Lecture 9

Why jumps processes?

The interest in Lévy processes arises from their being the fundamental building blocks in stochastic models whose evolution in time exhibits sudden changes in value. Many of these models have been suggested and extensively studied in the area of *mathematical finance* (see e.g [2] for an explanation of the necessity of considering jumps when modelling asset returns). However, they have played a central role in many other fields of science such as *physics*, *engineering*, *economics*, *actuarial science* among many others. Below we present some of the more common models with jumps.

Examples

Jump diffusion processes:

$$X_t = \gamma t + \sigma W_t + \sum_{i=1}^{N_t} Y_i,$$

where

- $(W_t)_{t\geq 0}$ is a standard Brownian motion;
- $(\sum_{i=1}^{N_t} Y_i)_{t\geq 0}$ is a compound Poisson process with intensity λ;
- the two processes are independent.

To specify the model we must still specify the law of Y_1. It is important to that regard to chose a law that reflects what we believe the behavior of extremal events is, since, as we have already seen, they are intimately linked with the Lévy measure: the tail behavior of the Lévy measure determines to a large extent the tail behavior of the probability distribution of the process. Let us quickly discuss two different situations:

1. **Merton model:** it corresponds to the case where $Y_1 \sim \mathcal{N}(\mu, \delta^2)$.

   In this particular case we have a series expansion for the probability density of X. Indeed, from

   $$\mathbb{P}(X_t \in A) = \sum_{k=0}^{\infty} \mathbb{P}(X_t \in A | N_t = k)\mathbb{P}(N_t = k)$$

we get that the probability density p_t of X_t satisfies

$$p_t(x) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k \exp\Big(- \frac{(x-\gamma t - k\mu)^2}{2(\sigma^2 t + k\delta^2)}\Big)}{k! \sqrt{2\pi(\sigma^2 t + k\delta^2)}},$$

that is the probability density of X_t can be written as a quickly converging series.

2. **Kou model:** it corresponds to the case where the law of Y_1 has a density with respect to the Lebesgue measure given by an asymmetric exponential of the form

$$f(x) = p\lambda_+ e^{-\lambda_+ x} \mathbf{1}_{(0,\infty)}(x) + (1-p)\lambda_- e^{-\lambda_- |x|} \mathbf{1}_{(-\infty,0)}(x),$$

   with $\lambda_+, \lambda_- > 0$ and $p \in [0,1]$.

   A formula for the probability distribution is not available in a closed form and in this case the tails are semi-heavy.

   For applications of these models in finance and an overview of their properties, see e.g. [9, 7].

α-stable processes

A remarkable property of Brownian motion is its selfsimilarity property: if W is a standard Brownian motion on $\mathbb{R}$ then

$$\forall a > 0, \quad \left(\frac{W_{at}}{\sqrt{a}}\right)_{t \geq 0} \overset{d}{=} (W_t)_{t \geq 0}.$$

A natural generalization is then the following.

Definition 1. *A real Lévy process X is said to be* selfsimilar *if*

$$\forall a > 0, \ \exists b(a) > 0 : \quad \left(\frac{X_{at}}{b(a)}\right)_{t \geq 0} \overset{d}{=} (X_t)_{t \geq 0}. \tag{1}$$

Let us now translate (1) in terms of characteristic functions. Let φ_{X_t} denote the characteristic function of X_t. By Proposition 1 in Lecture 7 we know that there exists a continuous function ψ such that $\varphi_{X_t}(z) = e^{-t\psi(z)}$. Therefore,

$$\mathbb{E}\Big[e^{\frac{iuX_{at}}{b(a)}}\Big] = \exp\left(- at\psi\Big(\frac{u}{b(a)}\Big)\right) = \Big(\varphi_{X_t}\Big(\frac{u}{b(a)}\Big)\Big)^a.$$

In other words, (1) is equivalent to

$$\forall a > 0, \ \exists b(a) > 0 : \quad \varphi_{X_t}(z)^a = \varphi_{X_t}(zb(a)), \quad \forall z \in \mathbb{R}.$$

Let us now highlight the strict connection that there exists between selfsimilar Lévy processes and stable distributions.

Definition 2. *A real random variable X is said to have a* stable distribution *if for every $a > 0$ there exist $b(a) > 0$ and $c(a) \in \mathbb{R}$ such that*

$$\varphi_X(z)^a = \varphi_X(zb(a)) + e^{izc(a)}, \quad \forall z \in \mathbb{R},$$

where φ_X denotes the characteristic function of X. It is said to have a strictly stable distribution *if for every $a > 0$ there exists $b(a) > 0$ such that*

$$\varphi_X(z)^a = \varphi_X(zb(a)), \quad \forall z \in \mathbb{R}.$$

Moreover, it can be shown that for every stable distribution there exists a constant $\alpha \in (0, 2]$, called the *index of stability*, such that $b(a) = a^{\frac{1}{\alpha}}$. Stable distributions with $\alpha = 2$ are only Gaussian.

By definition, it follows that a selfsimilar Lévy process has a strictly stable distribution for all times t. For this reason, they are often called *strictly stable Lévy processes* and, in particular, they satisfy

$$\forall a > 0, \quad \left(\frac{X_{at}}{a^{1/\alpha}}\right)_{t\geq 0} \stackrel{d}{=} (X_t)_{t\geq 0}. \tag{2}$$

More generally, an α-stable Lévy process satisfies (2) up to a trasnslation:

$$\forall a > 0,\ \exists c \in \mathbb{R} \quad (X_{at})_{t\geq 0} \stackrel{d}{=} \left(a^{1/\alpha} X_t + ct\right)_{t\geq 0}.$$

What is very useful in practice is the following characterization of the Lévy measure of a real-valued stable process. More precisely it can been proven that for any $0 < \alpha < 2$, the Lévy density (i.e. the density of ν with respect to the Lebesgue measure) of a real-valued stable process with index of stability α, also called α-stable processes, is of the form:

$$\frac{\nu(dx)}{dx} = \frac{A}{x^{\alpha+1}} \mathbf{1}_{x>0} + \frac{B}{|x|^{\alpha+1}} \mathbf{1}_{x<0}, \tag{3}$$

for some positive constants A and B.

From (3) joined with Proposition 1 in Lecture 8, we deduce that α-stable Lévy processes never admit a second moment (unless $\alpha = 2$) and they have a first moment only if $\alpha > 1$.

A detailed study of the properties of stable Lévy processes can be found in e.g. [10, 12].

Remark 1. *By definition, a Lévy measure is a measure concentrated on $\mathbb{R} \setminus \{0\}$ such that*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty \quad \text{and} \quad \int_{|x|\geq 1} \nu(dx) < \infty. \tag{4}$$

Let us notice that any transformation of ν respecting the integrability conditions in (4), will lead to the definition of a new Lévy measure and hence a new Lévy process. For instance, for any $\theta \in \mathbb{R}$ such that $\int_{|x|\leq 1} e^{\theta x} \nu(dx) < \infty$, the measure $\widetilde{\nu}$ defined by

$$\widetilde{\nu}(dx) := e^{\theta x} \nu(dx)$$

is a Lévy measure.

We may also consider asymmetric versions of this transformation by defining:

$$\widetilde{\nu}(dx) := \nu(dx) \Big(\mathbf{1}_{x>0} e^{-\lambda_+ x} + \mathbf{1}_{x<0} e^{-\lambda_- |x|} \Big), \tag{5}$$

for positive λ_+, λ_-. Then, if ν is a Lévy measure also $\widetilde{\nu}$ defined in (5) is. We are thus building a new Lévy process whose large jumps are "tempered".

Tempered stable processes:

A tempered stable process is a real Lévy process with a zero diffusion coefficient and a Lévy density of the form

$$\frac{\nu(dx)}{dx} = \frac{c_-}{|x|^{1+\alpha}} e^{-\lambda_-|x|}\mathbf{1}_{x<0} + \frac{c_+}{x^{1+\alpha}} e^{-\lambda_+ x}\mathbf{1}_{x>0}, \tag{6}$$

where $0 < \alpha < 2$ and the parameters $c_-, c_+, \lambda_-, \lambda_+$ are strictly positive. The process having ν as in (6) as a Lévy measure corresponds to a Lévy process obtained by multiplying the Lévy density of an α-stable process by a two-sided exponential factor. The exponentially decreasing factor has the effect of tempering the large jumps so that the process has finite moments, while retaining the same behavior for the small jumps. That is why this kind of processes are called *tempered stable processes.* They play a central role in mathematical finance. A possible reference among many others is [3].

Subordinators:

A *subordinator* is a real non-decreasing (a.s.) Lévy process. The following characterization holds (see e.g. [1]):

Theorem 1. *If X is a subordinator, then its characteristic exponent takes the form*

$$\psi(u) = -iu\gamma - \int_0^\infty (e^{iuy} - 1)\nu(dy), \tag{7}$$

where $\gamma \geq 0$ and the Lévy measure ν satisfies the additional requirements

$$\nu((-\infty, 0)) = 0 \quad \text{and} \quad \int_0^\infty (y \wedge 1)\nu(dy) < \infty.$$

Conversely, any mapping from $\mathbb{R}$ to $\mathbb{C}$ of the form (7) is the characteristic exponent of a subordinator.

Subordinators are very popular in finance mainly because they can be seen as a time deformation and thus used to construct the so called *time changed Lévy processes.* An example of subordinator is given by the Gamma process: that is a Lévy process having an absolutely continuous law with respect to the Lebesgue measure and with density given by:

$$p_{\alpha,\beta}(x) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x}\mathbf{1}_{(0,\infty)}(x),$$

for $\alpha, \beta > 0$. Here Γ denotes the Gamma function, i.e. $\Gamma(\beta) = \int_0^\infty t^{\beta-1}e^{-t}dt$. It can be proven that the Lévy measure of a Gamma process with parameters α and β is given by $\nu(dx) = \beta x^{-1}e^{-\alpha x}dx$ and more generally, the Lévy triplet is given by:

$$\Big(-\beta \int_0^1 e^{-\alpha x}dx, 0, \beta x^{-1}e^{-\alpha x}dx \Big).$$

Inference for Lévy processes

In the previous lectures we have seen that the law of a Lévy process X is uniquely determined by its Lévy triplet (γ, σ^2, ν). As statisticians, a first natural question is then how to estimate such a triplet from the data. The data in this case could be:

- The continuous observation of a trajectory of X on the time horizon $[0, T]$;
- The discrete observation of a trajectory of X on the time horizon $[0, T]$, i.e. we have at our disposal n observations of the process at times $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$:

  $$X_{t_1}, \dots, X_{t_n}.$$

  Intuitively, the estimation procedure should be different depending on the sampling rate of the observations, i.e. we expect that good estimators in the regime

  $$\max_{i=1,\dots,n} |t_i - t_{i-1}| \to 0, \qquad \text{as } n \to \infty$$

  should not perform as nicely when, for instance,

  $$\min_{i=1,\dots,n} |t_i - t_{i-1}| \to C, \qquad \text{as } n \to \infty$$

  for some constant $C > 0$.

  To better understand, let us suppose to have uniform discrete observations of X at sampling rate Δ on the time horizon $[0, T]$:

  $$X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$$

  with $n\Delta = T$. Then, two regimes are possible:

  - High frequency observations: $\Delta = \Delta_n \to 0$ as $n \to \infty$.
  - Low frequency observations: $\Delta = \Delta_n \to C > 0$ or $\Delta = \Delta_n \to \infty$ as $n \to \infty$.

Clearly, observing a trajectory of X in continuous time is more informative than having only access to discrete observations of X. However, in certain cases, if Δ goes to zero quickly enough, the statistical model associated with the continuous observation of X may contain asymptotically the same amount of information as the statistical model associated with the high frequency observations of X. Such a notion of "asymptotically contains the same amount of information as" can be made rigorous, for instance by considering the Le Cam theory of comparison between experiments. This is however beyond the scope of these lectures. Here, let us just stress that making inference for discretely observed Lévy processes is harder than making inference for continuously observed Lévy processes. Indeed, if we have at our disposal only discrete observations of X we do not know how many jumps have occurred between two observations and how much of the increment comes from the continuous dynamic of a diffusion component instead.

Certainly, a statistical model based on discrete observations of a Lévy process X is more realistic that the one based on the continuous observation of X. For that reason, even though the analysis in the first scenario is much harder, we will focus only on the case where the data are discrete observations of a Lévy process. More precisely, we will start our investigation by considering pure jumps Lévy processes.

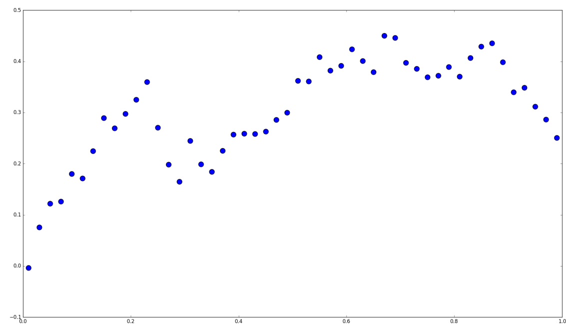

Figure 1: Do the data above come from the discrete observation of a continuous Lévy process (as in Figure 2) or from a pure jumps Lévy process (as in Figure 3)?

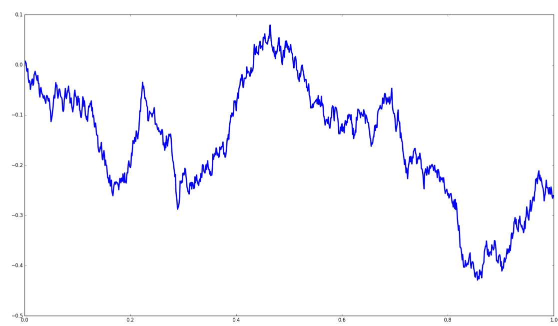

Figure 2: Trajectory of a Brownian motion

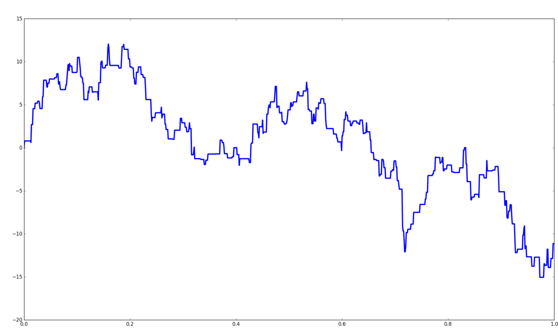

Figure 3: Trajectory of a compound Poisson process

Inference for pure jumps Lévy processes

As we have already pointed out in the previous lectures, the jumps dynamics of a Lévy process is entirely dictated by its Lévy measure. So, a natural question is how to estimate the Lévy measure of a pure jumps Lévy process X having at our disposal only discrete observations of X. From a mathematical point of view we have two powerful tools: the Lévy-Khintchine formula and the Lévy-Itô decomposition, that is a characterization of X via its characteristic function or a result on the structure of the paths of X. We will see how these two different approaches allow us to propose different estimators of the quantities of interest. We will begin by considering the easiest case where X is a compound Poisson process observed in a high frequency regime.

Inference for compound Poisson processes: a direct approach

Let X be a compound Poisson process with intensity λ and jump measure F. Let us suppose that F is absolutely continuous with respect to the Lebesgue measure and let us denote by $f = \frac{dF}{dx}$ the density. In other words

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0,$$

where N is a Poisson process with intensity λ independent of the sequence of i.i.d. random variables $(Y_i)_{i\geq 1}$ with common law F. In particular X is a Lévy process with Lévy measure $\nu = \lambda F$. Since F is absolutely continuous with respect to the Lebesgue measure, ν also is. Let us denote by ρ the Lévy density, i.e. $\rho = \frac{d\nu}{dx}$. Observe that $\lambda = \nu(\mathbb{R} \setminus \{0\})$.

Suppose now that we want to estimate ν, equivalently ρ, from high frequency observations of X. More precisely, suppose that we have at our disposal equidistant observations of X at the sampling rate Δ over the time interval T:

$$X_0, X_\Delta, X_{2\Delta}, \cdots, X_{\frac{n-1}{n}\Delta}, X_{n\Delta} \quad \text{with } n\Delta = T.$$

We shall consider the asymptotics $\Delta \to 0$ and $T \to \infty$ as $n \to \infty$. Our aim is to estimate ρ from the data $(X_{i\Delta})_{i=0}^n$. This is a non parametric estimation problem since ρ is an element of a functional space and hence an infinite dimensional space. Using the fact that $\rho = \lambda f$, a natural estimator for ρ is:

$$\widehat{\rho}_n(x) = \widehat{\lambda}_n \widehat{f}_n(x),$$

where $\widehat{\lambda}_n$ is a "good" estimator of λ and $\widehat{f}_n(x)$ is a "good estimator" of f. That is the problem of estimating ρ can be reduced to the problem of estimating the parameter λ and the density f of the random variables $(Y_i)_{i\geq 1}$. Set

$$Z_i := X_{i\Delta} - X_{(i-1)\Delta}, \quad i = 1, \dots, n.$$

Since X has stationary and independent increments, the random variables $Z_1, \dots, Z_n$ are i.i.d. with the same law as X_Δ. We will use the sample $(Z_i)_{i=1}^n$ to draw inference about λ and f. Let us begin by addressing the question of the (parametric) estimation of λ that is easier than the (non parametric) estimation of f.

In Lecture 8 we have seen that

$$\lim_{t\to 0}\frac{\mathbb{P}(X_t\neq 0)}{t}=\nu(\mathbb{R}\setminus\{0\})=\lambda.$$

This means that for Δ small enough $\lambda\Delta\approx\mathbb{P}(X_\Delta\neq 0)$. Now, a natural estimator of $\mathbb{P}(X_\Delta\neq 0)$ is given by its empirical counterpart:

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq 0}=:\frac{\widehat{n}(0)}{n}.$$

This suggests to define

$$\widehat{\lambda}_n:=\frac{\widehat{n}(0)}{n\Delta}. \tag{8}$$

Let us study the square-mean error of the estimator of λ defined in (8). Using the bias-variance decomposition we have that

$$\mathbb{E}[|\lambda-\widehat{\lambda}_n|^2]=\big(\mathbb{E}[\widehat{\lambda}_n]-\lambda\big)^2+\mathrm{Var}\big(\widehat{\lambda}_n\big).$$

- **Bias:** Since F is absolutely continuous with respect to the Lebesgue measure we have:

  $$\mathbb{P}(Z_i\neq 0)=\mathbb{P}(X_\Delta\neq 0)=\mathbb{P}(N_\Delta\neq 0)=1-e^{-\lambda\Delta}.$$

  Therefore

  $$\mathbb{E}[\widehat{\lambda}_n]=\frac{1}{n\Delta}\mathbb{E}\Big[\sum_{i=1}^n\mathbf{1}_{Z_i\neq 0}\Big]=\frac{1-e^{-\lambda\Delta}}{\Delta}=\lambda+O(\Delta).$$

- **Variance:** From the previous computations we know that $\mathbb{E}[\widehat{n}(0)]=n(1-e^{-\lambda\Delta})$. Moreover,

  $$\begin{aligned}\mathbb{E}[\widehat{n}(0)^2]&=\mathbb{E}\Big[\sum_{i,j=1}^n\mathbf{1}_{Z_i\neq 0}\mathbf{1}_{Z_j\neq 0}\Big]=n\mathbb{P}(Z_1\neq 0)+\mathbb{E}\Big[\sum_{i\neq j=1}^n\mathbf{1}_{Z_i\neq 0}\mathbf{1}_{Z_j\neq 0}\Big]\\&=n\mathbb{P}(Z_1\neq 0)+n(n-1)(\mathbb{P}(Z_1\neq 0))^2\\&=n(1-e^{-\lambda\Delta})(1+(n-1)(1-e^{-\lambda\Delta})).\end{aligned}$$

  Thus,

  $$\begin{aligned}\mathrm{Var}(\widehat{n}(0))&=n(1-e^{-\lambda\Delta})(1+(n-1)(1-e^{-\lambda\Delta}))-(n(1-e^{-\lambda\Delta}))^2\\&=n(1-e^{-\lambda\Delta})e^{-\lambda\Delta}.\end{aligned}$$

  It follows that

  $$\mathrm{Var}(\widehat{\lambda}_n)=\frac{\mathrm{Var}(\widehat{n}(0))}{n^2\Delta^2}=\frac{(1-e^{-\lambda\Delta})e^{-\lambda\Delta}}{n\Delta^2}=O\Big(\frac{1}{T}\Big).$$

To sum up, we have just proved that

$$\mathbb{E}[|\lambda - \widehat{\lambda}_n|^2] = O\Big(\frac{1}{T} + \Delta^2\Big),$$

that is the rate of convergence of $\widehat{\lambda}_n$ in mean-square risk is of order of $1/T + \Delta^2$.

Remark 2. *Another estimator of the intensity of a compound Poisson process can be proposed by directly using that*

$$\mathbb{P}(Z_i \neq 0) = 1 - e^{-\lambda\Delta}.$$

This leads to

$$\widetilde{\lambda}_n := -\frac{1}{\Delta} \ln\Big(1 - \frac{\widehat{n}(0)}{n}\Big). \tag{9}$$

For Δ small enough (9) and (8) are morally the same estimator but the advantage of (9) is that in principle it should be a good estimator of λ even in a low-frequency regime.

Let us now discuss the problem of the estimation of f, the common density of the i.i.d. random variables $(Y_i)_{i\geq 1}$. At first sight one may think that we only need to solve a "classical" non-parametric density estimation problem but the situation is more complicated than that since we do not observe the variables $(Y_i)_{i\geq 1}$ but only $(X_{i\Delta})_{i=0}^n$ or, equivalently, $(Z_i)_{i=1}^n$. In a high frequency regime the probability that there are two or more jumps between consecutive observations is very low, more precisely it is of order Δ^2 :

$$\mathbb{P}(N_\Delta \geq 2) = 1 - (e^{-\lambda\Delta} + \lambda\Delta e^{-\lambda\Delta}).$$

Heuristically we can then say that f is the density of X_Δ conditionally on the fact that $X_\Delta \neq 0$: in this way we can relate f to something that is observable so that we are finally left to solve a density estimation problem. We can then choose the way we prefer to estimate f: kernel estimators, projector estimators, etc... However, there is still a main difficulty, the sample size in this case is $\widehat{n}(0)$, that is the number of observations is random! This complication can be overcome by proving that $\widehat{n}(0)$ concentrates around T. So the rate of convergence of $\widehat{f}_n$ will be the same as the rate of convergence of an estimator of a density from T i.i.d. realizations of random variables having such a density. In particular, it depends on the regularity assumptions made on f (e.g. f belongs to a Hölder space with regularity s, or a Sobolev space, or a Besov space, etc...). A possible reference is [5].

References

[1] Bertoin, J., Lévy Processes, Cambridge University Press, 1996.

[2] Carr, P., Geman, H., Madan, D. B., and Yor, M., The fine structure of asset returns: An empirical investigation*, The Journal of Business, 2002.

[3] Cont, R., Bouchard, J.-P., and Potters, M., Scaling in financial data: Stable laws and beyond. Springer, 1997.

[4] Cont, R., and Tankov, P., Financial Modelling With Jumps Processes, Chapman & Hall/CRC, 2004.

[5] Duval, C., Density estimation for compound Poisson processes from discrete data, Stochastic Processes and their Applications, 2013.

[6] Duval, C., Mariucci, E., Compound Poisson approximation to estimate the Lévy density, ArXiv:1702.08787, 2017.

[7] Kou, S., A jump-diffusion model for option pricing, Management Science, 2002.

[8] Kyprianou, A., Fluctuations of Lévy Processes with Applications, Springer, 2104.

[9] Merton, R., Option pricing when underlying stock returns are continuous, J. Financial Economics, 1976.

[10] Samorodnitsky, G., and Taqqu, M. S., Stable Non-Gaussian Random Processes, Chapman and Hall, 1994.

[11] Sato, K., Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.

[12] Zolotarev, V. M., One-Dimensional Stable Distributions, American Mathematical Society, 1986.