicit density. One way to obtain this density is by starting with an ansatz $A(t, x, y)$ and then iterate the errors which can be discerned by computing $(\partial_t - L_x)A$. In a similar way, one can "probably" obtain an expansion of the density in the case that $a = 0$. These are principles that are related to the parametrix method.

In this case, one has to be careful with the time-space re-scaling properties (probably discussed in A. Kulik lectures).

From these considerations, one may expect to obtain theoretical properties related to the density of the solution process. Also, one should obtain a simulation method as proposed in the previous chapter may be obtained. There will still issues about infinite moments which will need to be treated. Some results on the strong approximation are available (see [48] and

[49]). Still, many equations of these type which explore the limits of coefficient irregularity in equations with jumps can be studied. Within a large literature we mention [43], [57] etc.

The Boltzmann equation is a non-linear equation related to the density of particles within a gas. We consider an imaginary situation where particles collide in a medium and we observe a section of it, say $\mathbb{R}^2$. These particles collide at a certain angle θ and velocity v which generates a certain force within a gas.

The Boltzmann equation main quantity of interest $f_t(v)$ describes the density of particles traveling at speed v at time $t > 0$ supposing an initial distribution f_0. We assume that these densities are equal all over the section and therefore independent of the position within the section.

The feature of interest to be proven here is that even if f_0 is a degenerate law in the sense that it is concentrated at some points, the noise in the corresponding equation will imply that for any $t > 0$ f_t will be a well defined function with some regularity.

The presentation format in this section follows closely the one presented in [6] with some simplifications. This field of research is growing very quickly and therefore even at the present time the results presented here may be outdated. Still, our intention is to provide an explicit example of application of the method presented in the previous chapter.

2. 2D homogeneous Boltzmann equations

In this section, we provide the basics of the Boltzmann equation needed in order to proceed without going into too many (interesting) physical and mathematical details. In particular, our physical explanations below are just in order to give a very loose idea of the physical set-up without claiming exactness.

Fix $T > 0$, $0 < \gamma \leq 1$ and $0 < \alpha < 1$ throughout the chapter. Define $b(\theta) = |\theta|^{-1-\alpha}$. In what follows, $f_t(v)$ denotes the density of particles with velocity v at time $t \geq 0$. Also let $\Pi := [-\pi/2, \pi/2]$ and $A(\theta) := (R_\theta - I)/2$, where I is the identity matrix and R_θ is the $2D$ rotation (particle collision) matrix of angle θ:

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Consider the 2-dimensional spatially homogeneous gas modeled by the Boltzmann equation:

$$\frac{\partial f_t(v)}{\partial t} = \int_\Pi \int_{\mathbb{R}^2} \left\{ f_t(v')\, f_t(v'_*) - f_t(v)\, f_t(v_*) \right\} B(|v - v_*|, \theta)\, dv_*\, d\theta, \tag{2.1}$$

where $B(|v-v_*|,\theta) = |v-v_*|^\gamma\, b(\theta)$, $b(\theta) = |\theta|^{-1-\alpha}$ for $\alpha \in (0,1)$ describes the cross section physical structure in terms of angle and speed change, and

$$v' = v + A(\theta)\,(v-v_*), \quad v'_* = v_* - A(\theta)\,(v-v_*).$$

The family of finite measures $\{f_t\,;\, 0 \le t \le T\}$ is called a weak solution to (2.1), if it satisfies

$$\frac{d}{dt}\int_{\mathbb{R}^2} \psi(v)\, f_t(dv) = \int_\Pi \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \big\{\psi\big(v + A(\theta)\,(v-v_*)\big) - \psi(v)\big\}\; B(|v-v_*|,\theta)\, f_t(dv)\, f_t(dv_*)\, d\theta$$

for any $t \in (0,T]$ and any globally Lipschitz continuous function $\psi : \mathbb{R}^2 \to \mathbb{R}$.

Loosely speaking, the equation (2.1) describes the time evolution of a gas due to the collision of particles which is given by the integral operator on the right side of (2.1). In this right side, the difference $f_t(v')\, f_t(v'_*) - f_t(v)\, f_t(v_*)$ describes essentially the change of density due to the result of the collision of particles. Some of them are lost as they gain other speeds and other are added as they achieve a certain speed. The function B describes the interaction between the particles, such as the strength of the collision. For more details and the exact physical meaning of the various terms, we refer the reader to [1] and [63].

The learned techniques for the α-stable CIR model are probably related to this model because on the one hand the function b corresponds to a stable measure close to zero and the function B is a power function. In fact, as Π is a compact set the behavior at infinity is not a problem in this case.

From the mathematical viewpoint, notice that the equation described in (2.1) is non-linear in f_t. In fact, it is linear in this variable on the left hand-side of the equation and quadratic on the right-hand side of the equation. This issue provokes additional problems that do not appear in the case of classical stochastic equations which have been considered so far.

Additionally, this problem is non-linear because we see the appearance of products of the density functions on the right hand side of 2.1. In this sense, one may say that this equation is of McKean-Vlasov type. We have also understood that one way to propose a solution is to find a good ansatz.

Moreover, if we assume the principles of conservation of mass, momentum and kinetic energy hold, then the following restrictions are imposed onto the solution of (2.1)

$$\int_{\mathbb{R}^2} f_t(dv) = \int_{\mathbb{R}^2} f_0(dv), \quad \int_{\mathbb{R}^2} v\, f_t(dv) = \int_{\mathbb{R}^2} v\, f_0(dv), \quad \int_{\mathbb{R}^2} |v|^2\, f_t(dv) = \int_{\mathbb{R}^2} |v|^2\, f_0(dv).$$

Without loss of generality, we shall suppose that $\int_{\mathbb{R}^2} f_0(dv) = 1$ and $\int_{\mathbb{R}^2} v\, f_0(dv) = 0$. Furthermore, we will suppose that f_0 is not a Dirac mass. In fact, we will suppose the slightly

stronger condition $\int_{\mathbb{R}^2} |v|^2 f_0(dv) > 0$. In fact, the conservation of momentum and kinetic energy are only heuristic in the above understanding of weak solutions. They are obtained by choosing the appropriate function ψ.

EXERCISE 16. *Prove that in the case that there is an $x \in \mathbb{R}^2$ such that $f_0(\{x\}) = 1$ then the solution to (2.1) is the trivial solution. That is, one needs at least two different speeds on the gas in order to create a non-trivial solution.*

The goal of the line of research we intend to introduce here is to prove that even in the case that f_0 does not have a smooth density, the collision strength characterized by B, implies the existence of a somewhat smooth density for f_t, for any $t > 0$, $\alpha \in (0, 1)$ and $\gamma \in (0, 1]$. Various of these cases have an associated physical interpretation. We refer the reader again to [63] for explanation and we just remark that there are the following cases in $\mathbb{R}^3$: for $\gamma = (s-5)/(s-1)$ and $\alpha = 2/(s-1)$,

- $s > 5$ is called the hard potential case,
- $s = 5$ is called the Maxwellian potential,
- $7/3 < s < 5$ is called the moderate soft potential case,
- $2 < s < 7/3$ is called as the very soft potential case,
- $s = 2$ is called the Coulomb potential,

where s is the parameter that characterizes the force exerted between particles, that is, two particles separated by a distance r in $\mathbb{R}^3$ exert each other a force of the order r^{-s} (which is expressed by B in (2.1)).

3. Stochastic representation

Here, we present a probabilistic representation for the solution of (2.1) which may be useful.

Let V_0 be an $\mathbb{R}^2$-valued random variable with the law $f_0(dv)$. Consider the $\mathbb{R}^2$-valued process $V = \{V_t\,;\, 0 \le t \le T\}$ with law $\{f_t;\, 0 \le t \le T\}$ determined by the stochastic differential equation:

$$V_t = V_0 + \int_0^t \int_\Pi \int_{\mathbb{R}^2} \int_0^{+\infty} A(\theta)\,(V_{s-} - v)\,\mathbf{1}_{[u,+\infty)}\big(|V_{s-} - v|^\gamma\big)\,d\mathcal{N}, \tag{3.1}$$

where $d\mathcal{N} = \mathcal{N}(ds, d\theta, dv, du)$ is the Poisson random measure over $[0,T] \times \Pi \times \mathbb{R}^2 \times [0, +\infty)$ with the intensity measure

$$d\widehat{\mathcal{N}} = \widehat{\mathcal{N}}(ds, d\theta, dv, du) := ds\, b(\theta) d\theta\, f_s(dv)\, du,$$

independent of V_0. This means that the space of jumps is a subset of $\mathbb{R}^4$.

EXERCISE 17. *Show that the total mass of the compensator is infinite. That is,* $\widehat{\mathcal{N}}([0,T] \times A \times \mathbb{R}^2 \times B) = \infty$ *if either* $A \subset \Pi$ *includes a neighborhood of* 0 *or* $B \subset [0, +\infty)$ *is a set of infinite measure. This implies that the construction of the above Poisson random measure has to be understood as the limit of Poisson random measures.*

EXERCISE 18. *Prove that if* V *is a solution for* (3.1) *then* $\mathbb{E}[V_t] = \mathbb{E}[V_0] = 0$ *and* $\mathbb{E}[|V_t|^2] = \mathbb{E}[|V_0|^2] > 0$ *for all* $t > 0$.

It is known that there exists a unique strong solution to the equation (3.1) such that the probability law of V_t is $f_t(dv)$ (cf. [6]-Proposition 2.1-(i), [62]-Theorems 4.1 and 4.2). In particular, we note that this is an infinite activity driving process, because $\int_0^\varepsilon b(\theta) d\theta = +\infty$ for any $\varepsilon > 0$. The reason why the stochastic integral in (3.1) is well defined in the variable θ is due to the behavior of A. In fact, $A_{ij}(\theta)\, b(\theta) = O(\theta^{1-\alpha})\delta_{ij} + O(\theta^{-\alpha})$ as $\theta \to 0$ for $i, j \in \{1, 2\}$. For the variable u, one has that the moments of the variable $|V_{s-} - v|^\gamma$ are bounded. In fact, even its exponential moments are bounded.

Therefore the goal is to provide simulation methods for f_t for $t > 0$.

CHAPTER 5

Lecture 3B: Using simulation methods to study densities

It is clear intuitively speaking that approximations could be used to study the laws of their limit processes.

We will try to be more specific about this topic.

Historically, Malliavin Calculus is a theory that was built in order to prove Hörmander's theorem about the regularity of the laws of hypoelliptic diffusions (although in the theory of partial differential equations is stated in a different but almost equivalent form). This differential calculus on Wiener-Poisson space is based on the fact that expectations are "integrals" over an infinite number of independent increment of a basic process, say X. The first step consists in introducing the differential operator:

$$D_s = \frac{\partial}{\partial \Delta X_s}$$

This derivative operator is closed using limit arguments. Then one creates its dual operator (equivalent to an extension of the stochastic integral). This dual operator gives rise to an integration by parts which can be used for studying the regularity of the law of the solution of a stochastic equations.

In the case of processes with jumps (usually called Malliavin Calculus for jump processes) one uses instead of X the jumps of the Lévy process. This has raised a number of theories that can be applied to a variety of problems (some references are Bichteler,Gravereaux and Jacod, Nualart-Nualart, Picard, Kunita, Ishikawa, Nualart, Bouleau-Denis). In each of these theories one chooses the appropriate random variables X as either jump sizes or times of jump. Then the definition of perturbation (derivative) has to be clearly defined. There are various approaches which are briefly explained at an undergraduate level in [46].

Recently, Fournier-Printems, Bally, V. and Caramellino have exploited the idea of using approximations as an interpolation procedure in order to obtain theoretical properties of densities of processes.

My intention in this chapter is to briefly explain what is the methodology used.

The most naive way to think about how to get information about limits from approximation is by sheer calculation strength.

For example, you may consider the Euler scheme for a diffusion. If we recall, (0.1), one may say that the density of $\bar{X}_{t_n}$ can be computed using n-convolutions. The problem is how to get information out of these n-convolutions. Let us fix a goal: Suppose that we want to obtain upper bound estimates for the densities of the limit X_T. Then, one would like to obtain upper bounds for transition densities and then take limits as $n \to \infty$. It is worth trying a bit to understand the difficulties. Anyway, it is a methodology that has not given too many results. Also we have to understand that we are operating under the belief that the increments $B_{t_{k+1}} - B_{t_k}$ should be generating the regularity of the law.

Actually in the Lévy case the richness of the laws creates a large number of possibilities.

EXERCISE 19. *Use the Box-Muller simulation method which is used to simulate standard Normal r.v.'s Z in order to create two integration by parts of the form*

$$\mathbb{E}[f'(Z)] = \mathbb{E}[f(Z)H],$$

for an appropriate random variable H and a class of test functions f.

Off course this is just a simple exercise as you can just use the law of the Normal random variable and obtain that $H = Z$.

Let U_1 and U_2 be independent samples chosen from the uniform distribution on the unit interval $(0, 1)$. Let

$$\begin{aligned} Z_0 =& R\cos(\Theta) = \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ Z_1 =& R\sin(\Theta) = \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{aligned}$$

Then Z_0 and Z_1 are independent random variables with a standard normal distribution. Prove that for bounded test functions with $f(r) = 0(r^\alpha)$, $\alpha > 0$ as $r \to 0$ one can choose $H = (\sin(\Theta))^{-1}(R - R^{-1})$. Here one should see that somehow R measures the length of the normal r.v.

This simple exercise presents a problem. When doing analysis, should you based your calculus on R or Θ? Or both?

In the history of the subject this was known as Partial Malliavin Calculus to indicate that one did not use all the randomness in the problem to analyze the regularity of certain random functional.

For example, one may want to analyze the the functional $F(X) = \sup_{t\in[0,T]} X_t$ which can be approximated using functions F_n

$$F_n \equiv F_n(E_1, ..., E_n, G_1, ..., G_n, U_1, ..., U_n) \overset{\mathcal{L}}{\Rightarrow} F(X)$$

One may try to do the analysis based on (some part of) the i.i.d. sequence $\{E_i, G_i, U_i\}_{i\in\mathbb{N}}$ if densities are explicit. I am choosing these symbols in order to explain an example later. To summarize.

Danger 1: The convergence rate is not good enough. R.N. Bhattacharaya and R. Ranga Rao, Normal Approximation and Asymptotic Expansions. SIAM Classics, 2010. Clearly, the simplest example is the random walk with Bernoulli r.v.'s. One can only obtain limited information about the limit (which in this simple example is the explicit Gaussian law) from the approximating sequence.

We will now consider another significant example: The supremum of a stable process. In order to study this functional we will use the approximations given by convex majorants for Lévy processes. This theory which started in the 50's$\sim$ has now been used for simulation purposes together with the multi-level method in such a way that the method converges exponentially fast.

This will be combined with Chambers-Mallows-Stuck decomposition method (recall that explicit stable laws are not available)('76, Kanter ('75)) which base the simulation on exponential r.v.'s $\{E_i\}_{i\in\mathbb{N}}$. That is, the "length" of stable increments.

The method of analysis will perform integration by parts using the i.i.d. exponentially distributed random variables $\{E_i\}$ and condition on the other r.v.'s $\{G_i, U_i\}_{i\in\mathbb{N}}$.

$$\mathbb{E}[f'(F_n)\Phi_n] = \mathbb{E}[D_E f(F_n)\Phi_n] = \mathbb{E}[f(F_n)H(F_n, \Phi_n)].$$

Therefore using the above formula for $f(x) = 1(x \geq K)$ will give us information about densities. But there is a danger here!

Danger 2: The used variables $\{E_i\}_{i\in\mathbb{N}}$ may not explain the main density behavior of $F(X)$.Because one concentrates on $F(X)$ one may loose track of other relevant path behavior

Danger 3: Need to analyze the explicit expression of $H(F_n, \Phi_n)$. In particular, one needs to see the optimal bounds and exploding moments of stable random variables may create problems.

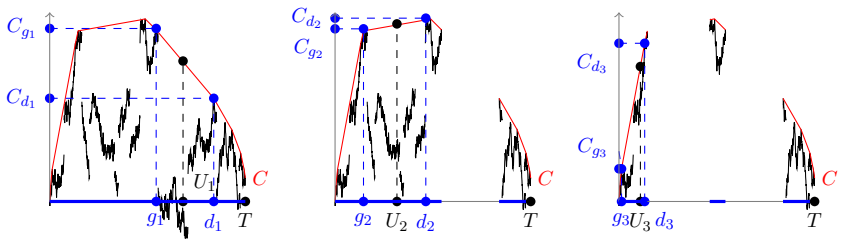


FIGURE 1.1. Selecting the first three faces of the concave majorant: U_1, U_2, U_3 are uniform with length T, $T-(d_1-g_1)$ and $T-(d_1-g_1)-(d_2-g_2)$, respectively. $C_{g_1}-C_{d_1}$ has the stable distribution with time g_1-d_1 with an associated positive or negative sign.

1. Convex majorants

2. Mathematical Definition of the convex majorant

Exponentially converging stick-breaking process: $\ell=(\ell_k)_{k\geq 1}$ on $[0,T]$, defined using $U_k\sim U[0,1]$

$$\begin{aligned}\ell_1 =& T(1-U_1)\\ \ell_k =& TU_1...U_{k-1}(1-U_k)\\ \mathbb{E}[\ell_k^{r/\alpha}] =& T^r\left(1+\frac{r}{\alpha}\right)^{-k}.\end{aligned}$$

i.i.d. stable r.v.'s $(S_k)_{k\geq 1}$ with parameters (α,ρ) (i.e. $S_k\overset{d}{=}X_1$).

$$\begin{aligned}\overline{X}_T =& X_+ = \sum_{k=1}^{\infty}\ell_k^{1/\alpha}[S_k]^+\\ X_T =& X_+ - X_- = X_+ - \sum_{k=1}^{\infty}\ell_k^{1/\alpha}[S_k]^-.\end{aligned}$$

Remark: One looses some information about the path!

Chambers-Mallows-Stuck/Kanter Next step: Probabilistic representation:(Chambers-Mallows-Stuck, $\alpha\in(0,1)\cup\{1\}\cup(1,2)$). E_k: length. G_k: oscillations

$$S_k\overset{\mathcal{L}}{=}E_k{}^{1-1/\alpha}G_k \qquad \text{and} \qquad G_k=g(V_k), \qquad k\in\mathbb{N},$$

for i.i.d. $Exp(1)\sim(E_k)_{k\geq 1}\perp(V_k)_{k\geq 1}\sim \mathrm{U}(-\frac{\pi}{2},\frac{\pi}{2})$

$$g(x)=\frac{\sin\left(\alpha\left(x+\pi\left(\rho-\frac{1}{2}\right)\right)\right)}{\cos^{1/\alpha}(x)\cos^{1-1/\alpha}\left((1-\alpha)x-\alpha\pi\left(\rho-\frac{1}{2}\right)\right)}, \qquad x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right).$$

Note that indeed $\mathbb{P}(S_k > 0) = \rho$.

$$\rho = \mathbb{P}(S_k > 0) \in [1 - 1/\alpha, 1/\alpha] \cap (0, 1).$$

semi-linear structure in the representation for $(X_T, \bar{X}_T) \equiv (X_+, X_-)$

$$\begin{aligned}\overline{X}_T =& X_+ = \sum_{k=1}^{\infty} \ell_k^{1/\alpha} E_k^{1-1/\alpha} [G_k]^+ \\ X_T =& X_+ - X_- = X_+ - \sum_{k=1}^{\infty} \ell_k^{1/\alpha} E_k^{1-1/\alpha} [G_k]^-.\end{aligned}$$

coordinate change and n-th order approx. to $X = (X_+, X_-)$: $X_n = (X_{+,n}, X_{-,n})$

$a_n := T^{1/\alpha}\kappa^n$ with $\kappa \in (0, 1)$. Let η_+ and $\eta_- \sim Exp(1)$

$$X_{\pm,n} = \sum_{k=1}^{n} \ell_k^{1/\alpha} E_k^{1-1/\alpha} [G_k]^{\pm} + a_n \eta_+^{1-1/\alpha}.$$

$n = 0$, $X_{\pm,0} \equiv 0$, $[x]^+ = \max\{x, 0\}$,

$0 \neq X_{\pm,n} - X_{\pm,n-1} = \ell_n^{1/\alpha}\ E_n^{1-1/\alpha}[G_n]^{\pm} + (a_n - a_{n-1})\eta_{\pm}^{1-1/\alpha}$

Intuitively, if the sequence $\{a_n\}$ decays too fast, then it will not serve its purpose. In particular, given the assumption below moments estimates will follow.

Assumption[A-κ] $a_n := T^{1/\alpha}\kappa^n$, $n \in \mathbb{N}$ where $\kappa^\alpha \in [\rho \vee (1 - \rho), 1)$.

Differential calculus based on the random variables $\{E_k\}_{k\in\mathbb{N}}$ gives the regularity of joint laws of $(X_T, \bar{X}_T)$. For explicit results, see [25].

Let $\mathcal{O} = \{(x, y) \in \mathbb{R}^2 : y > x \vee 0\}$, $n, m \geq 1$, $T > 0$.

THEOREM 2. *Let $F(x, y) := \mathbb{P}(X_T \leq x, \overline{X}_T \leq y)$ be the law of $(X_T, \overline{X}_T)$. Then $F \in C^\infty(\mathcal{O})$. Moreover, for any fixed $\alpha' \in [0, \alpha)$ there is $C > 0$ s.t. for all $(x, y) \in \mathcal{O}$,*

$$\begin{aligned}|\partial_x^n \partial_y^m F(x, y)| \leq\ & C y^{-m} (y - x)^{1-n-m} (2y - x)^{m-1} \\ & \times \min\left\{f_{\alpha'}^{00}(x, y), f_{\alpha'}^{01}(x, y), f_{\alpha'}^{10}(x, y), f_{\alpha'}^{11}(x, y)\right\}, \\ C \min\Big\{ & T^{2\frac{\alpha'}{\alpha}} (y - x)^{-n-\alpha'} y^{-m-\alpha'}, T^{\frac{\alpha'}{\alpha}(1-\rho)} (y - x)^{-n-\alpha'} y^{-m+\alpha'\rho}, \\ & T^{\frac{\alpha'}{\alpha}\rho} (y - x)^{-n+\alpha'(1-\rho)} y^{-m-\alpha'}, T^{-\frac{\alpha'}{\alpha}} (y - x)^{-n+\alpha'(1-\rho)} y^{-m+\alpha'\rho}\Big\}.\end{aligned}$$

4 Domains: Behavior at ∞. Behavior at 0. This result is optimal in time and almost optimal in space taking $\alpha' \approx \alpha$. The main reason is that there is some Chebyshev's type argument used.

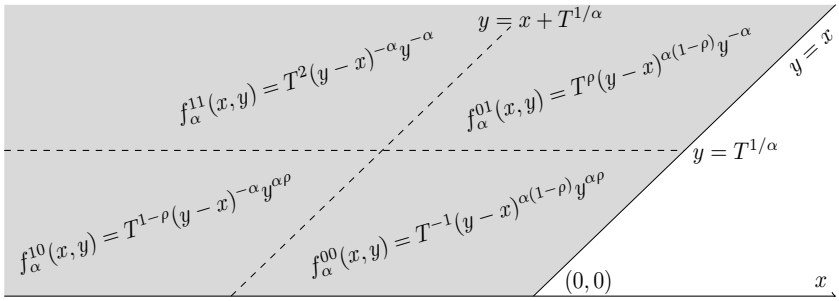


FIGURE 2.1. The set $\mathcal{O} = \{(x,y) \in \mathbb{R}^2 : y > \max\{x, 0\}\}$ (shaded in the figure) is the support of the joint density of $(X_T, \overline{X}_T)$. According to our Theorem, the support can be partitioned into 4 sub-regions according to which of the functions f_α^{ij}, $i,j \in \{0,1\}$, is the smallest in the (optimal) case $\alpha' = \alpha$. The dotted lines correspond to the change in time regime.

Bibliography

[1] R. Alexandre. A review of Boltzmann equation with singular kernels. *Kinet. Relat. Models*, 2(4):551–646, 2009.

[2] Patrik Andersson, Arturo Kohatsu-Higa, Tomooki Yuasa, Second order probabilistic parametrix method for unbiased simulation of stochastic differential equations. Stochastic Processes and their Applications, Volume 130, Issue 9, 2020, Pages 5543-5574,

[3] Approximations of Small Jumps of Lévy Processes with a View Towards Simulation. Søren Asmussen and Jan Rosiński Journal of Applied Probability Vol. 38, No. 2 (Jun., 2001), pp. 482-493

[4] Anis Al Gerbi, Benjamin Jourdain and Emmanuelle Clément Ninomiya–Victoir scheme: Strong convergence, antithetic version and application to multilevel estimators. Monte Carlo Methods and Its applications, 2016, 22, 3, 197-228.

[5] Rainer Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch.*, 13(3):381–401, 2009.

[6] V. Bally and N. Fournier. Regularization properties of the 2D homogeneous Boltzmann equation without cutoff. *Probab. Theory Related Fields*, 151(3-4):659–704, 2011.

[7] Vlad Bally and Arturo Kohatsu-Higa. A probabilistic interpretation of the parametrix method. *Ann. Appl. Probab.*, 25(6):3095–3138, 2015.

[8] Vlad Bally, Marie-Pierre Bavouzet, and Marouen Messaoud. Integration by parts formula for locally smooth laws and applications to sensitivity computations. *Ann. Appl. Probab.*, 17(1):33–66, 02 2007.

[9] Vlad Bally and Lucia Caramellino. *Stochastic Integration by Parts*. Advanced Courses in Mathematics - CRM Barcelona. Springer International Publishing, 2016.

[10] Bally, V. and Rey, C. (2016). Approximation of Markov semigroups in total variation distance. Electron. J. Probab., 21:Paper No. 12, 44

[11] D. Belomestny, T. Nagapetyan. 'Multilevel path simulation for weak approximation schemes with application to Lévy-driven SDEs' Bernoulli, 23(2):927-950, 2017.

[12] A Beskos, O Papaspiliopoulos, GO Roberts, P Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). Journal of the Royal Statistical Society: Series B. 2006. 68 Issue 3, Pages 333-382.

[13] Exact simulation of diffusions A Beskos, GO Roberts The Annals of Applied Probability 15 (4), 2422-2444.

[14] ϵ-Strong simulation of the Brownian path A Beskos, S Peluchetti, G Roberts Bernoulli 18 (4), 1223-1248, 2012

[15] Alexandros Beskos, Omiros Papaspiliopoulos, and Gareth O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

[16] Klaus Bichteler, Jean-Bernard Gravereaux, and Jean Jacod. *Malliavin Calculus for Processes with Jumps*. Stochastic Monographs : Theory and Applications of Stochastic Processes, Vol 2. Gordon and Breach Science Publishers, 1987.

[17] Jose Blanchet, Xinyun Chen, and Jing Dong. ϵ -Strong simulation for multidimensional stochastic differential equations via rough path analysis. Ann. Appl. Probab., Volume 27, Number 1 (2017), 275-336.

[18] Nicolas Bouleau and Laurent Denis. Energy image density property and the lent particle method for poisson measures. *Journal of Functional Analysis*, 257(4):1144 – 1174, 2009.

[19] Mark Broadie and Özgür Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.*, 54(2):217–231, 2006.

[20] Improved local approximation for multidimensional diffusions: The G-rates. S. Bodnarchuk, D. Ivanenko, A. Kohatsu-Higa and A. Kulik Theor. Probability and Math. Statist. 101 (2020), 13-38

[21] Thomas Bonis. Rates in the Central Limit Theorem and diffusion approximation via Stein's Method, 2018, arxiv1506.06966.

[22] Exact simulation of jump-diffusion processes with Monte Carlo applications. B Casella, GO Roberts. Methodology and Computing in Applied Probability 13 (3), 449-473, 2011

[23] S. Dereich, Multilevel Monte Carlo for Lévy-driven SDE's with Gaussian correction, Ann. Appl. Probab. 21 (1) (2011) 283–311

[24] Steffen Dereich, Felix Heidenreich, A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations, Stochastic Processes and their Applications, Volume 121, Issue 7, 2011, Pages 1565-1587.

[25] Joint density of a stable process and its supremum: regularity and upper bounds, Jorge González Cázares and Arturo Kohatsu-Higa and Aleksandar Mijatović, 2020 arXiv.

[26] Bin Chen, Cornelis W. Oosterlee, and Hans Van Der Weide. A low-bias simulation scheme for the sabr stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 15(02):1250016, 2012.

[27] Davis, M. H. A. "Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models." Journal of the Royal Statistical Society. Series B (Methodological), vol. 46, no. 3, 1984, pp. 353–388.

[28] Ronald A. Doney and Mladen S. Savov. The asymptotic behavior of densities related to the supremum of a stable process. *Ann. Probab.*, 38(1):316–326, 2010.

[29] N. Fournier. Simulation and approximation of Lévy-driven stochastic differential equations. ESAIM P&S, Vol. 15, 249-269, 2011.

[30] N. Fournier. On pathwise uniqueness for stochastic differential equations driven by stable Lévy processes. Ann. Inst. H. Poincar é Probab. Statist. 49(1), 138-159, 2013.

[31] Avner Friedman. *Partial differential equations of parabolic type.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

[32] Stochastic equations of non-negative processes with jumps. Zongfei Fu, Zenghu Li. Stochastic Processes and their Applications. Volume 120, Issue 3, March 2010, Pages 306-330

[33] Michael B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[34] Michael B. Giles and Lukasz Szpruch. Antithetic multilevel Monte Carlo estimation for multidimensional SDEs. In *Monte Carlo and quasi-Monte Carlo methods 2012*, volume 65 of *Springer Proc. Math. Stat.*, pages 367–384. Springer, Heidelberg, 2013.

[35] Mike Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[36] Peter W. Glynn. Randomized estimators for time integrals. Technical report, Mathematics Research Center, University of Wisconsin, Madison, 1983.

[37] Peter W. Glynn and Ward Whitt. The asymptotic efficiency of simulation estimators. *Oper. Res.*, 40(3):505–520, 1992.

[38] Exact simulation problems for jump-diffusions FB Gonçalves, GO Roberts Methodology and Computing in Applied Probability 16 (4), 907-930, 2014

[39] Lemaire, Vincent, Thieullen, Michèle and Thomas, Nicolas. Exact Simulation of the Jump Times of a Class of Piecewise Deterministic Markov Processes. Journal of Scientific Computing. 2018-3, 1776- 1807, vol. 75.

[40] Jorge I. González Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. Exact simulation of the extrema of stable processes. *Adv. in Appl. Probab.*, 51(4), 12 2019.

[41] István Gyöngy and Miklós Rásonyi. A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic Process. Appl.*, 121(10):2189–2200, 2011.

[42] Y. Ishikawa. *Stochastic Calculus of Variations: For Jump Processes*. De Gruyter Studies in Mathematics. De Gruyter, 2016.

[43] Peter Imkeller, Niklas Willrich, Solutions of martingale problems for Lévy-type operators with discontinuous coefficients and related SDEs, Stochastic Processes and their Applications, Volume 126, Issue 3, 2016, Pages 703-734,

[44] Alpha-CIR model with branching processes in sovereign interest rate modeling Finance and Stochastics. Issue 3/2017 Authors: Ying Jiao, Chunhua Ma, Simone Scotti

[45] K. Kawazu, Sh. Watanabe, "Branching processes with immigration and related limit theorems", Theory Probab. Appl., 16:1 (1971), 36–54

[46] Kohatsu-Higa A. and Takeuchi A. *Jump SDEs and the Study of Their Densities : A self-study book*. Universitext. Springer, 2019.

[47] Hiroshi Kunita. *Stochastic Flows and Jump-Diffusions*. Springer Singapore, 2019.

[48] On a positivity preserving numerical scheme for jump-extended CIR process: the alpha-stable case L Li, D Taguchi BIT Numerical Mathematics 59 (3), 747-774

[49] On the Euler–Maruyama scheme for spectrally one-sided Lévy driven SDEs with Hölder continuous coefficients L Li, D Taguchi Statistics &Probability Letters 146, 15-26

[50] Li, Zenghu; Mytnik, Leonid. Strong solutions for stochastic differential equations with jumps. Ann. Inst. H. Poincaré Probab. Statist. 47 (2011), no. 4, 1055–1067

[51] Don McLeish. A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.*, 17(4):301–315, 2011.

[52] David Nualart. *The Malliavin calculus and related topics. Probability and its Applications*. Springer-Verlag, New York, 2006.

[53] David Nualart and Eulalia Nualart. *Introduction to Malliavin Calculus*, page 158–181. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2018.

[54] Jean Picard. On the existence of smooth densities for jump processes. *Probab. Th. Rel. Fields*, 105:481–511, 1996.

[55] Jean Picard. Erratum to: On the existence of smooth densities for jump processes. *Probab. Th. Rel. Fields*, 147:711–713, 2010.

[56] Jim Pitman and Gerónimo Uribe Bravo. The convex minorant of a Lévy process. *Ann. Probab.*, 40(4):1636–1674, 2012.

[57] Enrico Priola. "Pathwise uniqueness for singular SDEs driven by stable processes." Osaka J. Math. 49 (2) 421 - 447, June 2012.

[58] On the exact and -strong simulation of (jump) diffusions M Pollock, AM Johansen, GO Roberts Bernoulli 22 (2), 794-856.

[59] Chang han Rhee and P.W. Glynn. A new approach to unbiased estimation for sde's. In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, pages 1–7, Dec 2012.

[60] Chang-Han Rhee and Peter W. Glynn. Unbiased estimation with square root convergence for sde models. *Oper. Res.*, 63(5):1026–1043, 2015.

[61] E. Rio, Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. 45 (2009) 802–817.

[62] H. Tanaka. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete*, 46(1):67–105, 1978/79.

[63] C. Villani. A review of mathematical topics in collisional kinetic theory. In *Handbook of mathematical fluid dynamics, Vol. I*, pages 71–305. North-Holland, Amsterdam, 2002.

[64] Zhang, Xilíng. A multi-dimensional central limit bound and its application to the euler approximation for Lévy-SDEs. ESAIM: PS, 2019, 23, 112-135.

[65] Rafał Weron. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics & Probability Letters*, 28(2):165 – 171, 1996.

[66] Jie Xiong, Jiayu Zheng, Xiaowen Zhou (2019) Unique strong solutions of Lévy processes driven stochastic differential equations with discontinuous coefficients, Stochastics, 91:4, 592-612

[67] Jiayu Zheng, Jie Xiong, Pathwise uniqueness for stochastic differential equations driven by pure jump processes, Statistics & Probability Letters, Volume 130, 2017, Pages 100-104,