POINT PROCESSES AND THEIR APPLICATION IN FINANCE

E. LÖCHERBACH

(1) Stochastic processes : first definitions and properties. The classical Poisson process; Thinning and superposition, Law of Large Numbers and Central Limit Theorem. Poisson processes with inhomogeneous rates.
(2) One-dimensional linear Hawkes process : definition, construction, properties, non-explosion, stationarity, law of large numbers, mean number of jumps, empirical covariation across time scales, longtime behavior.
(3) Multivariate linear Hawkes processes; Clustering representation.
(4) Application to financial time series.

1. Stochastic processes

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 1.1. *1) $X = (X_t, t \in I)$ is called* **stochastic process** *defined on $I \subset \mathbb{R}$, taking values in $\mathbb{R}^d$, if for all $t \in I$, X_t is a random variable.*
2) We call the application that maps, for fixed $\omega \in \Omega$, $I \ni t \mapsto X_t(\omega) \in \mathbb{R}^d$ the **trajectory of the process**.

Remark 1.2. *1) Most common choices are $I = \mathbb{R}_+, I = \mathbb{R}$, or $I = \mathbb{N}$ (for Markov chains, for instance).*
2) The trajectory of X is a random function from I to $\mathbb{R}^d$. Without any regularity properties for the moment.

Example 1.3. *1) Take $I = \mathbb{N}$ and $X = (X_n, n \geq 0)$ where the X_n are i.i.d.*
2) Markov chains.
3) Brownian motion.

In the sequel we shall be interested in modeling random events such as market buy or sell orders, price evolution, times at which an insurance has to pay claims of the policy holders, ...

Example 1.4 (Capital evolution of an insurance)**.** *Two types of random events : the times of the claims and the sizes of the claims.*

$$C_t = C_0 + rt - \sum_{k : T_k < t} U_k,$$

where r is the contribution rate of the policy holders and where the U_k are the losses.
To begin with, let us concentrate on the first source of randomness, the arrival times of the claims.

1.1. **Random arrival times.** We write $T_0 = 0 < T_1 < T_2 \ldots \leq T_n \leq \ldots$ for the successive random times of arrival (of a claim, of a market order...). Define for any $t \geq 0$,

$$N_t = \text{ number of claims before } \leq \text{ time } t \; = \#\{n \geq 1 : T_n \leq t\} = \sum_{n=1}^{\infty} 1_{\{T_n \leq t\}} = \sum_{k=0}^{\infty} k 1_{\{T_k \leq t < T_{k+1}\}}.$$

Proposition 1.5. *1) We have that* $\{T_n \le t\} = \{N_t \ge n\}$.
2) If $T_n \uparrow +\infty$ *then the trajectories of* $N = (N_t, t \ge 0)$ *are non-decreasing, piecewise constant, càdlàg, starting from* $N_0 = 0$, *having jumps or size* $+1$ *only. More precisely,*

(1) $N_0(\omega) = 0$ *for all* ω,
(2) *if* $s \le t$, *then* $N_s(\omega) \le N_t(\omega)$,
(3) *the process of jumps,* $\Delta N_t := N_t - N_{t-}$ *satisfies :* $\Delta N_t \in \{0, 1\}$, $\Delta N_t(\omega) = 1 \Leftrightarrow$ *there exists* $n \ge 1$, *such that* $t = T_n(\omega)$.

In the above scenario, we say that N is the counting process associated to the sequence of jump events $(T_n, n \ge 0)$.

Definition 1.6. *Let* N *be the counting process associated to the sequence of jump events* $(T_n, n \ge 0)$, *with* $T_0 = 0$. *We say that* N *is a* **Poisson process** *with intensity* $\lambda > 0$ *if* $\tau_n := T_n - T_{n-1}, n \ge 1$, *satisfy :* τ_n *are i.i.d.,* $\sim \exp(\lambda)$.

Remark 1.7. *We call* τ_n *the inter-arrival times (the waiting times for the next event).*

Remark 1.8. *Recall the loss of memory property for the exponential law.*

Definition 1.9 (Gamma distribution)**.** *Let* $\alpha, \beta > 0$. *A positive random variable* X *is said to follow the Gamma distribution* $\Gamma(\alpha, \beta)$ *if* X *has density*

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}$$

on $\mathbb{R}_+$. *In particular, the density of* $\Gamma(n, \lambda)$ *is given by*

$$f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}.$$

Proposition 1.10. *1)* $T_n \sim \Gamma(n, \lambda)$ *for all* $n \ge 1$.
2) In particular, $N_t \sim Poiss(\lambda t)$, $EN_t = Var N_t = \lambda t$ *for all* $t \ge 0$.
3) For all $t > 0$, $P(\Delta N_t = 1) = 0$.
4) $(N_t)_t$ *is a Lévy process, in particular it has independent and stationary increments.*

Proposition 1.11 (Large deviations)**.** *Let* N_t *be a Poisson process having intensity* 1. *Then*

$$P(\sup_{s \le t} |N_s - s| \ge \varepsilon) \le 2 \exp(-t h(\varepsilon / t)),$$

for all $\varepsilon < t$, *where* $h(x) = (1 + x) \log(1 + x) - x$.

Proof. We use the fact that $M_t := N_t - t$ is a martingale. We upper bound one part of the inequality, for any positive $\theta > 0$, by

$$P(\sup_{s \le t} N_s - s \ge \varepsilon) \le P(\sup_{s \le t} e^{\theta(N_s - s)} \ge e^{\theta \varepsilon}).$$

Using that $S_t := e^{\theta(N_t - t)}$ is a submartingale, plus Doob's maximal inequality for positive submartingales, we deduce from this that

$$P(\sup_{s \le t} N_s - s \ge \varepsilon) \le e^{-\theta \varepsilon} E(e^{\theta(N_t - t)}) = e^{-\theta \varepsilon - \theta t} E(e^{\theta N_t}).$$

But

$$E(e^{\theta N_t}) = e^{t(e^\theta - 1)}.$$

This upper bound is minimized (in θ) choosing $\theta = \ln(1 + \varepsilon/t)$. Plugging this into the above bounds, one gets the upper bound

$$e^{-th(\varepsilon/t)}.$$

For the other term

$$P(\sup_{s \le t} s - N_s \ge \varepsilon),$$

we get similarly the upper bound

$$e^{-th(-\varepsilon/t)}.$$

One concludes observing that $h(-x) \ge h(x)$ for all $x \in [0,1]$. $\square$

We conclude this section with a very useful characterisation of Poisson processes, which is due to Watanabe.

Theorem 1.12. *Let N_t be a simple and locally finite point process on $\mathbb{R}_+$, that is, for all $t \ge 0$,*
(i) $N_t < \infty$ almost surely.
(ii) $\Delta N_t \in \{0, 1\}$.
Suppose that N_t is $\mathbb{F}-$adapted and that for some fixed constant $\lambda > 0$,

$$E \int_{]0,T]} Z_t dN_t = E \int_0^T Z_t \lambda dt, \tag{1.1}$$

for all $T > 0$ and for all real valued, left-continuous processes Z_t which are $\mathbb{F}-$adapted. Then N is a Poisson process with intensity λ, and for all intervals $0 < a < b$, we have that $N_b - N_a$ is independent of $\mathcal{F}_a$.

1.2. **Poisson processes and random measures.** We start with a simple remark. Knowing the process N is equivalent to knowing the whole sequence of jump times (T_n) which is still equivalent to knowing the **random counting measure** on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ given by

$$\Pi(\omega) := \sum_{n=1}^{\infty} \delta_{T_n(\omega)}$$

and which acts in the following way : for all $C \in \mathcal{B}(\mathbb{R}_+)$, we have

$$\Pi(\omega, C) := \Pi(\omega)(C) = \sum_{n=1}^{\infty} 1_C(T_n(\omega))$$

which is nothing else then the number of jump times falling into the set C.

It is clear that if $T_n \uparrow \infty$, then for all fixed $\omega \in \Omega$, $\Pi(\omega, \cdot)$ is a $\sigma-$finite measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. Moreover, for all $C \in \mathcal{B}(\mathbb{R}_+)$ fixed, $\Pi(C)$ is a random variable. Π is also called **point process on $\mathbb{R}_+$.** In general, we will not distinguish between Π and N.

If the jump sequence (T_n) comes from a Poisson process, then Π is also called Poisson random measure on $\mathbb{R}_+$.

More generally we introduce the following notion of random measure.

Definition 1.13. *A random measure N on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ is a mapping $(\omega, C) \mapsto N(\omega, C)$ from $\Omega \times \mathcal{B}(\mathbb{R}_+)$ to $\mathbb{R}_+ \cup \{\infty\}$ such that*
(i) for each $\omega \in \Omega$, the application $\mathcal{B}(\mathbb{R}_+) \ni C \mapsto N(\omega, C)$ is a measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$.
(ii) for each $C \in \mathcal{B}(\mathbb{R}_+)$, $\omega \mapsto N(\omega, C)$ is measurable.
Moreover, N is called $\sigma-$finite if there exists a sequence of measurable sets K_n such that $\bigcup_n K_n = \mathbb{R}_+$ and $N(\omega, K_n) < \infty$ for all ω.

Remark 1.14. *Evidently, the above definition can be extended to any random measure on* $(E, \mathcal{E})$ *for any measurable space* $(E, \mathcal{E})$.

Definition 1.15. *Let* Π *be any random measure on* $\mathbb{R}_+$. *Define for any* $C \in \mathcal{B}(\mathbb{R}_+)$

$$\nu(C) := E(\Pi(C)) = \sum_{k=0}^{\infty} kP(\{\omega : \Pi(\omega, C) = k\}). \tag{1.2}$$

Exercise : Check that ν defines a positive measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. ν is called the *intensity measure* of Π (or of N).

Check that, if N is the Poisson process, then $\nu = \lambda \times$ Lebesgue.

Proposition 1.16. *Let* N *be a random measure on a measurable space* $(E, \mathcal{E})$ *having intensity measure* ν *and let* $\varphi : E \to \mathbb{R}$ *be measurable and positive (or* $\varphi \in L^1(\nu)$*). Then*

$$E(\int_E \varphi dN) = \int_E \varphi d\nu.$$

Proposition 1.17. *Let* N *be the Poisson process having intensity* λ *and let* Π *be the associated counting measure. Write* $\nu(dx) := \lambda dx$. *Then*

1. *for all* $C \in \mathcal{B}(\mathbb{R}_+)$, $\Pi(C) \sim Poiss(\nu(C))$.
2. *for all* $C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R}_+)$ *which are mutually disjoint,* $\Pi(C_1), \ldots, \Pi(C_n)$ *is an independent family of random variables.*

Proof. One checks that the assertion holds for $C_k =]t_{k-1}, t_k]$, for some $t_0 < t_1 < \ldots < t_n$. □

One can easily extend the above notion of a PRM from one dimension to two dimensions (or even more, but we shall not need this here).

1.3. **Two dimensional random measures.**

In the sequel, $X_n, n \geq 1$, will denote random variables taking values in $\mathbb{R}_+ \times \mathbb{R}_+$. We put

$$M := \sum_{n \geq 1} \delta_{X_n}.$$

Definition 1.18. M *is called PRM on* $\mathbb{R}_+ \times \mathbb{R}_+$ *having intensity* $\lambda dx_1 dx_2$ *if*
1) for all $C \in \mathcal{B}(\mathbb{R}_+^2)$, $M(C) \sim Poiss(\lambda |C|)$, *where* $|C|$ *denotes the Lebesgue measure of the set* C,
2) for all $C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$ *which are mutually disjoint,* $M(C_1), \ldots, M(C_n)$ *is an independent family of random variables.*

Proposition 1.19 (Which shows how to construct/simulate from $M_{|K}$ where $K \subset \mathbb{R}_+ \times \mathbb{R}_+$ is some compact set.)**.** *Choose* $N \sim Poiss(\lambda |K|)$ *and, conditionally on* $N = n$, n *i.i.d. random variables* $X_1, \ldots, X_n$ *which are uniformly distributed on* K *and independent of* N *(that is,* $P(X_1 \in C) = \frac{|C \cap K|}{|K|}$*). If we put*

$$\tilde{\Pi} := \sum_{k=1}^{N} \delta_{X_k},$$

then $\tilde{\Pi} \stackrel{\mathcal{L}}{=} M_{|K}$.

1.4. **Thinning procedures.** This section tries to answer to the question what $2-$d PRM are good for, if we are only interested in one-dimensional counting processes.

The answer is that we can actually use them to construct $1-$d counting processes.

Example 1.20. *Put* $N_t := M([0,t]\times[0,\lambda])$ *(make a picture to understand what we are just doing!). Show that the such constructed* N_t *is a Poisson process with intensity* λ.

Proof. Indeed, $N_t - N_s = M(]s,t]\times[0,\lambda]) \sim Poiss(\lambda(t-s))$. Apart from this we need to show that N has stationary and independent increments, but this follows directly from the properties of M. □

Example 1.21 (Poisson process with inhomogeneous rate). *Let* $\lambda_t : \mathbb{R}_+ \to \mathbb{R}_+$ *be a locally integrable function, that is,* $\int_0^t \lambda_s ds < \infty$ *for all* $t \geq 0$. *Define*

$$N_t := M(\{(s,x) : 0 \leq s \leq t, 0 \leq x \leq \lambda_s\}).$$

PICTURE !!!

Then we have that

$$P(N_t = 0) = P(T_1 > t) = e^{-\int_0^t \lambda_s ds}$$

and

$$N_t \sim Poiss(\int_0^t \lambda_s ds).$$

Jargon : *One says that* T_1 *occurs at rate* λ_t, *and that the process jumps at rate* λ_t. *At times* t *such that* $\lambda_t >> 1$, *jumps are more likely to occur than at times* s *such that* $\lambda_s \sim 0$.

Inhomogeneous rates can be used to model seasonalities.

Proposition 1.22. *Let* N_t *be an inhomogeneous Poisson process of rate* λ_t. *Then the joint law of its first* n *jumps* $T_1, \ldots, T_n$, *conditionally on* $\{N_T = n\}$, *is given by the order statistics of* n *i.i.d. random variables* $V_1, \ldots, V_n$, *chosen according to the density* $f(t) = \frac{\lambda(t) 1_{\{t \leq T\}}}{\int_0^T \lambda(s)ds}$.

Proof. It is easy to show that the joint law of $(T_1, \ldots, T_n)$ is given by

$$\lambda(t_1)\cdot\lambda(t_2)\ldots\lambda(t_n)e^{-\int_0^{t_n}\lambda(s)ds}1_{\{t_1\leq t_2\leq\ldots\leq t_n\}}.$$

Therefore, for any $g : \mathbb{R}_+^n \to \mathbb{R}_+$ measurable and bounded, and writing $\|\lambda\|_T := \int_0^T \lambda(s)ds$,

$$\begin{aligned} I\!E[g(T_1,,\ldots,T_n)1_{\{N_T=n\}}\ ] &= \int_{\mathbb{R}_+^n} 1_{\{t_1<t_2<\ldots<t_n<T\}} e^{-\int_0^T \lambda(s)ds} \prod_{i=1}^n \lambda(t_i) g(t_1,\ldots,t_n)dt_1\ldots dt_n \\ &= e^{-\|\lambda\|_T}\frac{\|\lambda\|_T^n}{n!}\int_{[0,T]^n}\left(\prod_{i=1}^n \frac{\lambda(t_i)}{\|\lambda\|_T}\right) g(t_1,\ldots,t_n)dt_1\ldots dt_n, \end{aligned}$$

implying the result, since $e^{-\|\lambda\|_T}\frac{\|\lambda\|_T^n}{n!} = I\!P(N_T = n)$.

□

1.5. **Poisson processes and martingales.** As in the preceding section, we take a PRM M on $\mathbb{R}_+^2$ having intensity measure dx_1dx_2 and a locally integrable function $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$. We define

$$N_t = M(\{(s,x) : 0 \leq s \leq t, 0 \leq x \leq \lambda_s\})$$

and

$$\mathcal{F}_t := \sigma(M(C) : C \subset [0,t]\times\mathbb{R}_+)$$

which can be interpreted as *past before time t*.

Theorem 1.23. *Then $N_t - \int_0^t \lambda_s ds$ is a martingale which is square integrable. In particular, we have the identities $EN_t = Var N_t = \int_0^t \lambda_s ds$.*

Proof. One uses that the centered random measure $\tilde M(ds, dz) := M(ds, dz) - dsdz$ is a martingale measure which implies that

$$M_t := \int_{[0,t]} \int_0^\infty \varphi(s, z) \tilde M(ds, dz)$$

is a martingale for all predictable processes φ satisfying that $E \int_0^t \int_0^\infty |\varphi(s, z)| dsdz < \infty$ for all $t > 0$. In this case, it is well-known (but has to be proved) that

$$Var M_t = \int_0^t \int \varphi^2(s, z) dsdz.$$

We apply this result with the particular choice

$$\varphi(s, z) = 1_{\{z \le \lambda_s\}}.$$

□

Finally, as a corollary of the above considerations, we are able to state a strong law of large numbers and an associated Central Limit Theorem.

Theorem 1.24. *Suppose that N is a time-inhomogeneous Poisson process such that*

$$\frac{1}{t} \int_0^t \lambda_s ds \to \sigma^2 > 0 \tag{1.3}$$

as $t \to \infty$. Then

$$\frac{1}{t} N_t \to \sigma^2$$

almost surely as $t \to \infty$ and

$$\sqrt{t} \left(\frac{N_t}{t} - \frac{1}{t} \int_0^t \lambda_s ds \right) \to \mathcal{N}(0, \sigma^2)$$

(weak convergence as $t \to \infty$). In particular, if $\lambda_t \equiv \lambda > 0$, we have convergence

$$\sqrt{t} \left(\frac{N_t}{t} - \lambda \right) \to \mathcal{N}(0, \lambda).$$

2. One-dimensional linear Hawkes processes

2.1. **Motivation.** The most important features of a Poisson process - independently of the fact whether it has a constant rate or a time inhomogenous one - is the fact that subsequent waiting times are independent the one of the others. Therefore, successive events are independent the one of the others. This is not what can be observed in most practical examples.

Consider for instance an order driven market (see e.g. [1]) where participants can submit orders of different types (limit orders, market orders, cancellations, ...). Just to clarify ideas, let us recall that

- **Limit orders** are orders that specify an upper/lower price limit at which one is willing to buy or to sell a certain number of shares. Of course, there is no certainty that this order will be executed. However, there will be some *market impact* of this order, since it changes the anticipation that other participants of the market may have concerning the future evolution of the market.
- **Market orders** are orders that trigger an immediate buy or sell transaction for a certain number of shares. Such orders are immediately executed.

- **Cancellations** Any of the above orders may be cancelled.

Several empirical facts :
– The Poisson hypothesis for the arrival of orders in time is not empirically satisfied.
– There are – of course ! – dependencies between order arrivals.
– More importantly : It is most likely that a given order, for example a buy limit order, even it is does not change the price of the share immediately, will **stimulate** other buy limit and market order events. And, for example, a buy limit order that changes the mid price (since the order price is higher than the best bid price) will trigger a strong increase in the probability of sell market orders.

Conclusion In general there will be strong temporal dependencies between events and some events **will trigger future events** - this is called *self-excitation*.

2.2. **Branching Poisson processes.** Ingredients :

– a fixed rate μ describing the arrival rate of **exogenous market events**.

– a function $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\int_0^\infty h(t)dt < \infty$ describing the **influence of an event on future events**. More precisely, $h(t)$ describes how an event at time s will trigger a future event at time $s+t$. h is also called *kernel function*, or *memory kernel* or *offspring kernel*.

Consider a homogeneous Poisson processes $(I_t, t \geq 0)$ having intensity μ and, independent of I, an i.i.d. collection $(Z^{n,k}, n, k \geq 1)$ of inhomogeneous Poisson processes having rate $h(t)$ at time $t \geq 0$ each. $I(t)$ describes the exogenous market events and the $Z^{n,k}$ will describe the subsequent events which are **triggered by previous events**. We fix a time horizon $T > 0$ and want to describe the order driven market on $[0, T]$. This is done in a hierarchical way as follows.

Generation 0: Simulate all exogenous market events on $[0, T]$, that is, a realization of I on $[0, T]$. Call the events $T_n^{(0)}, 0 \leq n \leq I_T$.

Generation 1: We work conditionally on the realization of the 0 generation. Suppose there is an exogenous market event at time $s = T_n^{(0)}$. This triggers future events on $]s, +\infty[$ according to a Poisson process having intensity $h(t-s)$ on $]s, T]$. If we consider the superposition of all such events, we obtain an overall intensity

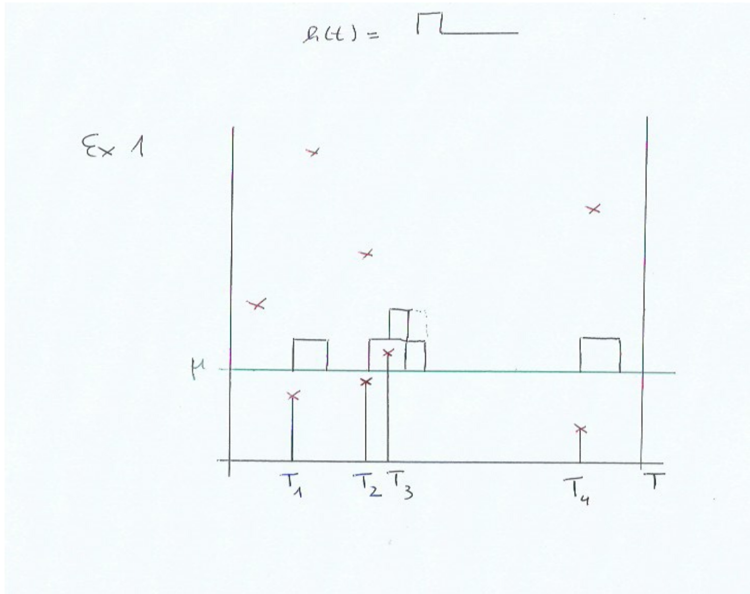
$$\lambda^{(1)}(t) = \sum_{n : T_n^{(0)} < t} h(t - T_n^0).$$

The associated counting process can we represented as

$$N_t^{(1)} = \sum_{n=1}^{I_t} Z^{n,1}_{t - T_n^{(0)}}.$$

Generation $k+1$: We work conditionally on the realization of generation k. Each point $T_n^{(k)}$ of generation k triggers future events on $]T_n^{(k)}, +\infty[$, independently of anything else, according to a Poisson process having intensity $h(t - T_n^{(k)})$ on $]T_n^{(k)}, +\infty[$. The superposition of all these events gives rise to a counting process $N^{(k+1)}$ which has an overall intensity

$$\lambda^{(k+1)}(t) = \sum_{n : T_n^{(k)} < t} h(t - T_n^{(k)}).$$



We have

$$N_t^{(k+1)} = \sum_{n=1}^{N_t^{(k)}} Z^{n,k+1}_{t-T_n^{(k)}}.$$

Definition 2.1. *We put $N_t := I_t + \sum_{k=1}^\infty N_t^{(k)}$. N is called linear Hawkes processes with immigration rate μ and memory kernel h.*

Put

$$\lambda_t := \mu + \int_{]0,t[} h(t-s) dN_s = \mu + \sum_{k=0}^\infty \sum_{n: T_n^{(k)} < t} h(t - T_n^{(k)}). \tag{2.4}$$

The relation with the just constructed process N_t is that we can show :

Theorem 2.2. *λ_t is the* **stochastic intensity** *of N_t in the sense that*

$$N_t = \int_{]0,t]} \int_0^\infty 1_{\{z \le \lambda_s\}} M(ds, dz),$$

where M is a 2−dimensional Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+$ having intensity $dsdz$.

Remark 2.3. *1). The picture shows clearly how the process is self-exciting : each jump increases the intensity of future jumps.*
2) The function $h(t) \ge 0$ models the **impact** *of every event on future events and* **how this impact vanishes with time.**
3) Hawkes processes model any sort of self-contamination : earthquakes (Hawkes), number of clicks on a webpage, market orders, spikes of neurons, ...

Remark 2.4. *Writing $T_n, n \geq 0$, for the successive jumps of the Hawkes process N_t, we can construct the intensity process in a recursive way. We have*

$$\lambda_t = \begin{cases} \mu, & \text{if } t \leq T_1 \\ \mu + h(t - T_1), & \text{if } T_1 < t \leq T_2 \\ \mu + h(t - T_1) + h(t - T_2), & \text{if } T_2 < t \leq T_3 \\ \vdots \end{cases}$$

We notice that λ_t does only depend on jumps of N **strictly** *before time t. This property is termed "predictable". Hence $t \mapsto \lambda_t$ is a predictable process. In fact, it is left-continuous, if the function h is continuous. (And one can show that all left-continuous processes are predictable. The proof that it is left-continuous relies also on the fact that N does not explode, that is, that $N_t < \infty$ almost surely for all $t-$ see later.)*

2.3. **Comparison to a branching process and total number of jumps.**

2.3.1. *Galton-Watson process.* Recall the classical construction : let $(X_{n,k})$ be i.i.d. with values in $\{0, 1, 2, \ldots\}$, with common mean m. Put $Z_0 = 1$ and

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n,k}.$$

Then $EZ_n = m^n$, and one knows about the dichotomy extinction/survival (explosion).

2.3.2. *Back to Hawkes.* Each "offspring"-process $Z^{n,k}$ gives rise to a total number of jump events which is given by $Z^{n,k}_\infty \sim Poiss(\int_0^\infty h(t)dt)$. Therefore, the average number of events triggered by a single "ancestor"-event is given by the associated expectation which equals $\|h\|_1 = \int_0^\infty h(t)dt$. Write for short $m := \|h\|_1$. Then m is the average number of orders directly triggered by a single event. In a similar way, m^2 is the number of grandchildren orders triggered by this event, and thus, if we call **cluster** all the descendants of a given event, the average size of a cluster is given by

$$\sum_{k \geq 1} m^k.$$

We clearly see that in the supercritical case $m > 1$ and in the critical case $m = 1$, the average size of a cluster is infinite, while in the subcritical case $m < 1$, the average size equals

$$\frac{m}{1-m} = \frac{\|h\|_1}{1 - \|h\|_1}.$$

Since only the first jump is an exogenous market order event, the **average proportion of endogenously triggered events** among all events, which can be interpreted as **degree of endogeneity of the market,** is given by

$$[\frac{m}{1-m}]/[1 + \frac{m}{1-m}] = m = \|h\|_1.$$

Remark 2.5. *Therefore, $\|h\|_1$ is a measure for the number of additional events that any single event triggers. The closer $\|h\|_1$ is to 1, the stronger the endogeneity of the market. Experts from financial statistics have therefore proposed to estimate $\|h\|_1$ from date, see Hardiman, Bercot and Bouchaud 2013.*

2.4. **Non-explosion and existence.**

Theorem 2.6. *If $h \in L^1_{loc}$, then the Hawkes process N_t with intensity λ_t given by* (2.4) *exists and we have $EN_t < \infty$ for all $t > 0$.*

Proof. The proof of existence is done by Picard iteration. Let M be the 2−dimensional Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+$ having intensity $dtdz$ (the read points in Figure 1). We define successively $\lambda_t^{(0)} := \mu$, $N_t^{(0)} = M(\{(s,x) : 0 \le s \le t, 0 \le x \le \lambda^{(0)}(s)\})$, and then more generally, for all $n \ge 0$,

$$\lambda_t^{(n+1)} := \mu + \int_{(0,t)} h(t-s) dN_s^{(n)},\ N_t^{(n+1)} = M(\{(s,x) : 0 \le s \le t, 0 \le x \le \lambda^{(n+1)}(s)\}).$$

By construction, for any fixed t, $n \mapsto \lambda_t^{(n)}$ and $n \mapsto N_t^{(n)}$ are increasing. Thus there exist the corresponding limits

$$\lambda_t := \lim_n \lambda_t^{(n)}, N_t = \lim_n N_t^{(n)}$$

and it is easy to show that

$$N_t = M(\{(s,x) : 0 \le s \le t, 0 \le x \le \lambda(s)\})$$

and that

$$\lambda_t = \mu + \int_{]0,t[} h(t-s) dN_s.$$

(One has to use that $M(\{(s,x) : x = \lambda_s\}) = 0$ almost surely.)

Therefore, we have shown that N_t exists. But it might have infinite expectation. Or even be infinite itself. To finish the proof and to show that $EN_t < \infty$, we will use a couple of lemmas. □

Recall that for a time-inhomogeneous Poisson process with rate λ_t we have that

$$EN_t = \int_0^t \lambda_s ds \text{ and } N_t - \int_0^t \lambda_s ds \text{ is a martingale.}$$

This remains true in the context of Hawkes processes. For that sake let

$$\mathcal{F}_t = \sigma\{N_s : s \le t\}$$

be the natural filtration of the process. Moreover we put

$$\Lambda_t := \int_0^t \lambda_s ds. \tag{2.5}$$

Theorem 2.7 (Another characterization of stochastic intensity)**.** (1)

$$(N_t - \Lambda_t, t \ge 0) \text{ is a local martingale.} \tag{2.6}$$

(2) *We have that for all $s < t$,*

$$E(N_t - N_s|\mathcal{F}_s) = E(\int_s^t \lambda_u du|\mathcal{F}_s). \tag{2.7}$$

Remark 2.8. *In mathematical terms, property* (2.6) *means that the increasing process Λ_t is the predictable compensator of N_t.* (2.7) *implies n particular that*

$$EN_{t\wedge T_n} = E\int_0^{t\wedge T_n} \lambda_s ds,$$

and, by monotone convergence,

$$EN_t = \int_0^t E(\lambda_s) ds.$$

Proof. The most elegant way of proving the Theorem is to use that the centered random measure $\tilde{M}(ds,dz) := M(ds,dz) - dsdz$ is a martingale measure which implies that

$$M_t := \int_{[0,t]} \int_0^\infty \varphi(s,z)\tilde{M}(ds,dz)$$

is a martingale for all predictable processes φ satisfying that $E\int_0^t\int_0^\infty |\varphi(s,z)|dsdz < \infty$ for all $t > 0$. Then one applies this to

$$\varphi(s,z) := 1_{\{z \le \lambda_s\}} 1_{\{s < T_n\}}.$$

We have

$$\int_0^t \int_0^\infty |\varphi(s,z)|dsdz = \mu(t \wedge T_n) + \int_0^{t\wedge T_n} \int_0^s h(s-u)dN_u ds$$

and we use Fubini (see exercises) and the fact that $N_s \le n$ on $s \le T_n$ to show that

$$\int_0^{t\wedge T_n} \int_0^s h(s-u)dN_u ds = \int_0^{t\wedge T_n} h(t\wedge T_n - s)N_s ds \le n\int_0^t h(s)ds < \infty$$

to verify the integrability condition on φ. □

Proof of the fact that $EN_t < \infty$ for all $t > 0$. Using the same argument, we obtain for $m_t := EN_{t\wedge T_n}$ the following inequality

$$m_t \le \mu t + \int_0^t h(t-s)m_s ds.$$

Since $\int_0^t h(s)ds < \infty$, we may choose A such that $\int_0^t h(s)1_{\{h(s)\ge A\}}ds < \frac{1}{2}$. We write

$$\int_0^t h(t-s)m_s ds = \int_0^t h(t-s)1_{\{h(t-s)\ge A\}}m_s ds + \int_0^t h(t-s)1_{\{h(t-s)\le A\}}m_s ds$$
$$\le \frac{1}{2}m_t + A\int_0^t m_s ds.$$

This implies that

$$m_t/2 \le \mu t + A\int_0^t m_s ds,$$

and the classical Gronwall lemma allows to conclude that

$$m_t \le 2\mu t e^{2At}.$$

The rhs of the above inequality does not depend on n, therefore, letting $n \to \infty$ and writing $T_\infty := \lim T_n = \sup T_n$, we obtain that

$$EN_{t\wedge T_\infty} \le 2\mu t e^{2At}.$$

Since $N_{t\wedge T_\infty} = N_{T_\infty} = \infty$ on $T_\infty \le t$, this clearly implies that $T_\infty > t$ almost surely (and thus, since this holds for all t, $T_\infty = \infty$ almost surely), whence

$$EN_t \le 2\mu t e^{2At}.$$

□

Corollary 2.9. *The above arguments show in particular that $m_t = EN_t$ satisfies*

$$m_t = \mu t + \int_0^t h(t-s)m_s ds. \tag{2.8}$$

Moreover, putting $M_t := N_t - \int_0^t \lambda_s ds$ (which is a martingale according to the proof of Theorem 2.7), we obtain that $X_t := N_t - EN_t$ satisfies

$$X_t = N_t - EN_t = M_t + \int_0^t h(t-s)X_s ds. \tag{2.9}$$

Remark 2.10. *If $\|h\|_1 < 1$, the above proof can be considerably shortened in the following way. Since $m_s \le m_t$ for all $s \le t$, we have that*

$$m_t \le \mu t + m_t \int_0^\infty h(t)dt$$

implying that

$$m_t \le \frac{\mu t}{1 - \|h\|_1}.$$

Corollary 2.11 (Corollary of Theorem 2.7)**.** *In the proof we have in particular used that*

$$\int_0^t \int_0^s h(s-u)dN_u ds = \int_0^t h(t-s)N_s ds.$$

Corollary 2.12 (Corollary of Theorem 2.7)**.** *We also have that*

$$E\int_0^t h(t-s)dN_s = E\int_0^t h(t-s)\lambda_s ds.$$

Proof. Take $\varphi(s,z) = h(T-s)1_{\{z \le \lambda_s\}}1_{\{s \le T\}}$ for some fixed $T > 0$. Then we obtain that for all $t \le T$,

$$\int_0^t \int \varphi(s,z)M(ds,dz) = \int_0^t h(T-s)dN_s$$

and on the other hand

$$\int_0^t \int \varphi(s,z)dsdz = \int_0^t h(T-s)\lambda_s ds.$$

The expectations of the two expressions are the same implying the assertion, if we take $T = t$. □

2.5. **Excursion on Hazard rates.** Let us come back to (2.7). We can rewrite this as follows :

$$P(N \text{ jumps within } (t, t+h]|\mathcal{F}_t) = E(N_{t+h} - N_t|\mathcal{F}_t) = \lambda_t h + o(h)$$

as $h \to 0$. We can even write the following :

Proposition 2.13 (Hazard rate)**.**

$$P(N_t - N_s = 0|\mathcal{F}_s) = e^{-\int_s^t \bar\lambda_u du}, \tag{2.10}$$

where $\bar\lambda_u = \mu + \int_{]0,s]} h(u-v)dN_v$ (which equals λ_u on $N \equiv 0$ on $]s,t]$). In particular, we obtain that

$$P(N_t - N_s = 0) = E\left[e^{-\int_s^t \bar\lambda_u du}\right].$$

Proof. We start with the following simple observation

$$1_{\{N_t - N_s = 0\}} = 1 - \int_{]s,t]} 1_{\{N(]s,u[)=0\}} dN_u.$$

For all $A \in \mathcal{F}_s$, the process

$$u \mapsto 1_{[s,t]}(u)1_A 1_{\{N(]s,u[)=0\}}$$

is predictable such that we can replace, under the expectation, the integral with respect to dN_u by $\lambda_u du$ and obtain

$$E1_A 1_{\{N_t - N_s = 0\}} = P(A) - \int_s^t E(1_A 1_{\{N(]s,u[)=0\}} \lambda_u) du.$$

We then use that for Lebesgue almost all u,

$$E(1_A 1_{\{N(]s,u[)=0\}} \lambda_u) = E(1_A 1_{\{N(]s,u])=0\}} \lambda_u) = E(1_A 1_{\{N(]s,u])=0\}} \bar{\lambda}_u),$$

where $\bar{\lambda}_u$ is $\mathcal{F}_s-$measurable. Therefore, we have just shown that

$$P(N_t - N_s = 0 | \mathcal{F}_s) = 1 - \int_s^t P(N_u - N_s = 0 | \mathcal{F}_s) \bar{\lambda}_u du.$$

Iterating this equation yields the result. □

Remark 2.14. *Equation* (2.10) *gives an other way of understanding a Hawkes process. Suppose we have already constructed the process up to time* s. *Write* $T_s := \inf\{t > s : N_t \neq N_s\}$ *for the first jump time of the process after time* s. *We thus have to simulate this random variable in order to continue the construction of* N *on* $]s, +\infty[$. (2.10) *means :*

$$P(T_s > t | \mathcal{F}_s) = e^{-\int_s^t \bar{\lambda}_u du}.$$

We first notice that $\bar{\lambda}_u \geq \mu$ *which implies that* $\int_s^\infty \bar{\lambda}_u du = \infty$ *and therefore*

$$P(T_s < \infty | \mathcal{F}_s) = 1.$$

Taking the derivative with respect to t *in the above expression and changing sign, we obtain the probability density of* T_s. *In fact,*

$$\mathcal{L}(T_s | \mathcal{F}_s)(dt) = \bar{\lambda}_t e^{-\int_s^t \bar{\lambda}_u du} 1_{\{t \geq s\}} dt.$$

Of course, the above defines a proper probability density, once we know the past before time s.

2.6. **On stationarity.** Here the idea is to construct a version of the Hawkes process which behaves in the same way all over its domain of definition.

Definition 2.15. *Let* $N = (N_t, t \geq 0)$ *be a point process on* $\mathbb{R}_+$. *We identify* N_t *and* $N(]0,t])$. N *is called stationary if for all* $s, t \geq 0$,

$$N_t = N(]0,t]) \stackrel{\mathcal{L}}{=} N(]s, s+t]) = N_{s+t} - N_s.$$

Remark 2.16. *People working in stochastic processes would say that such a process has stationary increments. But since we are thinking in terms of point processes or point measures, the above really says that the way the process distributes points over space is really homogeneous. Thus the process itself is stationary.*

How is it possible to construct a stationary version of the Hawkes process? The idea is simple : If the process started to evolve at time $t = -\infty$, then – since it has evolved over an infinite time interval – it must be in stationary regime (if it has not exploded so far). This is an old idea which is classical in the analysis of long-memory processes. To construct such a process starting from $-\infty$ one uses the same Picard iteration approach as before. We work on $\mathbb{R}$ instead of $\mathbb{R}_+$.

Generation 0 Put $\lambda_t^{(0)} = \mu$ and let $N^{(0)}$ be a point process on $\mathbb{R}$ having this intensity. Remember that this simply means that for any $C \in \mathcal{B}(\mathbb{R})$,

$$N^{(0)}(C) = M(\{(s,x) \in \mathbb{R} \times \mathbb{R}_+ : s \in C, 0 \leq x \leq \lambda_s^{(0)}\}).$$

Recursion Put

$$\lambda_t^{(n+1)} = \mu + \int_{]-\infty, t[} h(t-s) dN_s^{(n)}$$

(notice that the only difference with the proof of Theorem 2.6 is the fact that integration starts from $-\infty$ and not from 0) and let

$$N^{(n+1)}(C) = M(\{(s,x) \in \mathbb{R} \times \mathbb{R}_+ : s \in C, 0 \le x \le \lambda_s^{(n+1)}\}).$$

Then, passing to expectations, we obtain the following recursion, using Corollary 2.12,

$$E\lambda_t^{(n)} = \mu + \int_{-\infty}^t h(t-s) E\lambda_s^{(n-1)} ds.$$

Using that $\lambda_t^{(0)} \equiv \mu$ does not depend on t this shows recursively that for all $n \ge 1$, $E\lambda_t^{(n)} = E\lambda_0^{(n)} = \mu + \|h\|_1 E\lambda_0^{(n-1)}$, and thus

$$E\lambda_0^{(n)} = \mu(1 + \|h\|_1 + \ldots + \|h\|_1^n).$$

Then, as before, by monotone convergence, one obtains the existence of the limit point process N having intensity

$$\lambda_t = \mu + \int_{]-\infty, t[} h(t-s) dN_s$$

where

$$E\lambda_t = \frac{\mu}{1 - \|h\|_1}.$$

We have just proven the following theorem.

Theorem 2.17. *Under the sub criticality assumption $\|h\|_1 < 1$, there exists a unique stationary version of the Hawkes process with average intensity given by*

$$E\lambda_t = \frac{\mu}{1 - \|h\|_1}.$$

Remark 2.18. *The above construction shows even more : the process $\lambda_t = \mu + \int_{]-\infty,t[} h(t-s) dN_s$ is actually also stationary. Recall that in case $h(t) = \alpha e^{-\beta t}$, for some $\alpha < \beta$, λ is a Markov process with values in $\mathbb{R}_+$, having generator*

$$Ag(x) = -\beta(x-\mu) g'(x) + x[g(x+\alpha) - g(x)].$$

We thus have shown that this Markov process possesses a stationary distribution (an invariant probability measure). Of course, this can also be shown using standard theory of Markov processes.

Remark 2.19. *Prove that for the non-stationary version of the Hawkes process, under the sub criticality assumption $\|h\|_1 < 1$, we have the uniform bound*

$$E\lambda_t \le \frac{\mu}{1 - \|h\|_1}, \tag{2.11}$$

for all $t \ge 0$.

2.7. **Limit theorems.** Let us come back to the non-stationary Hawkes process starting to live at time 0.

In the subcritical regime, several limit theorems can easily be stated and proven. First of all, as for the classical Poisson process, we have a law of large numbers.

Theorem 2.20 (Law of large numbers)**.** *Suppose that $\|h\|_1 < \infty$. The we have that*

$$\frac{N_t}{t} \to \frac{\mu}{1 - \|h\|_1}$$

almost surely, as $t \to \infty$.

Sketch of proof. We use the cluster representation of the linear Hawkes process. Let $I = (I_t, t \geq 0)$ be the process of "immigrants" and let $C^{(n)}, n \geq 1$ be the sequence of independent clusters of point processes that are endogenously triggered by each such immigrant. Then we can write that

$$N_t = I_t + \sum_{n=1}^{I_t} C^{(n)}_{t - T_n^{(0)}}.$$

Step 1. As $t \to \infty$, $C^{(n)}_{t-T_n^{(0)}} \to C^{(n)}_\infty$, which is the total size of the cluster created by an immigrant at time $T_n^{(0)}$. Therefore (here, we are very sloppy...),

$$N_t \sim I_t + \sum_{n=1}^{I_t} C^{(n)}_\infty.$$

Step 2. Now we use that all these $C^{(n)}_\infty$ are i.i.d., having expectation $\frac{\|h\|_1}{1-\|h\|_1}$. Hence,

$$\frac{N_t}{I_t} \to 1 + \frac{\|h\|_1}{1 - \|h\|_1} = \frac{1}{1 - \|h\|_1},$$

almost surely, as $t \to \infty$. (Here, we have applied the usual SLLN, with the subsequence $I_t \to \infty$.)

Step 3. We finally use that the classical SLLN for Poisson processes implies that $I_t/t \to \mu$ almost surely as $t \to \infty$ (which is also the argument that we were missing in the previous step that ensures that indeed $I_t \to \infty$). $\square$

Remark 2.21. *The above result can be extended to*

$$\sup_{0 \leq s \leq 1} \left(\frac{N_{st}}{t} - \frac{s\mu}{1 - \|h\|_1} \right) \to 0$$

almost surely, as $t \to \infty$. For a proof and for a rigorous proof of Theorem 2.20, see [4].

We also have an associated Central Limit Theorem, under a slightly more restrictive condition on the function h.

Theorem 2.22 (Central Limit Theorem)**.** *Suppose that $\|h\|_1 < \infty$ and moreover that $\int_0^t h(t)\sqrt{t}dt < \infty$. Then*

$$\sqrt{t} \left(\frac{N_t}{t} - \frac{\mu}{1 - \|h\|} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

as $t \to \infty$, where $\sigma^2 = \frac{\mu}{(1-\|h\|_1)^3}$.

Remark 2.23. *In [4], a functional version of the above theorem is proved. It holds in fact that*

$$\left(\sqrt{t}\left(\frac{N_{st}}{t}-\frac{s\mu}{1-\|h\|}\right), 0\le s\le 1\right)\xrightarrow{\mathcal{L}}(\sigma B_s, 0\le s\le 1).$$

This shows that at large scales, a Hawkes process (if it is centered) behaves as a diffusion.

A sketch of a proof of Theorem 2.22 will be given in the next section, in a multivariate setting.

2.8. **Application : Estimation of $\|h\|_1-$ the degree of endogeneity of the market.** Suppose we know the rate of arrivals of exogenous market order events μ. Then we may use the above results to estimate $m := \|h\|_1$. We simply take

$$\hat{m}_t := 1-\mu\frac{t}{N_t}$$

and we know, by Theorem 2.20, that $\hat{m}_t \to m$ almost surely, as $t\to\infty$. Theoretically, Theorem 2.22 enables us also to provide confidence bands, but the associated variance depends on the unknown parameter, so we have to replace it with an estimator of it. And, due to continuity reasons, we have to exclude regions of the parameter space where m is too close to 1.

2.9. **Higher order moments.** In the following, N denotes the invariant Hawkes process defined on $\mathbb{R}$. In particular, we work under the condition of sub-criticality $\|h\|_1 < 1$. We introduce the second moment measure $M_2(dt, ds)$ on $\mathbb{R}\times\mathbb{R}$ by

$$I\!\!E\int\int g(t,s)N(dt)N(ds) =: \int\int g(t,s)M_2(dt,ds),$$

where g is any positive Borel measurable function. It is possible to show (but this is admitted here) that

$$M_2(dt,ds)=dt\sigma(du)\delta_{t+u}(ds)$$

where

$$\sigma(du)=\Lambda\delta_0(du)+\sigma^c(u)du,$$

with $\Lambda=\frac{\mu}{1-\|h\|_1}$ the expected value of the stationary intensity. Let us just briefly explain why a Dirac measure is necessarily appearing in the above formula. Indeed, we can write

$$\int\int g(t,s)N(dt)N(ds)=\sum_n\sum_m g(T_n,T_m)=\sum_n g(T_n,T_n)+\sum_n\sum_{m\neq n} g(T_n,T_m).$$

But

$$\sum_n g(T_n,T_n)=\int g(t,t)N(dt),$$

whose expectation is

$$\int g(t,t)I\!\!E(\lambda(t))dt=\Lambda\int g(t,t)dt.$$

So the sum over diagonal terms gives the Dirac measure, and the double sum over un-common jumps of N gives rise to a two dimensional Lebesgue density $\sigma^c(t-s)$.

Proposition 2.24. *In particular, we have that*

$$Cov\left(\int\varphi dN,\int\psi dN\right)=\int\int\varphi(t)\psi(t+s)\Gamma(ds)dt,$$

where the measure Γ is given by

$$\Gamma(du)=\sigma(du)-\Lambda^2du=\Lambda\delta_0+\sigma^c(u)du-\Lambda^2du.$$

The function $\gamma(u) := \sigma^c(u) - \Lambda^2$ is called the covariance density of the process.

Theorem 2.25. *We have that*

$$\gamma(\tau) = h(\tau)\Lambda + \int_{-\infty}^{\tau} h(\tau - u)\gamma(u)du.$$

Proof. First of all, by symmetry, $\gamma(\tau) = \gamma(-\tau)$. So in the following we only consider $\tau > 0$. The idea of the proof is the following : let $h > 0$ and introduce for any t, $\varphi_t(x) := \frac{1}{h} 1_{[t,t+h]}(x)$. Then, as $h \to 0$, for any $\gamma > 0$,

$$I\!E((\int \varphi_t dN)(\int \varphi_{t+\tau} dN)) \to \sigma^c(\tau).$$

So it is sufficient fo evaluate the expression on the lhs for sufficiently small values of h. Notice that $\int \varphi_t dN = (N_{t+h} - N_t)/h$ and $\int \varphi_{t+\tau} dN = (N_{t+\tau+h} - N_{t+\tau})/h$. We choose h sufficiently small such that $t + \tau > t + h$. Conditioning with respect to $\mathcal{F}_{\tau+t}$, we thus obtain

$$I\!E((\int \varphi_t dN)(\int \varphi_{t+\tau} dN)) = \frac{1}{h} I\!E((\int \varphi_t dN) \int_{\tau+t}^{\tau+t+h} \lambda(u)du),$$

which behaves, for $h \to 0$ as

$$\begin{aligned} I\!E((\int \varphi_t dN)\lambda(t+\tau)) &= \mu I\!E((\int \varphi_t dN)) + I\!E((\int \varphi_t dN) \int_{-\infty}^{(t+\ \tau)-} h(t+\tau-s)dN_s) \\ &\sim \mu\Lambda + I\!E((\int \varphi_t dN) \int_{-\infty}^{(t+\ \tau)-} h(t+\tau-s)dN_s). \end{aligned}$$

We have to control this last double integral. We can write

$$\begin{aligned} (\int \varphi_t dN) \int_{-\infty}^{(t+\ \tau)-} h(t+\tau-s)dN_s &= \int \varphi_t(s)h(t+\tau-s)dN_s \\ &+ \int \int \varphi_t(s)h(t+\tau-u)1_{u \le t+\tau} 1_{u \ne s} dN_s dN_u. \end{aligned}$$

But, as $h \to 0$,

$$I\!E \int \varphi_t(s)h(t+\tau-s)dN_s \to \Lambda h(\tau).$$

Moreover, we rewrite the second term as

$$\begin{aligned} &\int \int \varphi_t(s)h(t+\tau-u)1_{u \le t+\tau} 1_{u \ne s} dN_s dN_u \sim \\ &\qquad \int \varphi_t(s)dN_s \int_{-\infty}^{(t+\tau)-} h(t+\tau-u)du \int \varphi_u(v)1_{v \ne s} dN_v \\ &\qquad = \int_{-\infty}^{t+\tau} h(t+\tau-u)du \int \int \varphi_t(s)\varphi_u(v)1_{v \ne s} dN_s dN_v. \end{aligned}$$

We take expectation and exploit the fact that

$$I\!E \int \int \varphi_t(s)\varphi_u(v)1_{v \ne s} dN_s dN_v \sim \sigma^c(u-t)$$

such that the above expression finally gives

$$\int_{-\infty}^{t+\tau} h(t+\tau-u)\sigma^c(u-t)du = \int_{-\infty}^{\tau} h(\tau-s)\sigma^c(s)ds.$$

As a consequence,

$$\sigma^c(\tau) = \mu\Lambda + \Lambda h(\tau) + \int_{-\infty}^{\tau} h(\tau - s)\sigma^c(s)ds.$$

Subtracting Λ^2 and observing that $\mu\Lambda - \Lambda^2 = -\Lambda^2 \int_{-\infty}^{\tau} h(t-s)ds$, we obtain the desired recursion formula. □

3. Multivariate Hawkes processes

3.1. **Definition.** For applications, it is more convenient to consider multivariate Hawkes processes $N^1, \ldots, N^d$, $d \geq 1$, where each jump process accounts for the events associated to a given asset. The ingredients are as before

– positive constants $\mu_1, \ldots, \mu_d \geq 0$ – the arrival rates of exogenous market events;
– integrable interaction functions $h_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$, for all $1 \leq i, j \leq d$.

Definition 3.1. *We say that the multivariate counting process $(N_t^1, \ldots, N_t^d)$ is a linear Hawkes process if*

1. *Almost surely, for all $i \neq j$, N^j and N^i do not jump at the same times.*
2. *For all i, $N_t^i - \int_0^t \lambda_s^i ds$ is a martingale, where*

$$\lambda_t^i = \mu_i + \int_{]0,t[} \sum_j h_{ij}(t-s)dN_s^j. \tag{3.12}$$

Remark 3.2. *As before, we can construct the process N^i as the thinning of a two-dimensional Poisson point process $M^i(ds, dz)$ where $M^1, \ldots, M^d$ are i.i.d.*

It is possible to rewrite (3.12) in the following way. Let

$$\lambda_t := \begin{pmatrix} \lambda_t^1 \\ \vdots \\ \lambda_t^d \end{pmatrix}, \; \mu := \begin{pmatrix} \mu^1 \\ \vdots \\ \mu^d \end{pmatrix}, \; N_t := \begin{pmatrix} N_t^1 \\ \vdots \\ N_t^d \end{pmatrix}.$$

Write moreover

$$h(t) := (h_{ij}(t))_{1 \leq i,j \leq d} \in \mathbb{R}_+^{d \times d}.$$

Then

$$\lambda_t = \mu + \int_{]0,t[} h(t-s)dN_s,$$

and

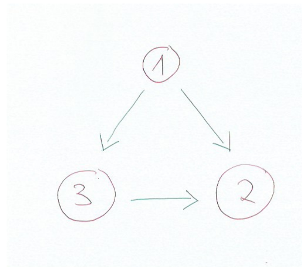
$$N_t - \int_0^t \lambda_s ds \; \in \mathbb{R}^d$$

is a martingale.

Example 3.3. *Let $d = 3$ and consider the following interaction graph. Thus, $h_{21} > 0, h_{12} = 0, h_{23} > 0, h_{32} = 0, \ldots$. (One has to read from the right to the left!) For example we can write*

$$\lambda_t^3 = \mu_3 + \sum_{T_n^3 < t} h_{33}(t - T_n^3) + \sum_{T_n^1 < t} h_{31}(t - T_n^1).$$

The first represents the exogenous events, the second one the self-interactions and the last one the interactions that N^3 feels due to events of type 1.



3.2. **First properties.** Define by matrix convolution the sequence of functions

$$h_n = h_{n-1} * h, \text{ that is, } (h_n)_{ij}(t) = \int_0^t \sum_k (h_{n-1})_{ik}(t-s) h_{kj}(s) ds,$$

for all $n \geq 1$. Moreover, define

$$\psi := \sum_{n \geq 1} h_n.$$

We will work under the usual sub criticality assumption which can be stated as follows. We put $K = \int_0^\infty h(t)dt$. Recall that this is a $d \times d-$matrix.

Assumption 3.4. *Let $\lambda_1, \ldots, \lambda_d$ be the (complex) eigenvalues of K. Then we put $\varrho(K) := \max |\lambda_i|$; $\varrho(K)$ is called the spectral radius of K. We suppose that $\varrho(K) < 1$.*

Notice that the above implies that $\|Kx\| \leq \varrho \|x\|$ for all $x \in \mathbb{R}^d$. In particular, $Id - K$ is invertible in this case, and ψ exists, that is, $\psi(t) < \infty$ for almost all t, since $\int_0^\infty \psi(t)dt = K + K^2 + \ldots = (Id - K)^{-1} - Id$.

Proposition 3.5. *Under Assumption 3.4, we have that*

$$EN_t = t\mu + \int_0^t \psi(t-s) s ds \mu. \tag{3.13}$$

Proof. Follows from the usual Fubini argument that implies that $m_t = EN_t$ satisfies

$$m_t = \mu t + \int_0^t h(t-s) m_s ds,$$

and the fact that the unique solution for this equation is given by the rhs of (3.13). □

It follows from the above considerations that the Strong Law of Large Numbers can be stated exactly as in the univariate case, except that one has to replace $\frac{1}{1 - \|h\|_1}$ by $(Id - K)^{-1}$. But the proof is essentially the same.

Theorem 3.6. *Under Assumption 3.4, we have that*

$$\frac{N_t}{t} \to (Id - K)^{-1} \mu$$

as $t \to \infty$, almost surely.

Let us now consider the martingale $M_t = N_t - \int_0^t \lambda_s ds$. We obtain

$$M_t = N_t - \mu t - \sum_j \sum_{n: T_n^j < t} \int_{T_n^j}^t h_{\cdot j}(s - T_n^j) ds = N_t - \mu t - K N_t + R_t = (Id - K) N_t - \mu t + R_t,$$

where $R_t^i = \sum_j \sum_{n:T_n^j<t} \int_t^\infty h_{ij}(s-T_n^j)ds + [K(N_t - N_{t-}]^i$ and where $1 = (1,\ldots,1) \in \mathbb{R}^d$.

It follows from this that

$$\frac{N_t}{\sqrt{t}} - \sqrt{t}(Id-K)^{-1}\mu \sim (Id-K)^{-1}\frac{M_t}{\sqrt{t}} + \frac{R_t}{\sqrt{t}}.$$

It is relatively easy to show that $R_t/\sqrt{t} \to 0$ in probability. Therefore, the weak convergence of $\frac{N_t}{\sqrt{t}} - \sqrt{t}(Id-K)^1\mu$ is implied by the one of $M_t/\sqrt{t}$.

To prove this last result, it is possible to use a classical martingale convergence theorem. Since M has uniformly bounded jumps and since different components of M do never jump together, it is sufficient to prove that the corresponding quadratic variations converge. But $[M^i]_t = N_t^i$, and by the SLLN, $N_t^i/t \to \sigma^i := [(Id-K)^{-1}\mu]^i$. So, writing Σ for the diagonal matrix having all diagonal elements equal to $\Sigma_{ii} = \sqrt{\sigma^i}$, we have that

$$M_t/\sqrt{t} \xrightarrow{\mathcal{L}} \Sigma B_1,$$

where $B = (B^1,\ldots,B^N)$ is Brownian motion, and thus

Theorem 3.7. *Suppose that $\int_0^\infty \sqrt{t}h(t)dt$ converges.*

$$\sqrt{t}\left(\frac{N_t}{t} - (Id-K)^{-1}\mu\right) \xrightarrow{\mathcal{L}} (Id-K)^{-1}\Sigma B_1,$$

as $t \to \infty$.

Proof of Theorem 3.7. It remains to prove that $R_t/\sqrt{t} \to 0$ in probability (we will show that this convergence actually holds in L^1). It is clear that the term $K(N_t - N_t)$ does not pose any problem. For the remaining terms and for the simplicity of notation, let us do the proof in the uni-variate case. We have $R_t = \sum_{T_n<t} g(t-T_n) = \int_{[0,t[} g(t-s)dN_s$, where $g(t) = \int_t^\infty h(s)ds$. Therefore, applying Corollary 2.12,

$$ER_t = \int_0^t E(\lambda_s)g(t-s)ds \le \sup_{s\le t} E(\lambda_s)\int_0^t g(s)ds.$$

Now, use that, by sub criticality, $\sup_{s\le t} E(\lambda_s) \le \frac{\mu}{1-\|h\|_1}$ (uniformly in t, see (2.11)), whence

$$\frac{1}{\sqrt{t}}ER_t \le \frac{\mu}{1-\|h\|_1}\frac{\int_0^t g(s)ds}{\sqrt{t}}.$$

We use de l'Hopital's rule and obtain

$$\lim_{t\to\infty}\frac{\int_0^t g(s)ds}{\sqrt{t}} = \lim_{t\to\infty}\frac{g(t)}{\frac{1}{2}t^{-1/2}} \le 2\int_t^\infty s^{1/2}h(s)ds \to 0$$

as $t \to \infty$.

□

Remark 3.8. *In [4], a functional version of the above result is proved: It actually holds that*

$$\left(\sqrt{t}\left(\frac{N_{tv}}{t} - v(Id-K)^1\mu\right), 0 \le v \le 1\right) \xrightarrow{\mathcal{L}} \left((Id-K)^{-1}\Sigma B_v, 0 \le v \le 1\right).$$

3.3. **Covariation across scales.** Within this section we consider a multivariate Hawkes process in stationary regime (that is, started from $t=-\infty$, under our sub criticality condition Assumption 3.4). Recall firstly that

$$E(N_t - N_s) = E\int_s^t \lambda_u du = (t-s)\boldsymbol{\lambda},$$

where

$$\boldsymbol{\lambda} = E\lambda_t = (Id - K)^{-1}\mu.$$

We shall also write

$$\boldsymbol{\lambda} = \frac{E(dN_t)}{dt}.$$

The *autocovariance density* is defined, for a time lag $\tau > 0$, to be

$$R(\tau) = Cov(\frac{dN_{t+\tau}}{d\tau}, \frac{dN_t}{dt}).$$

The above expression might look a bit mysterious. What it just means is that

$$R_{ij}(\tau) = \lim_{h\to 0} \frac{P(\Delta N^i_{t+\tau+h} = 1, \Delta N^j_{t+h} = 1) - P(\Delta N^i_{t+\tau+h} = 1)P(\Delta N^j_{t+h} = 1)}{h^2}. \tag{3.14}$$

By symmetry, we have that $R(\tau) = R(-\tau)$ for all $\tau > 0$. However, for $\tau = 0$, the limit in (3.14) does not exist for $i = j$, since

$$P(\Delta N^i_{t+h} = 1) \sim hE\lambda^i = h\boldsymbol{\lambda}^i, \text{ as } h \to 0.$$

Therefore, the complete covariance density is defined by

$$R^c(\tau) = R(\tau) + \boldsymbol{\lambda}\delta_0(\tau).$$

Theorem 3.9. *We have that*

$$R(\tau) = h(\tau)D + \int_{-\infty}^{\tau} h(\tau - u)R(u)du,$$

where D is the $d \times d-$diagonal matrix having entries $D_{ii} = \boldsymbol{\lambda}^i$.

Sketch of the proof of Theorem 3.9. Since $P(\Delta^i_t = 1|\mathcal{F}_t) = \lambda^i_t h + o(h)$, we can write that

$$\frac{1}{h^2}P(\Delta N^i_{t+\tau+h} = 1, \Delta N^j_{t+h} = 1) \sim \frac{1}{h}E(\lambda^i_{t+\tau}1_{\{\Delta N^j_{t+h}=1\}}) \sim E\left(\left[\mu^i + \int_{-\infty}^{t+\tau} h_{i\cdot}(t+\tau-s)dN_s\right]\frac{dN^j_t}{dt}\right).$$

On the other hand, $\boldsymbol{\lambda}$ satisfies the equation

$$\boldsymbol{\lambda} = \mu + \int_{-\infty}^{t+\tau} h(t+\tau-s)\boldsymbol{\lambda} ds.$$

Since

$$\lim_{h\to 0}\frac{P(\Delta N^i_{t+\tau+h} = 1)P(\Delta N^j_{t+h} = 1)}{h^2} = \boldsymbol{\lambda}^i\boldsymbol{\lambda}^j,$$

this gives

$$R_{i,j}(\tau) = E\left(\left[\int_{-\infty}^{t+\tau} h_{i\cdot}(t+\tau-s)d\bar{N}_s\right]\frac{d\bar{N}^j_t}{dt}\right),$$

where $\bar{N}_t = N_t - EN_t$. Now we use Fubini and obtain

$$R_{ij}(\tau) = \int_{-\infty}^{t+\tau} h_{i\cdot}(t+\tau-s)E\left(\frac{d\bar{N}_s}{ds}\frac{d\bar{N}^j_t}{dt}\right)ds.$$

But

$$E\left(\frac{d\bar{N}_s}{ds}\frac{d\bar{N}_t}{dt}\right) = R^c(s-t) = diag(\boldsymbol{\lambda})\delta(0) + R(s-t).$$

As a consequence, we obtain that

$$R_{ij}(\tau) = h_{ij}(\tau)\boldsymbol{\lambda}^j + \int_{-\infty}^{t+\tau}[h(t+\tau-s)R(s-t)]_{ij}ds = h_{ij}(\tau)\boldsymbol{\lambda}^j + \int_{-\infty}^{\tau}[h(\tau-u)R(u)]_{ij}du,$$

which is the assertion. □

4. Application to financial statistics

In this section we follow closely the paper by Bacry, Delattre, Hoffmann and Muzy [5].

We introduce a univariate price process given by

$$P_t = N_t^1 - N_t^2, \tag{4.15}$$

where N_t is a multivariate Hawkes process in dimension $d = 2$ and where

$$\lambda_t^1 = \mu + \int_{]0,t[} h(t-s)dN_s^2, \ \lambda_t^2 = \mu + \int_{]0,t[} \varphi(t-s)dN_s^1.$$

Thus the two processes are mutually exciting, and the associated interaction matrix is given by

$$h(t) = \begin{pmatrix} 0 & \varphi(t) \\ \varphi(t) & 0 \end{pmatrix}.$$

The fact that the price process takes values in $\mathbb{Z}$, that is, takes values on a lattice, corresponds to the tick-grid in financial statistics accounting for the discreteness of the price formation at fine scales.

Microstructure noise is a stylized fact of high-frequency financial analysis that corresponds to the observation that the observed daily variance increases as the time step of observation tends to 0. This observed daily variance is defined as follows. We introduce a time step $\Delta > 0$ and put

$$\hat{C}(\Delta) := \frac{1}{T}\sum_{i=1}^{[T/\Delta]}(P_{i\Delta} - P_{(i-1)\Delta})^2.$$

This is the sum of squared increments of the price process over small time intervals of length Δ. The total observation interval is given by $[0, T]$. We renormalize by T to obtain an expression which should be of order 1 with respect to the total observation time.

We call

$$C(\Delta) := E\hat{C}(\Delta)$$

the *mean signature* plot of the price process.

The fact that $C(\Delta)$ increases as $\Delta \to 0$ can not (!) be explained with a standard diffusion model. But it can be explained by the fact that an upwards jump of the price will be more likely be followed by a downward jump and vice versa, which is included in our model due to the mutual excitation of the two processes.

Proposition 4.1. *Suppose that $\varphi(t) = \alpha e^{-\beta t}$, for $t \geq 0$ and for $0 \leq \alpha < \beta$. Then, if the process is in stationary regime,*

$$C(\Delta) = \Lambda\left(\kappa^2 + (1-\kappa^2)\frac{1-e^{-\gamma\Delta}}{\gamma\Delta}\right),$$

where $\Lambda = 2\mu/(1-\|\varphi\|)$, $\kappa = 1/(1+\|\varphi\|)$, $\gamma = \alpha + \beta$.

Proof. We want to apply Theorem 3.9. Firstly, by symmetry of N^1 and N^2, we certainly have that $N_t^1 \sim N_t^2$, thus they have same expectation and variance. We start computing the mean average jump intensity $\boldsymbol{\lambda}^1 = EN_t^1/t$ in stationary regime. Observe that $\boldsymbol{\lambda}^1 = \boldsymbol{\lambda}^2 =: \boldsymbol{\lambda}$. We have

$$\boldsymbol{\lambda}_t := E\lambda_t(t) = \mu + \alpha \int_0^t e^{-\beta(t-s)} \boldsymbol{\lambda}_s ds.$$

Since $\boldsymbol{\lambda}_t \to \boldsymbol{\lambda}$, as $t \to \infty$, we obtain

$$\boldsymbol{\lambda} = \mu + (\alpha/\beta)\boldsymbol{\lambda}$$

and thus

$$\boldsymbol{\lambda} = \frac{\mu\beta}{\beta - \alpha}.$$

By stationarity, we also have that

$$C(\Delta) = \frac{2}{\Delta}(E(N_\Delta^1)^2 - EN_\Delta^1 N_\Delta^2).$$

Then $E((N_t^1)^2) = VarN_t^1 + \boldsymbol{\lambda}^2 t^2$. But

$$VarN_t^1 = Cov(N_t^1, N_t^1) = \int_0^t \int_0^t Cov(\frac{dN_s^1}{ds}, \frac{dN_u^1}{du}) dsdu = \boldsymbol{\lambda} t + \int_0^t \int_0^t R_{11}(s-u) dsdu.$$

And in the same way, we obtain that

$$E(N_t^1 N_t^2) = \boldsymbol{\lambda}^2 t^2 + \int_0^t \int_0^t R_{12}(s-u) dsdu.$$

Thus,

$$C(\Delta) = 2\left(\boldsymbol{\lambda} + \frac{1}{\Delta} \int_0^\Delta \int_0^\Delta (R_{11} - R_{12})(s-u) dsdu\right).$$

Now we exploit the findings of Theorem 3.9. We have that

$$R_{11}(\Delta) = \int_{-\infty}^{\Delta} \varphi(\Delta - s) R_{12}(s) ds.$$

And moreover,

$$R_{12}(\Delta) = \boldsymbol{\lambda}\varphi(\Delta) + \int_{-\infty}^{\Delta} \varphi(\Delta - s) R_{11}(s) ds.$$

Consequently, $r(t) := R_{11}(t) - R_{12}(t)$ satisfies

$$r(t) = -\varphi(t)\boldsymbol{\lambda} - \int_{-\infty}^{t} \varphi(t-s) r_s ds = -\varphi(t)\boldsymbol{\lambda} - \int_0^t \varphi(t-s) r_s ds - \int_0^\infty \varphi(t+s) r_s ds,$$

since $r_s = r_{-s}$.

Writing $\hat{r}(\mu) := \int_0^\infty e^{-\mu t} r_t dt$, we obtain

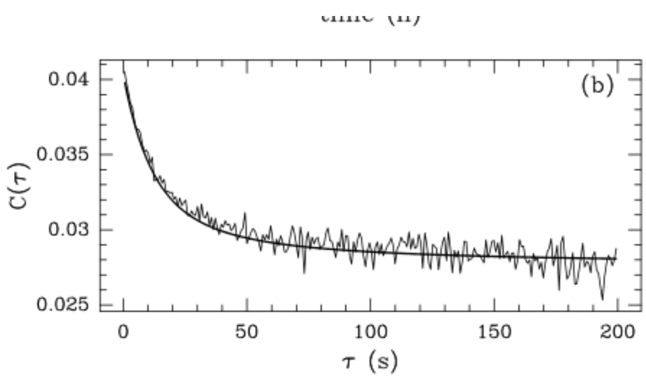
$$\hat{r}(\mu) = -\boldsymbol{\lambda}\hat{\varphi}(\mu) - \hat{r}(\mu)\hat{\varphi}(\mu) - \int_0^\infty e^{-\mu t} \int_0^\infty \varphi(t+s) r_s dsdt.$$

We calculate the last expression, taking into account that $\varphi(t) = \alpha e^{-\beta t}$.

$$\int_0^\infty e^{-\mu t} \int_0^\infty \varphi(t+s) r_s dsdt = \alpha \int_0^\infty e^{-\mu t} e^{-\beta t} \int_0^\infty e^{-\beta s} r_s dsdt = \frac{\alpha}{\mu + \beta} \hat{r}(\beta),$$

where we have used that $\varphi(\mu) = \frac{\alpha}{\mu+\beta}$. Therefore,

$$\hat{r}(\mu) = -\boldsymbol{\lambda} \frac{\alpha}{\mu + \beta} - \hat{r}(\mu) \frac{\alpha}{\mu + \beta} - \frac{\alpha}{\mu + \beta} \hat{r}(\beta). \tag{4.16}$$



Taking $\mu = \beta$, we obtain

$$\hat{r}(\beta) = -\frac{1}{2}\frac{\boldsymbol{\lambda}\alpha}{\alpha+\beta}.$$

Replacing in (4.16), this yields

$$\hat{r}(\mu) = \frac{\boldsymbol{\lambda}\alpha}{\mu+\beta+\alpha}[-1+\frac{\alpha}{2(\alpha+\beta)}]$$

which is the Laplace transform of

$$r(x) = ae^{-bx}, b = \alpha+\beta, a = -\boldsymbol{\lambda}\alpha\frac{2\beta+\alpha}{2(\alpha+\beta)}.$$

Finally, we obtain that

$$C(\Delta) = 2\left(\boldsymbol{\lambda} + 2\frac{a}{b} - 2\frac{a}{b^2}\frac{1-e^{-b\Delta}}{\Delta}\right)$$

which gives the result.

□

Remark 4.2. *Notice that, as* $\Delta \to 0$, $C(\Delta) \to 2\boldsymbol{\lambda}$ *and that* $C(\Delta)$ *increases as* Δ *decreases to* 0. *This can also be seen on the next picture which is taken from* [5].

References

[1] Abergel, F., Lu, X. Limit order book modelling with high dimensional Hawkes processes. hal-01512430, 2017.
[2] Bacry, E., Muzy, J.-F. Hawkes model for price and trades high-frequency dynamics. arxiv.org/pdf/1301.1135.
[3] Bacry, E., Mastromatteo, I., Muzy, J.-F. Hawkes processes in finance. arxiv.org/pdf/1502.04592
[4] Bacry, E., Delattre, S., Hoffmann, M. Muzy, J.-F. Scaling limits for Hawkes processes and application to financial statistics. arxiv.org/abs/1202.0842
[5] Bacry, E., Delattre, S., Hoffmann, M. Muzy, J.-F. Modeling microstructure noise with mutually exciting point processes. arxiv.org/abs/1101.3422
[6] Brémaud, P., Massoulié, L. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3) (1996) 1563-1588.

E. Löcherbach: SAMM, Université de Paris1 Panthéon Sorbonne
Email address: eva.locherbach@univ-paris1.fr