# High-frequency data and limit order books

—

## Lecture 4 - March 18th, 2025

Ioane Muni Toke

MICS Laboratory and Chair of Quantitative Finance
CentraleSupélec – Université Paris-Saclay, France

CentraleSupélec

U. Paris-Saclay CentraleSupelec cursus Ingénieur 3A Mathématiques et Data Science
IP Paris ENSAE M2 Statistics, Finance and Actuarial Science
U. Paris-Saclay Evry M2 Quantative Finance

# Table of contents

# Table of contents

# Log-likelihood of a point process I

## Proposition (Log-likelihood of a point process)

*The log-likelihood of a counting process $(N_t)_{t\geq 0}$ with conditional intensity $(\lambda^*_t)_{t\geq 0}$ observed on $[0, T]$ is*

$$\log \mathcal{L}_T = \int_0^T \log \lambda^*_s \, dN_s - \int_0^T \lambda^*_s \, ds.$$

*Sketch of the proof*: See e.g., [8, 2]. We use successive conditioning (same notation as in the slides on conditional intensity in the previous lecture). Let $0 < t_1 < \ldots < t_n < T$ be an observation of the counting process $N$ on $[0, T]$. The likelihood is

$$\mathcal{L}_T = p_1(t_1)\, p_2(t_2|t_1) \, \ldots \, p_n(t_n|t_1, \ldots, t_{n-1})\, S_{n+1}(T, t_1, \ldots, t_n).$$

# Log-likelihood of a point process II

Given our construction of the conditional intensity, setting $t_0 = 0$,

$$\mathcal{L}_T = \prod_{i=1}^{n} \left[ \lambda^*(t_i) \exp\left( -\int_{t_{i-1}}^{t_i} \lambda^*(u)\, du \right) \right] \times \exp\left( -\int_{t_n}^{T} \lambda^*(u)\, du \right)$$

$\square$

- *Example:* In the case of a time-homogeneous Poisson process, we have for an observed sample $0 < t_1 < \ldots < t_n < T$:

$$\log \mathcal{L}_T = n \log \lambda - \lambda T.$$

The maximum-likelihood estimator of the intensity is thus (as expected) $\hat{\lambda}_{MLE} = \frac{n}{T}$.

# Table of contents

# One-dimensional Hawkes processes

## Definition (Linear self-exciting process)

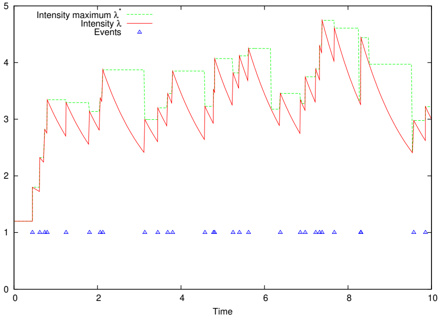
A counting process $N$ is called a (linear, one-dimensional) Hawkes process if it is a counting process with conditional intensity:

$$\lambda_t = \lambda_0(t) + \int_0^t \nu(t-s)dN_s = \lambda_0(t) + \sum_{0<t_i<t} \nu(t-t_i),$$

where $\lambda_0 : \mathbb{R} \mapsto \mathbb{R}_+$ and $\nu : \mathbb{R}_+ \mapsto \mathbb{R}_+$ are deterministic functions.

- $\nu$ expresses the positive influence of the past events $t_i$ on the current value of the intensity process.
- Hawkes processes are self-exciting processes.
- $\lambda_0$ is the baseline intensity ($\lambda \geq \lambda_0$). From now on, $\lambda_0 > 0$ constant.
- Original paper by [4]. Used in geology (earthquakes replicas), finance (see later in the course), biology (neuron excitation, multidimensional case, see below).

# Sample path, kernels




Sample path of a 1D-Hawkes process with a single exponential kernel ($P = 1$) and parameters $\lambda_0 = 1.2, \alpha_1 = 0.6, \beta_1 = 0.8$.

- Exponential kernel commonly used:

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \mathbf{1}_{\mathbb{R}_+}(t).$$

- Exponential kernels allow for efficient computations, especially for maximum-likelihood estimation (*see later*), but do not exhibit long-memory properties.

# Stability of a Hawkes process

If the process has reached a stationary regime, the average intensity value is: $\dfrac{\lambda_0}{1 - \int_0^\infty \nu(x)dx}$.

- The average number of points of a Hawkes process on a time interval of length $T$ is thus in a stationary regime $\dfrac{\lambda_0 T}{1 - \|\nu\|_1}$.
- If $\|\nu\|_1 < 1$, the Hawkes process is sub-critical / stationary. If $\|\nu\|_1 = 1$ (resp. $> 1$), the process is called critical (resp. supercritical) (*See discussion in finance later in the course*).
- In the case of exponential kernels, the stability condition is written $\sum_{j=1}^{P} \dfrac{\alpha_j}{\beta_j} < 1$.

# Thinning for stochastic intensities

- [6] extends the thinning method for non-homogeneous Poisson processes introduced by [5] to point processes with stochastic intensities.

## Theorem (Thinning)

*Consider point process $N$ with intensity $\lambda_t$ on an interval $(0, T]$. Suppose we can find a process $\lambda_t^*$ such that $\lambda_t \leq \lambda_t^*$ ($\mathbf{P}$-a.s.). Let $t_1^*, t_2^*, \ldots t_n^* \in (0, T]$ be the points of the process $N^*$ with intensity $\lambda^*$. For each of the points, attach a mark $p = 1$ with probability $\frac{\lambda(t_j^*)}{\lambda^*(t_j^*)}$, else $p = 0$. Then the points with marks $p = 1$ form a point process which is the same as $N$.*

- If $\lambda_t^*$ can be constructed pathwise as a piecewise constant process, we can thus simulate the point process $N$ with simple exponential random variables.

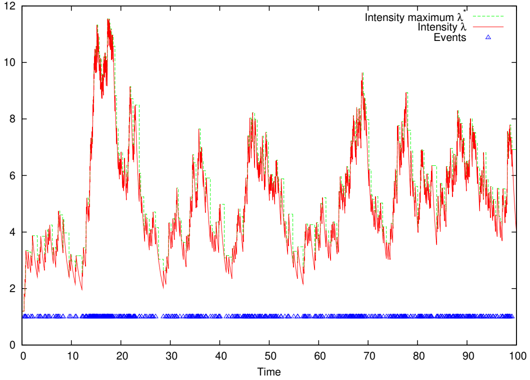
# Simulation algorithm for a one-dimensional Hawkes process

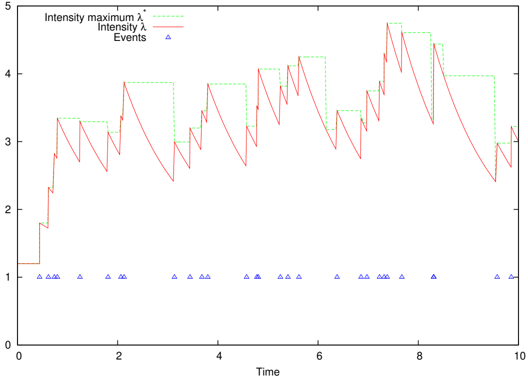
1. Initialization : Set $t \leftarrow 0$, $\lambda^* \leftarrow \lambda_0(0)$ and an empty list of events.
2. Main loop : **While** $t < T$:
   (a) New event candidate : Generate $U \rightsquigarrow \mathcal{U}_{[0,1]}$ and set $t \leftarrow t - \frac{1}{\lambda^*} \log U$.
   **If** $t \geq T$, **Then** return the list of events.
   (b) Thinning : Generate $D \rightsquigarrow \mathcal{U}_{[0,1]}$.
   **If** $D \leq \frac{\lambda(t)}{\lambda^*}$,
   **Then** New event: add $t$ to the list of events and update the maximum intensity $\lambda^* \leftarrow \lambda(t) + \sum_{j=1}^{P} \alpha_j$ ;
   **Else** No new event: update the maximum intensity $\lambda^* \leftarrow \lambda(t)$.
3. Output: return the list of events.

It is crucial to observe that exponential kernels lead to an easy update of the jump part of the intensity, without having to parse all previous events.

# Path of a Hawkes process and its intensity

- Simulation of a one-dimensional Hawkes process with a single exponential kernel (parameters $P = 1, \lambda_0 = 1.2, \alpha_1 = 0.6, \beta_1 = 0.8$).



# Log-likelihood of a Hawkes process I

- In the case of a Hawkes process with $P$ exponential kernels, the log-likelihood of an observation $0 < t_1 < \ldots < t_n < T$ is:

$$\log \mathcal{L}_T = \sum_{i=1}^{n} \log \left[ \lambda_0(t_i) + \sum_{j=1}^{P} \sum_{k=1}^{i-1} \alpha_j e^{-\beta_j (t_i - t_k)} \right] - \Lambda(0, T).$$

- The double sum can be bypassed by a recursive computation ([6])

$$\log \mathcal{L}_T = \sum_{i=1}^{n} \ln \left[ \lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i) \right] - \Lambda(0, T),$$

where $R_j(1) = 0$ and $R_j(i) = e^{-\beta_j (t_i - t_{i-1})} (1 + R_j(i-1))$, for all $j = 1, \ldots, P$.

# Log-likelihood of a Hawkes process II

**Proposition (Log-likelihood of a Hawkes process, exponential kernels)**

*The log-likelihood of a Hawkes process* $(N_t)_{t\geq 0}$ *with* $P$ *exponential kernels observed on* $[0, T]$ *is*

$$\log \mathcal{L}_T = -\int_0^T \lambda_0(s)ds - \sum_{i=1}^{n}\sum_{j=1}^{P}\frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(T-t_i)}\right) + \sum_{i=1}^{n}\log\left[\lambda_0(t_i) + \sum_{j=1}^{P}\alpha_j R_j(i)\right].$$

- Maximum-likelihood estimator numerically available.
- Analytical gradients available for better optimization.
- In the case of non-exponential kernel, the absence of recursive evaluations may lead to prohibitive computational times.

# Maximum-likelihood estimation

- [7]: MLE estimator $\hat{\theta}_T = (\hat{\lambda}_0, \hat{\alpha}_j, \hat{\beta}_j))$ is
  - *consistent*, i.e. converges in probability to the true values $\theta^* = (\lambda_0, \alpha_j, \beta_j))$ as $T \to \infty$ :

$$\forall \epsilon > 0, \quad \lim_{T \to \infty} P[|\hat{\theta}_T - \theta^*| > \epsilon] = 0.$$

  - *asymptotically normal*, i.e.

$$\sqrt{T}\left(\hat{\theta}_T - \theta^*\right) \to \mathcal{N}(0, I^{-1}(\theta^*))$$

    where $I(\theta) = \left(\mathrm{E}\left[\frac{1}{\lambda}\frac{\partial \lambda}{\partial \theta_i}\frac{\partial \lambda}{\partial \theta_j}\right]\right)_{i,j}$.
  - *asymptotically efficient*, i.e. asymptotically reaches the lower bound of the variance of an estimate.

# Change of time

- The "change of time" theorem proved for the non-homogeneous Poisson process can be extended in the general case.

### Theorem (Change of time of a point process)

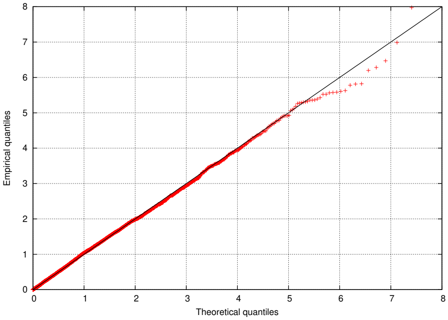
*Let $(\tilde{N}_t)_{t\geq 0}$ be a point process with intensity $(\lambda_t)_{t\geq 0}$. Define for each $t$ the stopping time $\tau(t)$ such that*

$$\int_0^{\tau(t)} \lambda_s \, ds = t$$

*Then $N_t = \tilde{N}_{\tau(t)}$ is a time-homogeneous Poisson process with intensity 1.*

- Consequence: If the $\tilde{t}_i$'s are the points of $\tilde{N}$, then the quantities $\int_{\tilde{t}_{i-1}}^{\tilde{t}_i} \lambda_s \, ds$ are exponentially distributed with parameter 1.

# Goodness-of-fit




- For a one-dimensional Hawkes process with exponential kernels:

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda_0(s)ds$$
$$+ \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \left(1 - e^{-\beta_j(t_i - t_{i-1})}\right) A_j(i-1),$$

where $A_j(1) = 1$ and
$A_j(i-1) = 1 + e^{-\beta_j(t_{i-1} - t_{i-2})} A_j(i-2)$.

- Quantile plots available as a *visual* indication of goodness-of-fit

# Table of contents

# Multidimensional Hawkes processes

## Definition (Multidimensional Hawkes processes)

A $M$-dimensional Hawkes process is a point process $\mathbf{N}_t = (N_t^1, \ldots, N_t^M)$ with intensities $(\lambda_t^m)_{t \geq 0}$, $m = 1, \ldots, M$ such that:

$$\lambda^m(t) = \lambda_0^m(t) + \sum_{n=1}^{M} \int_0^t g_{mn}(t-s) dN_s^n,$$

where $\lambda_0 : [0, \infty) \to \mathbb{R}_+$ and $g_{mn} : [0, \infty) \to \mathbb{R}_+$, $m, n = 1, \ldots, M$ are deterministic functions.

- We may write using a vector notation: $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 + \int_0^t \mathbf{G}(t-s) \cdot d\mathbf{N}_s$.
- In the case of exponential kernels, $\mathbf{G}(t) = \left(\alpha^{mn} e^{-\beta^{mn} t}\right)_{m,n=1,\ldots,M}$.

# Stability of multidimensional Hawkes processes

- If the process is in a stationary state, the average intensities $\boldsymbol{\mu}^t \in \mathbb{R}^M$should satisfy:

$$\boldsymbol{\mu} = \left(\mathbf{I} - \int_0^\infty \mathbf{G}(u)du\right)^{-1} \boldsymbol{\lambda}_0$$

- If the spectral radius of the matrix $\boldsymbol{\Gamma} = \int_0^\infty \mathbf{G}(u)du$ is strictly smaller than 1, then the $M$-dimensional Hawkes process is stable.
- In the exponential kernel case, $\boldsymbol{\Gamma} = \left(\frac{\alpha^{mn}}{\beta^{mn}.}\right)_{m,n=1,\dots,M}$

# Simulation of multidimensional Hawkes processes

- The algorithm is formally identical to one in one dimension.
- Let $I_k(t) = \sum_{n=1}^{k} \lambda^n(t)$, $k = 1, \ldots, M$.
- The total intensity is now $I_M(t)$, bounded above by $I_M^*$, and the next candidate event must not be generated with $I_M^*$.
- The thinning step is now:
  Generate $D \rightsquigarrow \mathcal{U}_{[0,1]}$.
  **If** $D \leq \frac{I_M(s)}{I_M^*}$,
  **Then** set $s$ as a point of the process $n_0$, where $n_0$ is such that $\frac{I_{n_0-1}(s)}{I_M^*} < D \leq \frac{I_{n_0}(s)}{I_M^*}$, and go through the main loop again ;
  **Else** update $I_M^* \leftarrow I_M(s)$ and generate a new candidate point.

# Log-likelihood of multidimensional Hawkes processes

- Log-likelihood of a multidimensional Hawkes process:

$$\log \mathcal{L}_T = \sum_{m=1}^{M} \log \mathcal{L}_T^m,$$

with: $\log \mathcal{L}_T^m = \int_0^T \log \lambda_s^m \, dN_s^m - \int_0^T \lambda_s^m \, ds.$

# Table of contents

# Branching representation of a Hawkes process I

## Proposition (Branching representation of a Hawkes process)

*Let $T > 0$. Let $\lambda_0 > 0$ and $\nu : [0, \infty) \to \mathbb{R}_+$ a deterministic function. Consider the following procedure:*

1. *Generate a time-homogeneous Poisson process $\{t_i^{(0)}\}$ with intensity $\lambda_0$ on $[0, T]$.*
2. *For each $t_i^{(0)}$, generate a non-homogeneous Poisson process $\{t_i^{(1)}\}$ with deterministic intensity $t \mapsto \nu(t - t_i^{(0)})$ on $[t_i^{(0)}, T]$.*
3. *Repeat operation 2. for each point of all the sets $\{t_i^{(1)}\}$, then for each point of all the sets $\{t_i^{(2)}\}$, etc. until there is no new point generated in $[0, T]$.*

*Then the ordered set of all the points generated in $[0, T]$ is a Hawkes process $N$ on $[0, T]$ with intensity $\lambda_t = \lambda_0 + \int_0^t \nu(t - s)\, dN_s$.*

- Link with Galton-Watson processes

# Branching representation of a Hawkes process II

- The average number of points (*offspring*) generated by one point (*parent*) is $\|\nu\|_1$. We retrieve the three regimes (sub-critical, critical, supercritical) discussed above depending on the value of $\|\nu\|_1$ compared to 1..
- Provides a new algorithm for the simulation of Hawkes processes
- Immediate extension to the case of time-dependent baseline intensity and/or the case of a multidimensional Hawkes process.

# Table of contents

# Second-order characterization of Hawkes processes

## Proposition (Second-order characterization of Hawkes processes [1])

*A multivariate Hawkes process with stationary increments $N_t = (N^i_t)_{i=1,\dots,D}$ is uniquely defined by its first-order statistics (the expectation of the intensity) and its second-order statistics (the covariance structure), i.e. by*

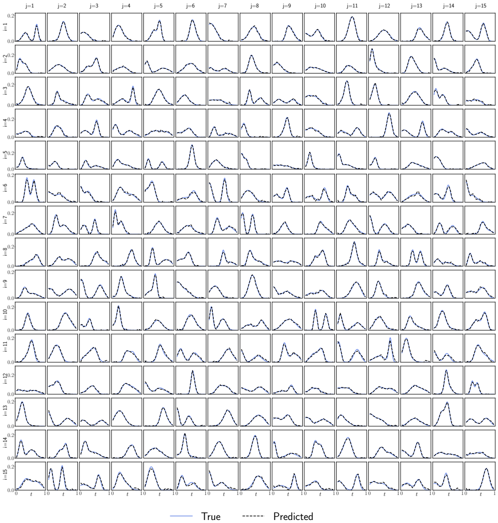
$$\begin{cases} \Lambda^i = \mathbf{E}[\lambda^i_t], \\ \mathrm{Cov}(dN^i_t, dN^j_{t'}) = \mathbf{E}(dN^i_t dN^j_{t'}) - \Lambda^i \Lambda^j \, dt \, dt'. \end{cases}$$

- More precisely, the kernel matrix $\varphi(t)$ of the Hawkes process is the unique solution of the matrix equation
$$g(t) = \varphi(t) + (\varphi \star g)(t), t \geq 0.$$
where the operator $\star$ stands for regular matrix multiplications where all the multiplications are replaced by convolutions, and the matrix $g(t)$ has components satisfying
$$g^{ij}(t)dt = \mathbf{E}(dN^i_t | dN_0 = 1) - \mathbf{1}_{\{i=j\}}\delta(t) - \Lambda^i dt.$$

# Non-parametric estimation of Hawkes processes

- A characterization equation can also be developed for multidimensional marked Hawkes processes.
- The matrix convolution equation can be solved by numerical quadrature integration (original method proposed in [1]) or more recently via Deep Galerkin-like networks ([3], physics-informed networks).

Source : [3]



# Table of contents

# Lab 3 - Questions I

Preliminary remarks:

- In this lab you need an external Python library capable of maximum likelihood estimation of one-dimensional Hawkes processes. We will use this one: `https://pypi.org/project/hawkes/`.
- Executing this lab may take a long time. Start by using small horizons when coding, and then run the notebook with larger horizons when you are satisfied with your preliminary results.
- You may write temporary results to files to avoid re-running simulations or estimations. To avoid any trouble during the evaluation execution, make sure that the names of your temporary result files start with the full name of your notebook. Temporary result files cannot be uploaded when submitting the lab.

# Lab 3 - Questions II

Start by writing two functions:

- a function that simulates one path of a Hawkes process with a *generic* decreasing kernel using the *thinning* algorithm ;
- a function that simulates one path of a Hawkes process with a *generic* decreasing kernel using the *branching* algorithm.

In each case, inputs should be the constant baseline intensity, the kernel function and the horizon.

# Lab 3 - Questions III

1. **Properties of Hawkes MLE estimates**. Check that MLE estimators computed with the Hawkes library on samples simulated by your simulators exhibit expected statistical properties.
2. **Computational cost of Hawkes simulators**. Compare for an exponential kernel the computational cost of your thinning algorithm, your branching algorithm, and the simulation of the Hawkes library. Estimate the complexity of these algorithm w.r.t. the horizon. Explain.
3. **A Hawkes process for trades.** Is a Hawkes process a good model for the time dynamics of the trades reported in your dataset ? Use statistical arguments to support your answers.

# References I


[1] E. Bacry and J.-F. Muzy. "First-and second-order statistics characterization of Hawkes processes and non-parametric estimation". In: *IEEE Transactions on Information Theory* 62.4 (2016), pp. 2184–2202.

[2] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

[3] T. Fabre and I. Muni Toke. "Neural Hawkes: Non-Parametric Estimation in High Dimension and Causality Analysis in Cryptocurrency Markets". In: *Quantitative Finance* (2025). In press.

[4] A. G. Hawkes. "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1 (1971), pp. 83–90.

[5] P. W. Lewis and G. S. Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Naval research logistics quarterly* 26.3 (1979), pp. 403–413.

# References II


[6] Y. Ogata. "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1 (1981), pp. 23–31.

[7] Y. Ogata. "The asymptotic behaviour of maximum likelihood estimators for stationary point processes". In: *Annals of the Institute of Statistical Mathematics* 30.1 (1978), pp. 243–261.

[8] I. Rubin. "Regular point processes and their detection". In: *IEEE Transactions on Information Theory* 18.5 (1972), pp. 547–557.