Hawkes Processes
From Definition To Modeling In Finance

Study carried out by the Data Science Practice
Special thanks to Pierre-Edouard THIERY
1 Introduction

Hawkes processes have been first used in seismology modeling by Ogata, to model epidemics of earthquakes’ aftershocks. But they are not limited to this field: they are a mathematical tool which turns out to be also relevant when it comes down to modeling in finance. First of all, it is important to understand the staple definitions of point processes and multivariate processes in order to grab the features of Hawkes processes [1]. This will be our goal in the first part of this note. Then, in a second part, we will present a simulation algorithm to generate times and events that are linked via a Hawkes model. In our third and fourth parts, we will focus much more on modeling: Hawkes processes are indeed a powerful tool to represent order book dynamics at high-frequency. This will lead us to quickly expose some important property concerning the random times $(T_n)$ is the following: the difference between two consecutive times $T_n - T_{n-1}$ is a random variable whose law is an exponential one with parameter $\lambda$.

We now extend the definition of a point process in order to take into account several kinds of events: this leads us to define the multivariate point process.

Definition 2 (Multivariate Point Process)
We consider a set $X = 1, 2, \ldots, d$ where $d$ denotes the number of different kinds of events we would like to follow. A multivariate point process is a sequence of random variables $(T_n, X_n) \in \mathbb{R} \times X,$ such that the sequence of the times $(T_n)$, called the ground process, is a point process. We can define the counting process as:

$$\forall \ C \in B(\mathbb{R} \times X), \ N(C) = \sum_{n \in \mathbb{N}} 1_{(T_n, X_n) \in C}$$

We naturally denote:

$$N^j(t) = \sum_{n \in \mathbb{N}} 1_{T_n \leq t} 1_{X_n = j}$$

$N_j^i \in \mathbb{R}$ counts the number of events which are of type $j$ and which occur before $t$.

To a point process or a multivariate process, we can associate a conditional intensity, denoted $\lambda$. In the theory of those processes, this intensity plays a significant role; we only mention here two main results about it. First, for any probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$, it is the only progressively measurable and locally integrable process which compensates the counting process, i.e. such that $N^j_t - \int_0^t \lambda^j_s ds$ is a $(\mathcal{F}_t)$ martingale under $\mathbb{P}$. But the converse is also true: for any progressively measurable and locally integrable process $(\lambda_t)$ there exists a unique probability $\mathbb{P}$ such that $\lambda$ is the compensator of the counting process $N$. $\lambda$ is a way to characterize the process $N$. For more details, the reader can see Jacod [2].

Here we only give the definition of the conditional intensity.

Definition 3 (Conditional Intensity)
We define the conditional intensity of a simple point process denoted $(N_t)$, for a given filtration $(\mathcal{F}_t)$, as

$$\lambda_t(\mathcal{F}_t) = \lim_{h \to 0} \mathbb{E}[N_{t+h} - N_t | \mathcal{F}_t] / h$$

2 Definitions: from Point Processes to Hawkes Processes

A point process is a mathematical tool that is used to represent random dates at which events may occur. As we will see, it is possible to extend the staple definition to enable us to follow various kinds of events; this will bring us to introduce multivariate Hawkes processes. In this note, we work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 1 (Simple Point Process on the Real Line)
A point process on the real line is a strictly increasing sequence of random times denoted $(T_n)$. Denoting $B(\mathbb{R})$ the Borel set of $\mathbb{R}$, we can easily define the counting process of such a point process as

$$\forall \ C \in B(\mathbb{R}), \ N(C) = \sum_{n \in \mathbb{N}} 1_{T_n \in C}$$

Very often we consider $C = [\infty, t]$ or $C = [0, t]$ and we denote $N(C) = N_t$. So $N$ defines a process that counts the number of events that occur during a given time period. It is very common to refer to a point process by its counting process, and we will do so in the rest of the note.

A simple example which can prove to be helpful to understand the above notions is the famous Poisson process. A Poisson process is a point process $(T_n)_{n \in \mathbb{N}}$, with its counting process denoted $(N_t)_{t \geq 0} \in \mathbb{N}$, which is defined by two properties:

- for $t_0 < t_1 < \cdots < t_n \in \mathbb{R}^+$, the variables $(N_{t_j} - N_{t_{j-1}})_{1 \leq j \leq n}$ are independent;
- for $0 \leq s < t$, the law of $N_{t-s}$ only depends on the difference $t-s$, so $N_t - N_s$ and $N_{t-s}$ have the same law.

Such a process is called a Poisson process since we can find $\lambda > 0$ such that, for $0 \leq s < t$, $N_t - N_s$ is a Poisson variable with parameter $\lambda(t-s)$.

We consider a set $X = 1, 2, \ldots, d$ where $d$ denotes the number of different kinds of events we would like to follow. A multivariate point process is a sequence of random variables $(T_n, X_n) \in \mathbb{R} \times X,$ such that the sequence of the times $(T_n)$, called the ground process, is a point process. We can define the counting process as:
We can make a few remarks. First the numerator $E[N_{t+h} - N_t]$ counts the average number of jumps over a period of size $h$; we divide by $h$ to normalize. The higher the intensity, the higher the probability to observe jumps.

Then the filtration which is usually considered for the conditional intensity is the natural filtration of the process, $\mathcal{F}_t = \sigma(N_s, s \leq t)$. In this case, the conditional intensity is merely denoted $\lambda_t$.

The definition 3 can easily be extended for a multivariate process. In this case, the conditional intensity at $t$ becomes a vector of size $d$; we measure the intensity for all kinds of events. So we have

$$\lambda_t = (\lambda^1_t, \ldots, \lambda^d_t)^T$$

**Definition 4 (Multivariate Hawkes Process)**

We consider a dimension $d$, a vector $\mu \in \mathbb{R}^d_+$, and a matrix $d \times d$ of functions $(\phi^{ij}(\bullet))_{i,j \leq d}$ where $\phi^{ij} : \mathbb{R}^+ \to \mathbb{R}^+$ for every $i$ and $j$. A multivariate Hawkes process is a multivariate process where, $\forall i \in 1, \ldots, d$:

$$\lambda^i_t = \mu^i + \sum_{n : T_n \leq t} \sum_{j=1}^d 1_{\lambda_n < a} \phi^{ij}(t - T_n)$$

$\mu \in \mathbb{R}^d_+$ is called the vector of exogenous intensities, and the functions $\phi^{ij} : \mathbb{R}^+ \to \mathbb{R}^+$ the kernel functions.

A common assumption is that the kernels $\phi^{ij}$ are decreasing functions on a compact interval $[0, K^{ij}]$ with $K^{ij} > 0$. For example, if we work with decreasing exponential functions, we assume that this is the case, due to the rapidly decreasing slope of the exponential function. For the rest of this paper, we work under this assumption.

Figure 1 shows what the conditional intensity of a Hawkes process looks like in a simple unidimensional case. When an event occur, there is a jump in the intensity, which then decreases. The jump is the intensity means that it is more likely to observe new events when an event has just occurred: that is why Hawkes processes are sometimes called self-exciting processes.

![Figure 1: Conditional intensity for a 1-D Hawkes process](image)

### 3 Simulation

The existence of at least one simulation algorithm for a process amounts to a proof that such a process exists. We provide here the main simulation algorithm for Hawkes processes, known as the thinning algorithm. Of course, simulation is not only important to assert a process exists, but it is also pivotal to evaluate the features of a model and to check the quality of an estimation technique.

We develop here the thinning algorithm [3], but the reader should know that this is not the only technique which exists. There also exists the inverse method simulation algorithm, which is mainly based on the idea that $X = F(Y)$, where $Y$ is a random variable with a given law and $F$ its distribution function, follows a uniform law on the unit interval. This result can be adapted for point processes: for a point process $N$, it is possible to show, under the right assumptions, that $N_{t+1}$, with $\Lambda_t = \int_0^t \lambda_s \, ds$, is a Poisson process with unit rate. This inverse method is nonetheless never used due to the computational effort needed.

**Algorithm (Thinning Algorithm For Hawkes Processes)**

We assume we would like to simulate a Hawkes process with kernel functions denoted $\phi^{ij}$ for $i, j \in 1, \ldots, d$, where $d$ is the dimension of the output process. $N_{\text{events}}$ denotes the maximum number of events we would like to generate; $T_{\text{max}}$ the upper bound for the $T_n$ variables. The simulation algorithm is based on two loops, an outer loop on $n$, which contains an inner loop on $r$.

- **Initialization of the outer loop ($n = 0$):**
  $$T_0 = 0, \forall j \in 1, \ldots, d, \lambda^j_0 = \mu^j$$

- **Then we iterate on $n$: for $n \geq 1$:**
  - **Initialization of the inner loop ($r = 0$)**
    $$T^0_n = T_{n-1}, \quad M^0_n = \sum_{i=1}^d \lambda^i_{T^0_{n-1}}$$

  - **Then we iterate on $r \geq 1$: we sample $E^r_n \equiv \text{Exp}(1)$, and we define**
    $$\tau^r_n = \tau^{r-1}_n + \frac{E^r_n}{M^r_n - 1}$$

    $$M^r_n = \sum_{i=1}^d \lambda^i_{\tau^r_n}$$

  - For each $r$, we also sample a uniform variable $U^r_n \equiv U[0, 1]$; it gives a criterion to stop the inner loop. When $U^r_n \geq \frac{M^r_n}{M^{r-1}_n}$, we continue the iteration. Otherwise we stop the inner loop on $r$ and we set:
    $$T_n = \tau^r_n$$
To get the type of the event occurring at date \( T_0 \), we use the density defined as:

\[
  f_m(i, t) = \frac{\lambda_i}{\sum_{j=1}^{d} \lambda_j}
\]

These weights define a discrete random variable; we only have to sample a value following this law, using \( t = T_0 \). The conditional intensity is updated by applying the formula of the Hawkes process:

\[
  \lambda_{T_i} = \mu^i + \sum_{k=1}^{n} \phi(t-T_k)
\]

We stop the outer loop when \( n > N_{\text{event}} \) or \( T_n > T_{\text{max}} \).

To illustrate this algorithm, and to show that, by construction, the output process is a Hawkes process, we develop a simple example: a Hawkes process with dimension \( d = 1 \), kernel function \( \phi(t) = ae^{-bt} \), and exogeneous intensity \( \mu \).

Step 1: we start from \( T_0 = 0 \) and \( \lambda_0 = \mu \). By applying the desired formula for the conditional intensity,

\[
  \lambda_t = \mu + \sum_{n : T_n \leq t} \phi(t-T_n)
\]

We know that, until the first event occurs, \( \lambda_t = \mu \).

Step 2: to build \( T_1 \), we have to iterate on the inner loop on \( r \). We start with \( \tau_0 = T_0 = 0 \), and \( M_{\tau_0} = \lambda_{\tau_0} = \mu \). We sample \( E_1 \), following an exponential law. \( \tau_1 \) is equal to \( \tau_0 + E_1/M_0 \), so:

\[
  \tau_1 = \frac{E_1}{\mu}
\]

Step 3: we check whether or not we keep \( \tau_1 \) to define \( T_1 \), otherwise we have to continue the iteration on \( r \). We sample \( U_1 \equiv U[0,1] \). We check the condition (obvious in this case):

\[
  U_1 \leq \frac{M_1}{M_0} \iff U_1 \leq 1
\]

We stop the iteration on \( r \) after only one iteration and we can define:

\[
  T_1 = \tau_1
  \lambda_{T_1} = \mu + \phi(0) = \mu + a
\]

It is interesting to go a few steps further and to consider the construction of \( T_2 \).

Step 4: by applying the conditional intensity formula, we know that, until a second event occurs, it is worth, for \( t \) after \( T_1 \):

\[
  \lambda_t = \mu + ae^{-bt}
\]

We set \( \tau_2 = T_1 \) and \( M_{\tau_2} = \mu + a \).

Step 5: we iterate on \( r \). We sample \( E_2 \) and we define \( \tau_2 = \tau_2 + E_2/M_2 \), so:

\[
  \tau_2 = T_1 + \frac{E_2}{\mu + a}
  M_2 = \lambda_{\tau_2} = \mu + ae^{-b(T_1 + E_2/\mu + a)}
\]

Step 6: now we check the condition to know whether we keep \( \tau_2 \) to define \( T_2 \). We sample \( U_2 \):

\[
  U_2 \leq \frac{M_2}{M_0} \iff U_2 \leq \frac{\mu + ae^{-b(T_1 + E_2/\mu + a)}}{\mu + a}
\]

If this condition is verified, we define \( T_2 = \tau_2 \), otherwise we keep iterating on \( r \) until the above condition is fulfilled.

### 4 Modeling in finance

The theory of point processes can be used to model order book dynamics. We present here such a framework.

As it is fairly difficult to represent the entire order book, we can dwell on certain quantities that are particularly interesting. We can for example consider a 4-dimension process:

\[
  P_t = \begin{pmatrix} T_{t^-}, T_{t^+}, N_t, N_t^- \end{pmatrix}^T
\]

All the elements in this vector are counting processes. \( T_{t^-} \) and \( T_{t^+} \) account for the number of anonymous market orders that arrive before \( t \) either a the best ask (\( T_{t^+} \)) or at the best bid (\( T_{t^-} \)). \( N_t \) and \( N_t^- \) account for the number of upward and downward price jumps before \( t \).

So far, the conditional intensity of the counting process denoted \( \lambda_t \) has not been specified. We assume that \( P_t \) is a Hawkes process with dimension 4. If we denote \( \Phi \) the matrix of the kernel functions, we know that, in \( \mathbb{R}^4 \):

\[
  \lambda_t = \mu + \int_{-\infty}^{t} \Phi(t-s) dP_s
\]

\( \Phi \) represents the influence of each component of \( P \) on each component of \( \lambda \). The kernel matrix is decomposed into four 2x2 symmetrical submatrices:

\[
  \Phi(t) = \begin{pmatrix} \Phi_F(t) & \Phi_F(t) \\ \Phi_F(t) & \Phi_F(t) \end{pmatrix}
  \quad \Phi_s(t) = \begin{pmatrix} \phi^{s}(t) & \phi^{s}(t) \\ \phi^{s}(t) & \phi^{s}(t) \end{pmatrix}
\]
Each submatrix represents the influence of past trades/jumps in prices on the intensity of the future trades/jumps in prices. If we dive into the details:

- $\Phi^T$ accounts for the influence of trades on the intensity of trades: $\phi^{T,s}$ accounts for the influence of buying (or selling) market orders on the intensity of future buying (selling) market orders, whereas $\phi^{T,c}$ accounts for the influence of buying (or selling) market orders on the intensity of future selling (buying) market orders;
- in the same way, $\Phi^N$ accounts for the influence of price jumps on the future intensity of price jumps.
- $\Phi^I$ accounts for the influence of trades on price jumps: $\phi^{I,s}$ represents this influence between buying (selling) market orders and upward (downward) price jumps, whereas $\phi^{I,c}$ represents the influence between buying (selling) market orders and downward (upward) price jumps;
- $\Phi^F$ accounts for the influence of price jumps on the intensity of trades.

For general culture, we can keep in mind that Hawkes processes have been first used in seismology modeling, by Ogata, to model epidemics of earthquakes’ aftershocks, which are highly clustered in time. In our problem, the relevance of Hawkes processes stems from that order arrivals are also highly clustered in time, according to empirical observations. That is why we would like to use this very general class of processes. Nonetheless we still have to verify that such a process is good at representing some of the stylized facts of the high-frequency dynamics of an order book.

5 A first insight into high-frequency stylized facts

In this part, we provide the reader with a first insight into high-frequency stylized facts; this will allow to mention some models that exist, their limits, and the advantage that we can retrieve from using a Hawkes model.

**Definition 5 (Cumulated Price)**
The cumulated price is defined by:

$$X_t = N_t^+ - N_t^-$$

The cumulated price accounts for the price of the considered asset, as it represents at each date $t$ the number of price jumps, either upward or downward, since $t = 0$.

This gives us immediately a first stylized fact of high-frequency dynamics modeling: the price evolves on a discrete grid due to the existence of a tick, i.e. a minimum upward or downward movement in the price. By construction $(X_t)_{t\geq 0}$ is a discrete process which evolves on $\mathbb{Z}$.

**Stylized fact 1 (Microstructure Price)**
The high-frequency price evolves on a discrete grid due to the existence of a tick, i.e. a minimum upward or downward movement in the price.

Nonetheless, if a model is good at representing prices at high-frequency, it should also be good to represent prices at low-frequency: who can do more can do less. So we should find an Ito process-like behavior at low-frequency. This is typically the kind of observation that is possible when drawing the cumulated price process of our model at low-frequency.

![Figure 2: Evolution of $X_t$ at high-frequency](image)

**Stylized fact 2 (High-Frequency Prices at Low-Frequency)**
A model for price at high-frequency has to be also good at representing the price at low-frequency.

Theoretically, Bacry and al [4] have derived a kind of central limit theorem for the counting process $(P_t)$ of our model:

$$\frac{1}{h} (P_{ht} - \mathbb{E}(P_{ht})) \overset{Law}{\rightarrow} C_0 W_t$$

when $h \to \infty$.

![Figure 3: Evolution of $X_t$ at low-frequency](image)
**Stylized fact 3** (Microstructure Noise And Signature Plot)

If we denote \((X_t)\) a price process at high-frequency, as we have done above for the cumulated price in case of a Hawkes process, the signature plot on the interval \([0, T]\) with step \(\tau\) \((T = N \times \tau)\) is defined as:

\[
V_T(\tau) = \frac{1}{T} \sum_{n=0}^{N-1} |X_{(n+1)\tau} - X_{n\tau}|^2
\]

It is merely the realized volatility over a time period \([0, T]\), using a step \(\tau\).

We can then draw the signature plot by considering the function \(\tau \rightarrow V_T(\tau)\). The stylized fact of microstructure noise is that the signature plot is a decreasing function of \(\tau\) which is stable when \(\tau\) is above one-five minute(s). At ultra high-frequency, we observe that \(V_T\) is finite, so:

\[
V_T(\tau) \xrightarrow{\tau \to 0} V_T(0) < \infty
\]

This stylized fact rules out the additive model for microstructure price; indeed the latter is sometimes used as a first approach to grab the features of microstructure price dynamics. It states that the price which is observed is noisy:

\[
Y_t = X_t + \epsilon_t,
\]

With \(t\) the observation times, and \(X\) the true price. The variables \(\epsilon_t\) are assumed to be independent noises with \(\mathbb{E}[\epsilon_t] = 0\), and \(\mathbb{E}[\epsilon^2_t] = \sigma^2\). So we see that:

\[
\mathbb{E}[V_T(\tau)] = \frac{1}{T} \sum_{n=0}^{N-1} \mathbb{E}[(X_{(n+1)\tau} + \epsilon_{(n+1)\tau} - X_{n\tau} - \epsilon_{n\tau})^2]
\]

\[
= \mathbb{E} \left[ \frac{1}{T} \sum_{n=0}^{N-1} (X_{(n+1)\tau} - X_{n\tau})^2 \right] + \frac{2N\sigma^2}{\tau} + \frac{\sigma^2}{\beta}
\]

Under the right assumptions, the quantity \(A\) converges when \(\tau \to 0\), but then \(B \to \infty\). This rules out the additive model to represent price dynamics at high-frequency.

Bacry and al [5] have derived the mean signature plot in the case of our model, whose form is of the kind:

\[
\mathbb{E}[V(\tau)] = \alpha + \beta \frac{1 - e^{-\gamma \tau}}{\gamma \tau}
\]

We refer to the above-mentioned article for the details and proof of such a result.

Last but not least, we mention, mainly for the reader’s interest, the Epps effect [6] [7], which is another important stylized fact of microstructure price.

**Stylized fact 4** (Epps Effect)

We consider two assets, denoted \(X^1\) and \(X^2\); the correlation between those two assets is defined by

\[
\rho_T(\tau) = \frac{C_{1,2,T}(\tau)}{\sqrt{C_{1,1,T}(\tau)C_{2,2,T}(\tau)}}
\]

with

\[
C_{1,2,T} = \frac{1}{T} \sum_{n=0}^{N-1} (X^1_{(n+1)\tau} - X^1_{n\tau})(X^2_{(n+1)\tau} - X^2_{n\tau})
\]

We called Epps effect the fact that \(\rho\) is an increasing function which vanishes when \(\tau \to 0\).

As for the signature plot, Bacry and al [5] have derived a closed formula for the correlation \(\rho_T(\tau)\) for the above-mentioned model of price. They showed that \(\rho_T(\tau) \to 0\) when \(\tau \to 0\), and that \(\rho_T(\tau) \to \rho_\infty\) when \(\tau \to \infty\). So, when modeling the prices with Hawkes processes, the empirical correlation between two assets is decreasing to 0 at ultra high frequency.

6 Conclusion

In this note, we have presented the main definitions that lead to the notion of Hawkes process, along with a simulation algorithm to generate such random times and events. Hawkes processes are very helpful in finance when it comes to modeling microstructure price and order book dynamics. A major advantage of such of model is that it respects the most important stylized facts of high-frequency: discrete price at high-frequency, which tends to an Ito-like diffusion when considered at low-frequency; a signature plot which is a decreasing function with finite value in 0; the Epps effect.

Nonetheless, this is only a first step. Indeed, if Hawkes processes are eligible processes to model data at high-frequency, it is then interesting to see how we can use them to get some powerful results. The market impact, for instance, is a quantity which can be estimated within the framework we have presented in our third part.
References


