Measuring Discrepancies Between Poisson and Exponential Hawkes Processes

Rachele Foschi

ONLINE
FIRST

Springer

Measuring Discrepancies Between Poisson and Exponential Hawkes Processes

Rachele Foschi[1] (ORCID)

© The Author(s) 2020

Abstract

Poisson processes are widely used to model the occurrence of similar and independent events. However they turn out to be an inadequate tool to describe a sequence of (possibly differently) interacting events. Many phenomena can be modelled instead by Hawkes processes. In this paper we aim at quantifying how much a Hawkes process departs from a Poisson one with respect to different aspects, namely, the behaviour of the stochastic intensity at jump times, the cumulative intensity and the interarrival times distribution. We show how the behaviour of Hawkes processes with respect to these three aspects may be very irregular. Therefore, we believe that developing a single measure describing them is not efficient, and that, instead, the departure from a Poisson process with respect to any different aspect should be separately quantified, by means of as many different measures. Key to defining these measures will be the stochastic intensity and the integrated intensity of a Hawkes process, whose properties are therefore analysed before introducing the measures. Such quantities can be also used to detect mistakes in parameters estimation.

1 Introduction

A Poisson process models, by means of the unordered vector of its jump times, the occurrence of similar and independent events. This means that the jumps of a Poisson process are in a sense unexpected and the Poisson process, despite its mathematical tractability, is an inadequate tool when we believe that some connection exists among events; in particular when some events are caused by some previous events and it becomes therefore possible to predict or even to prevent them.

✉ Rachele Foschi
    rachele.foschi@unipi.it

1   Department of Economics and Management, Università degli Studi di Pisa, Via Cosimo Ridolfi 10, 56124 Pisa, Italy

A widespread tool for modelling this kind of phenomenon is through the use of Hawkes processes.

The name Hawkes processes is due to the seminal work (Hawkes 1971), setting the theoretical basis for the study of self-exciting processes (see also Daley and Vere-Jones 2008; Bacry et al. 2015 and references therein), that were actually already used in engineering and reliability theory (see e.g. Rangan and Grace 1988 and references therein). Currently Hawkes processes are applied in a number of fields: in geology, to earthquakes or volcanic eruptions, in biology, to population growth, spread of infections, neuronal activity, in computer science or social sciences, to networks and social interactions, and in finance, to order book dynamics, defaults and so on (see e.g. Zhuang et al. 2002; Reynaud-Bouret et al. 2013; Delattre et al. 2016; Hawkes 2018).

The main feature of this class of processes is that they exhibit self-excitation behaviour, that is the jump intensity increases immediately after each jump. It is reasonable to require that such a self-excitation effect is limited in time and decays as time elapses. It is also reasonable and very frequent in the above applications, assuming the decay to have an exponential form. This features correspond to a specific expression of the stochastic intensity in the exponential Hawkes model, that will be our main object of investigation:

$$\bar{\lambda}_t = \lambda_0 + \alpha \sum_{T_i < t} e^{-\beta(t-T_i)},$$

where $t \in \mathbf{R}_+$ denotes time, T_i the random time of the i-th jump and λ_0, α, β are deterministic, constant and positive parameters. The exponential Hawkes model also benefits of a particular mathematical tractability, mainly due to its Markov property.

We are interested in quantitatively describing this stochastic intensity and further characteristics, with the aim of distinguishing between Hawkes and Poisson processes. Towards achieving this aim, we present and study a variety of measures, quantifying the shift between the two processes.

In general, there are three possible scenarios, that would have to be separately discussed. We may deal with measuring:

– discrepancy between a Hawkes process and a Poisson one;
– discrepancy between a Hawkes process and a process not having a specific or a known law;
– discrepancy between a Poisson process and a process not having a specific or a known law.

In the present paper, we focus on the first situation.

Knowing how far a process is from a Poisson process may be of interest due to many aspects: from a theoretical point of view, for problems of model choice, or from a decision-making perspective, if we think for example that a cost is associated with jumps. A large distance from a Poisson process corresponds to the fact that a system is very sensitive to past shocks (or there is a contagion among them) and/or it is scarcely resilient to shocks.

The measures we define will also be helpful in detecting inaccuracies in parameters estimation. It may happen, in fact, that the estimated parameters provide no evidence against the hypothesis that a process is Poisson. We specifically refer to the case when α is close to the "critical" value 0.

By definition, a Hawkes process with $\alpha = 0$ is a Poisson process.

From data, instead, it is not easy to establish whether $\alpha = 0$ is the true value of the parameter, therefore it is important studying the behaviour of the process, in terms of its distance from a Poisson process, when $\alpha \approx 0$.

In fact, if it is not exactly $\alpha = 0$, the parameter β too has an effect on such a distance.

On the other hand, we can state that, as $\beta \longrightarrow +\infty$, the Hawkes process tends to a Poisson one, but there are not results quantifying their similarity, for finite values of β.

Indeed, just the parameters α, β provide a first quantification of the distance between Hawkes and Poisson processes: they respectively measure instantaneous impact and persistence of each jump and, (together with λ_0) allow us to write and compute the intensity, that completely describe the process, also dynamically.

Intuitively, when α is very small and β is very large, we expect to deal with a process that is close to a Poisson one. If β is fixed, the larger α, the farther the process is from a Poisson; viceversa; if α is fixed, the larger β, the closer the process is to a Poisson one.

In the other cases, a variety of not "ordered" situations may present, each one depending on the interaction of those particular values of α, β (and possibly λ_0).

In fact, the two parameters in themselves do not provide a sufficient information, we are rather interested in their total effect; therefore we need a function combining them. The intensity already describes such a global effect, but it depends on time; instead we need a more synthetic quantity.

However, since the difference between Hawkes and Poisson processes is complex and does not involve only one aspect, we illustrate several measures, each one quantifying the Hawkes–Poisson distance with respect to a different characteristic feature.

The paper is structured as follows: in Section 2, we recall basic notation and results on Hawkes processes. Section 3 is devoted to study some properties both of stochastic intensity and integrated intensity; these theoretical properties justify the definition, in Section 4, of two measures respectively based on stochastic intensity and integrated intensity, namely the dispersion $\mathcal{M}_{\mathrm{disp}}$ of the values of the intensity just before jump times and the relative difference $\mathcal{M}_\Lambda$ between the integrated intensities of a Hawkes and of a Poisson process. Again in Section 4, we define further measures, derived from another relevant feature of Hawkes processes, that is the frequency distribution of inter-arrival times. Section 5 contains numerical examples and an application to data. Section 6 concludes.

2 Notation and Basic Results

Definition 1 A simple counting process $N = \{N_t\}_{t \geq 0}$ with natural filtration $\mathcal{F} = \sigma(N_s, 0 \leq s \leq t)$ is a Hawkes process if it has a stochastic intensity $\bar{\lambda}$ given by

$$\bar{\lambda}_t = \lambda_0 + \int_0^t \Phi(t - s) dN_s, \tag{1}$$

where $\lambda_0 \in [0, +\infty)$ is constant, and the kernel $\Phi \geq 0$ is a deterministic function, locally integrable on $\mathbf{R}_+$, wrt the Lebesgue measure.

Conventionally, we set $N_0 = 0$, while we denote by $\{T_n\}_{n \in \mathbf{N}}$ the random jump times of N.

In particular, we consider an exponential decay of self-excitation. Hence the intensity can be written as

$$\bar{\lambda}_t = \lambda_0 + \alpha \sum_{T_i < t} e^{-\beta(t - T_i)};$$

note that $\bar{\lambda}$ turns out to be a càglàd process.

Remark 1 In order to deal with a process with a finite number of jumps on limited intervals, we set $\alpha < \beta$. Otherwise, the excitation effect would overcome the decay and the process would be explosive. For processes on $\mathbf{R}$, the condition $\alpha < \beta$ is equivalent to stationarity. However, when the process starts at λ_0 from $t = 0$, then $\alpha < \beta$ implies asymptotic stationarity.

Given $\mathcal{F}_{t-}$, we know exactly when the jumps before t occurred and, to indicate that they are no longer random, we name them t_n rather than T_n. Given $\mathcal{F}_{t-}$, we are able to reconstruct the path realized by the intensity up to $t-$, and we call λ_t the intensity process conditional to $\mathcal{F}_{t-}$. In particular λ_t is such that

$$P(N_{t+dt} - N_t = 1|\mathcal{F}_{t-}) = \lambda_t dt + o(dt), \tag{2}$$

i.e.

$$\lambda_t \equiv \lim_{dt \to 0} \frac{1}{dt} P(N_{t+dt} - N_t = 1|\mathcal{F}_{t-})$$

and

$$\lambda_t = E[\bar{\lambda}_t|\mathcal{F}_{t-}] = \lambda_0 + \alpha \sum_{t_i < t} e^{-\beta(t - t_i)}.$$

We recall some formulas that will be used or generalized in the following.

For notational simplicity, since we will deal with absolutely continuous distributions, we will write the probability densities as $P(T = t)$ instead of $P(T \in [t, t + dt))$. Reference (Rangan and Grace 1988) provides an expression for the conditional density

$$P(T_2 = t_2|T_1 = t_1)$$
$$= \lim_{dt_2 \to 0} \frac{1}{dt_2} P \text{ (jump in } [t_2, t_2 + dt_2)|T_1 = t_1, \text{ no jumps in } (t_1, t_2))$$
$$P \text{ (no jumps in } (t_1, t_2)|T_1 = t_1)$$
$$= \lambda_{t_2} \exp\left(-\int_{t_1}^{t_2} \lambda_s ds\right). \tag{3}$$

By recursively applying (3), the finite-dimensional distribution of the process can be computed as

$$P(T_1 = t_1, ..., T_n = t_n) = \lambda_{t_1} \cdots \lambda_{t_n} \exp\left(-\int_0^{t_n} \lambda_t dt\right),$$

and, by changing variables, the distribution of the first n inter-arrival times,

$$P(T_1 = s_1, T_2 - T_1 = s_2, ..., T_n - T_{n-1} = s_n) =$$
$$\lambda_{s_1} \cdots \lambda_{s_1 + \cdots + s_n} \exp\left(-\int_0^{s_1 + \cdots + s_n} \lambda_t dt\right).$$

In parallel with the intensity (conditional or stochastic), a cumulative or integrated intensity is defined as

$$\Lambda(t) = \int_0^t \lambda_s ds$$

(or $\bar{\Lambda}(t) = \int_0^t \bar{\lambda}_s ds$). See Daley and Vere-Jones (2008), Karr (1991), and Rasmussen (2011). We recall two relevant properties of $\Lambda(t)$:

– $\Lambda(t)$ can be seen as the expected number of jumps of the process N on $[0, t]$. In case of stationarity,

$$\frac{\Lambda(t)}{t} = \mathbf{E}[\bar{\lambda}_t] = \frac{\lambda_0}{1 - \frac{\alpha}{\beta}}.$$

- $\Lambda(t_{i-1}, t_i) \equiv \Lambda(t_i) - \Lambda(t_{i-1})$, $i \in \mathbf{N}$, are the inter-arrival times of a standard Poisson process.

3 Analytical and Distributional Properties of Stochastic Intensity and of the Integrated Intensity

As mentioned above, the first important difference between a (homogeneous) Poisson process and a Hawkes process concerns the characterization in terms of the jump intensity: a constant for the first one, a collection of random variables with distribution depending on the process itself, in the second case.

The intensity of a Hawkes process has also a specific form, highlighting the main difference of behaviour of such a process with respect to a Poisson one, that is self-excitation, i.e. the occurrence of a jump increases the probability of further jumps.

In the literature, in particular in the reliability field, the intensity also has a key role in describing dynamic dependence properties among events.

Therefore we believe that those aspects of diversity of a process from a Poisson one, that concern dependence among jumps, may be well described and measured at first by quantities derived from the intensity.

We start this section by studying some properties of the distribution of the intensity of a Hawkes process, in particular of the distribution of the intensity just before jumps.

First of all, we recall a recursive formula for λ_t: let $\mathcal{F}_{t-}$ be given and, in particular, t_n be the time of the last jump before t. Then

$$\lambda_t = \lambda_0 + (\lambda_{t_n} - \lambda_0 + \alpha)e^{-\beta(t-t_n)} \qquad (4)$$

(see Foschi et al. 2019).

Remark 2 We notice that applying (4) does not require the knowledge of $\mathcal{F}_{t-}$. For computing λ_t, it is sufficient to know the value of λ_{t_n} and that t_n is the time of the last jump before t. Therefore, given λ_{t_n}, the process after t_n is independent of the history of the process before t_n

3.1 Intensity Values Just Before a Jump Time

Let T_n denote the random time of the n-th jump, $\bar{\lambda}_{T_n}$ the intensity immediately before the n-th jump and E_t^n the event {no jumps in $(T_n, T_n + t)$} and let $\bar{\lambda}_{T_n}$ be given. Then an analogue of Eq. 4 holds for the stochastic intensity as well:

$$\bar{\lambda}_{T_n+t|E_t^n} = \lambda_0 + (\bar{\lambda}_{T_n} - \lambda_0 + \alpha)e^{-\beta t} \qquad (5)$$

Remark 3 Equation 5 can be generalized to compute $\bar{\lambda}_{\tau+t|E}$, with τ a given time instant, not necessarily a jump time, and E a suitable event (see Foschi et al. 2019).

By Eq. 5 and Remark 2, it follows that

$$P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1}, ..., \bar{\lambda}_{T_2} = \ell_2) = P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1}),$$

that is

Lemma 1 *The process $\{\bar{\lambda}_{T_n}\}_{n\in\mathbf{N}}$ is markovian.*

By Lemma 1, we can prove

Theorem 1 *The following recursive formula holds:*

$$
\begin{cases}
P(\bar{\lambda}_{T_n} = \ell_n) = \displaystyle\int_{\lambda_0}^{\lambda_0 + (n-2)\alpha} \ell_n \left(\frac{\ell_n - \lambda_0}{\ell_{n-1} - \lambda_0 + \alpha} \right)^{\frac{\lambda_0}{\beta}} \\
\quad \exp\left(\frac{\ell_n - \alpha - \ell_{n-1}}{\beta} \right) P(\bar{\lambda}_{T_{n-1}} = \ell_{n-1}) d\ell_{n-1} \\
P(\bar{\lambda}_{T_2} = \ell_2) = \ell_2 \left(\frac{\ell_2 - \lambda_0}{\alpha} \right)^{\frac{\lambda_0}{\beta}} \exp\left(\frac{\ell_2 - \alpha - \lambda_0}{\beta} \right)
\end{cases}
\tag{6}
$$

Proof As a first step, we compute the distribution of $\bar{\lambda}_{T_2}$, partitioning wrt the variable T_1, i.e.

$$
P(\bar{\lambda}_{T_2} = \ell_2) = \int_0^{+\infty} P(\bar{\lambda}_{T_2} = \ell_2 | T_1 = t_1) P(T_1 = t_1) dt_1.
\tag{7}
$$

It can be easily obtained that $P(T_1 = t_1) = \lambda_0 e^{-\lambda_0 t_1}$.

By changing variable,

$$
P(\bar{\lambda}_{T_2} = \ell_2 | T_1 = t_1) = P(\lambda_0 + \alpha e^{-\beta(T_2 - t_1)} = \ell_2 | T_1 = t_1)
$$

can be traced back to the (conditional) jump time distribution in Eq. 3:

$$
P\left(T_2 = t_1 - \frac{1}{\beta} \log\left(\frac{\ell_2 - \lambda_0}{\alpha} \right) \middle| T_1 = t_1 \right) = \ell_2 \exp\left(-\int_{t_1}^{t_1 + \sigma_1} \lambda_\tau d\tau \right),
$$

where $\sigma_1 = -\frac{1}{\beta} \log\left(\frac{\ell_2 - \lambda_0}{\alpha} \right)$.

For $\tau \in (t_1, t_1 + \sigma_1)$, $\lambda_\tau = (\lambda_{t_1} - \lambda_0 + \alpha)e^{-\beta(\tau - t_1)} + \lambda_0$, that is, for $s \in (0, \sigma_1)$, $\lambda_{s+t_1} = \alpha e^{-\beta s} + \lambda_0$; hence

$$
\int_{t_1}^{t_1 + \sigma_1} \lambda_\tau d\tau = \int_0^{\sigma_1} (\alpha e^{-\beta s} + \lambda_0) d\tau = -\frac{\lambda_0}{\beta} \log\left(\frac{\ell_2 - \lambda_0}{\alpha} \right) - \frac{\alpha}{\beta} \left(\frac{\ell_2 - \lambda_0}{\alpha} - 1 \right).
$$

Therefore, by substituting in Eq. 7,

$$
P(\bar{\lambda}_{T_2} = \ell_2) = \int_0^{+\infty} \ell_2 \exp\left(\frac{\lambda_0}{\beta} \log\left(\frac{\ell_2 - \lambda_0}{\alpha} \right) + \frac{\alpha}{\beta} \left(\frac{\ell_2 - \lambda_0}{\alpha} - 1 \right) \right) \lambda_0 e^{-\lambda_0 t_1} dt_1
$$

$$
= \ell_2 \left(\frac{\ell_2 - \lambda_0}{\alpha} \right)^{\frac{\lambda_0}{\beta}} \exp\left(\frac{\ell_2 - \lambda_0 - \alpha}{\beta} \right) \int_0^{+\infty} \lambda_0 e^{-\lambda_0 t_1} dt_1.
$$

Since $\int_0^{+\infty} \lambda_0 e^{-\lambda_0 t_1} dt_1 = 1$, the second statement in the thesis is obtained.

In order to compute $P(\bar{\lambda}_{T_n} = \ell_n)$, we apply the total probabilities formula with the partition defined by the variable $\bar{\lambda}_{T_{n-1}}$, i.e.

$$
P(\bar{\lambda}_{T_n} = \ell_n) = \int_{\lambda_0}^{\lambda_0 + \alpha(n-2)} P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1}) P(\bar{\lambda}_{T_{n-1}} = \ell_{n-1}) d\ell_{n-1},
\tag{8}
$$

where the integration limits are set in view of Foschi et al. (2019), Proposition 1.

Since $\bar{\lambda}_{T_n} = \lambda_0 + (\bar{\lambda}_{T_{n-1}} - \lambda_0 + \alpha)e^{-\beta(T_n - T_{n-1})}$,

$$
P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1}) = P\left(T_n - T_{n-1} = -\frac{1}{\beta} \log\left(\frac{\ell_n - \lambda_0}{\ell_{n-1} - \lambda_0 + \alpha} \right) \middle| \bar{\lambda}_{T_{n-1}} = \ell_{n-1} \right),
$$

that, by Foschi et al. (2019), Proposition 2, and by some algebra, can be written as

$$P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1}) = \ell_n \exp\left(\frac{\ell_n - \ell_{n-1} - \alpha}{\beta}\right) \left(\frac{\ell_n - \lambda_0}{\ell_{n-1} - \lambda_0 - \alpha}\right)^{\frac{\lambda_0}{\beta}}. \quad (9)$$

By substituting in Eq. 8, the thesis is proven. □

Remark 4 Notice that, by definition, $\bar{\lambda}_{T_1} = \lambda_{T_1}$ is the intensity immediately before the first jump and therefore it coincides with λ_0, i.e.

$$P(\bar{\lambda}_{T_1} = \ell_1) = \begin{cases} 1 \text{ if } \ell_1 = \lambda_0 \\ 0 \text{ if } \ell_1 \neq \lambda_0 \end{cases}$$

In view of this fact, $P(\bar{\lambda}_{T_2} = \ell_2)$ can be read as $\ell_2 \left(\frac{\ell_2 - \lambda_0}{\ell_1 - \lambda_0 + \alpha}\right)^{\frac{\lambda_0}{\beta}} \exp\left(\frac{\ell_2 - \alpha - \ell_1}{\beta}\right)$.

Corollary 1 *For $n \in \mathbf{N}$, $n \geq 2$, $P(\bar{\lambda}_{T_n} = \ell_n) =$*

$$\left(\prod_{i=2}^{n-1} \int_{\lambda_0}^{\lambda_0 + (i-1)\alpha} d\ell_i\right) \left[\prod_{i=2}^{n} \ell_i \left(\frac{\ell_i - \lambda_0}{\ell_{i-1} - \lambda_0 + \alpha}\right)^{\frac{\lambda_0}{\beta}} \exp\left(\frac{\ell_i - \alpha - \ell_{i-1}}{\beta}\right)\right]. \quad (10)$$

Proof By the total probabilities formula and in view of Lemma 1,

$$P(\bar{\lambda}_{T_n} = \ell_n) = \int_{\lambda_0}^{\lambda_0 + (n-2)\alpha} d\ell_{n-1} \cdots \int_{\lambda_0}^{\lambda_0 + 2\alpha} d\ell_3 \int_{\lambda_0}^{\lambda_0 + \alpha} d\ell_2 P(\bar{\lambda}_{T_n} = \ell_n | \bar{\lambda}_{T_{n-1}} = \ell_{n-1})$$

$$\cdot P(\bar{\lambda}_{T_{n-1}} = \ell_{n-1} | \bar{\lambda}_{T_{n-2}} = \ell_{n-2}) \cdots P(\bar{\lambda}_{T_3} = \ell_3 | \bar{\lambda}_{T_2} = \ell_2) P(\bar{\lambda}_{T_2} = \ell_2).$$

The thesis follows by Remark 4 and Eq. 9. □

Remark 5 This result is particularly useful and of straight application: in fact, recursively writing the distribution of $\bar{\lambda}_{T_n}$ allows us to easily implement it. Furthermore, e.g. in Reliability, the function $\bar{\lambda}_t$ can be interpreted as an analogue of the instantaneous wear of a system (see e.g. Cha and Finkelstein 2012), causing the failure of that system when it exceeds a certain (deterministic) threshold. By applying (6) or (10), we are now able to compute the probability of reaching such a threshold. Similarly the (stochastic) integrated intensity $\bar{\Lambda}(t)$ can be used to represent the cumulative wear of a system up to time t and an analogue of Eqs. 6 or 10 can be applied to it.

Step-Wise Convexity of λ_t For any $t \in \mathbf{R} \setminus \{t_1, t_2, \dots\}$, λ_t is infinitely many times differentiable. In particular,

$$\frac{d\lambda_t}{dt} = -\alpha\beta \sum_{t_i < t} e^{-\beta(t - t_i)} < 0, \quad \frac{d^2\lambda_t}{dt^2} = \alpha\beta^2 \sum_{t_i < t} e^{-\beta(t - t_i)} > 0,$$

that is λ_t is decreasing in the intervals between two jumps and $\frac{d\lambda_t}{dt}$ is negative but increasing, meaning that it is decreasing in absolute value.

In a neighbourhood of two consecutive jumps times, we have

$$\lim_{t \to t_{j-1}^+} \frac{d\lambda_t}{dt} = -\alpha\beta \sum_{t_i \leq t_{j-1}} e^{-\beta(t_{j-1} - t_i)}, \quad \lim_{t \to t_j^-} \frac{d\lambda_t}{dt} = e^{-\beta(t_j - t_{j-1})} \lim_{t \to t_{j-1}^+} \frac{d\lambda_t}{dt},$$

$$\lim_{t \to t_j^+} \frac{d\lambda_t}{dt} = -\alpha\beta \sum_{t_i \le t_j} e^{-\beta(t_j - t_i)} = \lim_{t \to t_j^-} \frac{d\lambda_t}{dt} - \alpha\beta,$$

$$\lim_{t \to t_j^-} \frac{d\lambda_t}{dt} = -\alpha\beta \sum_{t_i < t_j} e^{-\beta(t_j - t_i)} = -\alpha\beta e^{-\beta(t_j - t_{j-1})} \sum_{t_i \le t_{j-1}} e^{-\beta(t_{j-1} - t_i)}.$$

Hence $\lim_{t \to t_j^+} \frac{d\lambda_t}{dt} < \lim_{t \to t_j^-} \frac{d\lambda_t}{dt}$, that implies that, after the jump, the slope is still negative but more pronounced than before.

3.2 Jumps' Concentration Impact on Intensity and Integrated Intensity

Unlike for Poisson and mixed Poisson processes, for Hawkes processes, the observation of the exact configuration of jumps (e.g. till time t) has a greater information content than the observation of the number of jumps occurred till time t.

This fact and the non-homogeneous concentration of jumps are traits distinguishing a Hawkes process from a Poisson one.

In this subsection, we aim at showing how a different position or a non-homogeneous concentration may affect the intensity and the integrated intensity.

Let us fix $\delta > 0$, $k, m \in \mathbf{N}$ with $m > k$, and define $\lambda_{(n+m)\delta}^{(k)}$ as the intensity at time $(n+m)\delta$ conditional on the fact that on $[n\delta, (n+m)\delta)$ a jump was registered on k different intervals of length δ, and on $m - k$ intervals there were no jumps. Given k, a bigger value of m would indicate that the k jumps can possibly be less concentrated, while for given m a higher value of k means that on the same time period $[n\delta, (n+m)\delta)$ more jumps occurred. One would expect that, given k, $\lambda_{(n+m)\delta}^{(k)}$ is non-increasing in m, because the impact of each jump had more time to decay, while, for a given m, $\lambda_{(n+m)\delta}^{(k)}$ is increasing in k. However this last fact turns out not to be necessarily true, because it depends on how the k intervals, where the jumps occurred, are distributed among the m ones.

To prove this fact, we split k as $k' + k''$, where k' is the number of consecutive jumps occurring on the time interval $[(n+m-k')\delta, (n+m)\delta)$, i.e. just before the observation time $(n+m)\delta$, and k'' is the number of jumps distributed in some way on the remaining $n - k' - 1$ intervals. Note that, by construction, we have no jumps on $[(n+m-k'-1)\delta, (n+m-k')\delta)$.

We then show that, when k increases in such a way that k' decreases and k'' increases, we can find configurations such that $\lambda_{(n+m)\delta}^{(k)}$ is not increasing.

In fact, for a given m, we make an increase of k by some amount h. Let $h = h'' - h' > 0$, where h' is the decrease in the number of recent consecutive jumps, so that k' passes to $k' - h'$, and h'' is the increment of k''. We have to account for the following constraints: $h \le k'$, $1 \le h' \le m - k - 1$. Then we compare the two conditional intensities $\lambda_{(n+m)\delta}^{(k+h)}$ and $\lambda_{(n+m)\delta}^{(k)}$. We point out that we represent by k a same number of jumps, but not the same jumps for both intensities. The term $\lambda_{(n+m)\delta}^{(k+h)} - \lambda_{(n+m)\delta}^{(k)}$ may be negative and its minimum value is attained when the h' jumps are the last ones of the k' ones and the h'' are positioned starting from the $k' - h' + 2$-th place.

Then a sufficient condition guaranteeing that $\lambda_{(n+m)\delta}^{(k)}$ be increasing in k, regardless of the allocation (k', k''), is

$$\sum_{i=0}^{h-1} e^{-\beta\delta(k'-i)} - \sum_{i=0}^{h'+h-1} e^{-\beta\delta(m-i)} < 0,$$

which is equivalent to

$$e^{-\beta\delta k'}(1 - e^{\beta\delta h}) - e^{-\beta\delta m}(1 - e^{\beta\delta(h+h')}) > 0,$$

or also to

$$h < \frac{1}{\beta\delta} \log\left(\frac{1 - e^{\beta\delta(m-k')}}{1 - e^{\beta\delta(m-k'-h')}}\right) - \beta\delta h'.$$

While the intensity value may be affected by the jumps configuration even in opposition to a greater number of jumps, as concerns the integrated intensity, we choose to analyse the impact on Λ of a different concentration of the same number of jumps on a fixed interval $I = [a, b)$.

Proposition 1 *Let $k \geq 2$ be the number of jumps occurred in a fixed interval I. The sparser the configuration of such jumps, the smaller the generated $\Lambda(I) \equiv \Lambda(a, b)$.*

Proof We consider the value of $\bar{\lambda}_a$ is given and distinguish two cases:

- k equidistant jumps on I;
- k jumps in $[a, a + \delta)$, with $\delta \in (0, b - a)$.
   In view of the arbitrariness of I, the case "k jumps in $[b - \delta, b)$" is analogous.

We compute the cumulative intensity in the two cases, respectively $\Lambda_{eq}(I)$ and $\Lambda_\delta(I)$.

Equidistant jumps on I means that the jumps have occurred at times $a + j\frac{b-a}{k}$, for $j = 0, ..., k - 1$.

In view of Eq. 4, this fact allows us to write

$$\Lambda_{eq}(I) = \sum_{i=0}^{k-1} \int_0^{\frac{b-a}{k}} \left(\lambda_0 + (\bar{\lambda}_{a+i\frac{b-a}{k}} - \lambda_0 + \alpha)e^{-\beta\tau}\right) d\tau$$

The integrand can be simplified as follows:

$$\lambda_0 + (\bar{\lambda}_a - \lambda_0)e^{-\beta(\tau+i\frac{b-a}{k})} + \alpha \sum_{j=0}^{i-1} e^{-\beta(\tau+j\frac{b-a}{k})} =$$

$$\lambda_0 + (\bar{\lambda}_a - \lambda_0)e^{-\beta\tau}e^{-\beta i\frac{b-a}{k}} + \alpha e^{-\beta\tau}\frac{1 - e^{-\beta i\frac{b-a}{k}}}{1 - e^{-\beta\frac{b-a}{k}}}.$$

By integrating wrt τ on $\left[0, \frac{b-a}{k}\right)$, we obtain

$$\lambda_0\frac{b-a}{k} + \left[(\bar{\lambda}_a - \lambda_0)e^{-\beta i\frac{b-a}{k}} + \alpha\frac{1 - e^{-\beta i\frac{b-a}{k}}}{1 - e^{-\beta\frac{b-a}{k}}}\right]\int_0^{\frac{b-a}{k}} e^{-\beta\tau} d\tau =$$

$$\lambda_0\frac{b-a}{k} + \frac{1}{\beta}(1 - e^{-\beta\frac{b-a}{k}})\left[(\bar{\lambda}_a - \lambda_0)e^{-\beta i\frac{b-a}{k}} + \alpha\frac{1 - e^{-\beta i\frac{b-a}{k}}}{1 - e^{-\beta\frac{b-a}{k}}}\right].$$

Therefore, we can compute $\Lambda_{eq}(I)$ as

$$\sum_{i=0}^{k-1}\lambda_0\frac{b-a}{k} + \sum_{i=0}^{k-1}\frac{\bar{\lambda}_a - \lambda_0}{\beta}(1 - e^{-\beta\frac{b-a}{k}})e^{-\beta i\frac{b-a}{k}} + \sum_{i=0}^{k-1}\frac{\alpha}{\beta} - \sum_{i=0}^{k-1}\frac{\alpha}{\beta}e^{-\beta i\frac{b-a}{k}},$$

finally obtaining

$$\Lambda_{\text{eq}}(I) = \lambda_0(b-a) + \frac{1}{\beta}(\bar{\lambda}_a - \lambda_0)(1 - e^{-\beta(b-a)}) + \frac{\alpha}{\beta}\left(k - \frac{1 - e^{-\beta(b-a)}}{1 - e^{-\frac{\beta}{k}(b-a)}}\right). \quad (11)$$

Let us split $\Lambda_\delta(I) = \Lambda_{\text{eq}}(a, a+\delta) + \Lambda_{\text{decay}}(a+\delta, b)$, by supposing that the k jumps are equidistant on the interval $[a, a+\delta)$, so that we can apply (11) to compute $\Lambda_{\text{eq}}(a, a+\delta)$, while Λ_{decay} is the integrated intensity on the interval $[a+\delta, b]$ without jumps.

$$\Lambda_{\text{eq}}(a, a+\delta) = \lambda_0\delta + \frac{1}{\beta}(\bar{\lambda}_a - \lambda_0)(1 - e^{-\beta\delta}) + \frac{\alpha}{\beta}\left(k - \frac{1 - e^{-\beta\delta}}{1 - e^{-\frac{\beta}{k}\delta}}\right).$$

By recursively apply (4) (see also Foschi et al. 2019),

$$\bar{\lambda}_{a+\delta} = \lambda_0 + (\bar{\lambda}_a - \lambda_0)e^{-\beta\delta} + \alpha\sum_{j=1}^{k}e^{-\beta j\frac{\delta}{k}}$$

$$= \lambda_0 + (\bar{\lambda}_a - \lambda_0)e^{-\beta\delta} + \alpha e^{-\beta\frac{\delta}{k}}\frac{1 - e^{-\beta\delta}}{1 - e^{-\beta\frac{\delta}{k}}}$$

and, for $\tau \in [0, b - a - \delta)$,

$$\bar{\lambda}_{a+\delta+\tau} = \lambda_0 + (\bar{\lambda}_{a+\delta} - \lambda_0)e^{-\beta\tau};$$

hence

$$\Lambda_{\text{decay}}(a+\delta, b) = \int_0^{b-a-\delta}\bar{\lambda}_{a+\delta+\tau}d\tau = \lambda_0(b-a-\delta) + \frac{1}{\beta}(1 - e^{-\beta(b-a-\delta)})(\bar{\lambda}_{a+\delta} - \lambda_0)$$

$$= \lambda_0(b-a-\delta) + \frac{1}{\beta}(1 - e^{-\beta(b-a-\delta)})\left((\bar{\lambda}_a - \lambda_0)e^{-\beta\delta} + \alpha e^{-\beta\frac{\delta}{k}}\frac{1 - e^{-\beta\delta}}{1 - e^{-\beta\frac{\delta}{k}}}\right).$$

By summing up the two terms, we obtain

$$\Lambda_\delta(I) = \lambda_0(b-a) + \frac{1}{\beta}(\bar{\lambda}_a - \lambda_0)(1 - e^{-\beta(b-a)}) + \frac{\alpha}{\beta}k - \frac{\alpha}{\beta}\frac{1 - e^{-\beta\delta}}{1 - e^{-\frac{\beta}{k}\delta}}(1 - e^{-\beta\delta} + e^{-\beta(b-a-\delta+\frac{\delta}{k})}).$$

At first, we want to check that, for any $\delta \in (0, b-a)$, $\Lambda_\delta(I) \geq \Lambda_{\text{eq}}(I)$. At this aim, we compute

$$\Lambda_\delta(I) - \Lambda_{\text{eq}}(I) = \frac{\alpha}{\beta}\left[\frac{1 - e^{-\beta(b-a)}}{1 - e^{-\frac{\beta}{k}(b-a)}} - \frac{1 - e^{-\beta\delta}}{1 - e^{-\frac{\beta}{k}\delta}}\left(1 - e^{-\frac{\beta}{k}\delta} + e^{-\beta(\frac{b-a+\delta}{k} - \delta)}\right)\right].$$

We obtain that $\Lambda_\delta(I) - \Lambda_{\text{eq}}(I) > 0$, $\forall\alpha, \beta > 0$, when $(b-a-\delta)(k+1) > 0$; this last condition is satisfied $\forall\delta \in (0, b-a)$ and $\forall k \in \mathbf{N}$.

Actually the same kind of condition guarantees that

$$\Lambda_\delta(I) \geq \Lambda_\eta(I) \quad \forall\eta, \delta \in (0, b-a), \ \eta \geq \delta.$$

In fact, $\Lambda_\delta(I) - \Lambda_\eta(I) =$

$$\frac{\alpha}{\beta}\left[\frac{1 - e^{-\beta\eta}}{1 - e^{-\frac{\beta}{k}\eta}}\left(1 - e^{-\frac{\beta}{k}\eta} + e^{-\beta(\frac{b-a+\eta}{k} - \eta)}\right) - \frac{1 - e^{-\beta\delta}}{1 - e^{-\frac{\beta}{k}\delta}}\left(1 - e^{-\frac{\beta}{k}\delta} + e^{-\beta(\frac{b-a+\delta}{k} - \delta)}\right)\right],$$

that is positive $\forall\alpha, \beta > 0$, when $(k+1)(\eta - \delta) > 0$. Since this last condition is satisfied $\forall\eta > \delta$ and $\forall k \in \mathbf{N}$, the thesis is proven. $\square$

4 Non-poissonianity Measures

This section is devoted to the definition of three kinds of measures and to illustrating their specific properties and aims.

Our measures naturally apply to single paths, that we may observe in real situations. However, they can be computed as well on samples of simulated paths of a process. This allows us to quantify how much a Hawkes process with given parameters is far from a Poisson process (even if no data are available) and also provides us with confidence intervals for the different measures, that are useful to possibly compare the values of measures computed on a sample with the theoretical ones.

We introduce the following notation:

Definition 2 Given $\lambda_0, \alpha, \beta > 0$, with $\alpha < \beta$, we denote by $N^{(\lambda_0, \alpha, \beta)}$ the exponential Hawkes process with parameters $(\lambda_0, \alpha, \beta)$ and by $H^{(\lambda_0, \alpha, \beta)} = \left(H_h^{(\lambda_0, \alpha, \beta)} \right)_{h=1,\dots,d}$ a set of d realizations of $N^{(\lambda_0, \alpha, \beta)}$.

4.1 Intensity-Based Measure

By definition, the intensity of a Poisson process is a deterministic constant λ_0, i.e. a degenerate random variable, that therefore has a standard deviation equal to zero.

As we illustrated above, for a Hawkes process $N^{(\lambda_0, \alpha, \beta)}$, both $\{\bar{\lambda}_t\}_{t \in \mathbf{R}_+}$ and $\{\bar{\lambda}_{T_n}\}_{n \in \mathbf{N}}$ are stochastic processes.

We aim at quantifying how much the intensity values are dispersed with respect to a constant. In order to do that, we consider the frequency distribution of $\{\bar{\lambda}_{T_n}\}_{n \in \mathbf{N}}$. To obtain the values of $\{\bar{\lambda}_{T_n}\}_{n \in \mathbf{N}}$, we need to know (or to estimate) the parameters of the Hawkes process; we can alternatively directly estimate the empirical intensity from data with any known method.

Since the intensity of a Poisson process should attain only one value, under this aspect, the distance of a Hawkes process from a Poisson process can be expressed by the dispersion $\mathcal{M}_{\text{disp}}$ of the values attained by $\bar{\lambda}_{T_n}$ for $n \in \mathbf{N}$. Namely

Definition 3 Given a realization H of the Hawkes process $N^{(\lambda_0, \alpha, \beta)}$, let j be the number of its jumps and ℓ_n denote, for $n = 1, \dots, j$, the realizations of $\bar{\lambda}_{T_n}$ on the path H. By setting $\ell_0 = \bar{\lambda}_{T_0} \equiv \lambda_0$,

$$\mathcal{M}_{\text{disp}} = \frac{1}{j+1} \sum_{i=0}^{j} (\ell_i - \bar{\ell})^2, \quad \bar{\ell} = \frac{1}{j+1} \sum_{i=0}^{j} \ell_i.$$

Notice that the dispersion, computed as the sample standard deviation, is meant with respect to the mean value of the $\bar{\lambda}_{T_n}$'s and not with respect to the "Poisson part" of the Hawkes intensity, λ_0.

This measure allows us to get the general idea of how much the intensity values are far from λ_0, but it does not explicitly quantify this distance. Specifically, $\mathcal{M}_{\text{disp}}$ allows us to compare a Hawkes process with any Poisson process, regardless of its intensity, i.e. it answers the more general question whether a Poisson process exists fitting well with the observed phenomenon. In fact, the values of $\bar{\lambda}_{T_n}$ may be concentrated around a $\hat{\lambda}$ much larger than λ_0: the process cannot be approximated with a Poisson process with intensity λ_0, but it may be close to a different Poisson process (the one with intensity $\hat{\lambda}$). Otherwise,

the values of $\bar{\lambda}_{T_n}$ may be in average closer to λ_0, but too sparse, so that neither the Poisson process with intensity λ_0 nor any Poisson process are a good model for the observed process.

$\mathcal{M}_{\text{disp}}$ is susceptible to large fluctuations of the values of $\bar{\lambda}_{T_n}$, i.e. $\mathcal{M}_{\text{disp}}$ is large even if $\bar{\lambda}_{T_n} \approx \lambda_0$ for a high percentage of observations and few times $\bar{\lambda}_{T_n}$ attains very large values.

In other words, $\mathcal{M}_{\text{disp}}$ is more susceptible to the presence of large values of $\bar{\lambda}_{T_n} - \lambda_0$ than to the number of observations leading to a large $\bar{\lambda}_{T_n} - \lambda_0$. For this reason, in some cases, it is discordant with the other measures; or rather we should say that, while other measures quantify an effect (e.g. the expected number of extra-Poisson jumps), which may even be quite weak, $\mathcal{M}_{\text{disp}}$ catches the strength of a potential cause of the effect.

4.2 Integrated Intensity-Based Measure

Another measure of the departure of a process from a Poisson process can be derived by the integral of the "non-Poisson" term of intensity. In particular, in the case of a Hawkes process, the non-Poisson term coincides with the self-excitation term of the intensity. In order to compute this quantity, we need to fix a time horizon T and consider the process for $t \in [0, T]$.

$\Lambda(T) - T\lambda_0$ captures the mean excess of jumps of a Hawkes process with parameters $(\lambda_0, \alpha, \beta)$ with respect to a Poisson process with intensity λ_0. However, in order to make this quantity independent of the time horizon T and of the baseline intensity λ_0, we divide it by $T\lambda_0$ obtaining

$$\mathcal{M}_\Lambda(N^{(\lambda_0,\alpha,\beta)}) = \frac{\Lambda(T) - T\lambda_0}{T\lambda_0}.$$

Hence, along an observed path H, the expected number of non-Poisson jumps is $\mathcal{M}_\Lambda(H)$ times the expected number of Poisson jumps.

$\Lambda(T)$ is computed by applying its definition, as the integral of the conditional intensity λ_t on $[0, T]$.

In view of its theoretical properties, we expect that $\mathcal{M}_\Lambda$:

- can detect the different behaviour of a Hawkes process from a Poisson process in those limit cases, when α is very small or β is very large, when commonly we conclude that the Hawkes process is very similar or tends to a Poisson, or quantify such a similarity;
- is sensitive to clusters, that are a distinguishing feature of a Hawkes process.

We consider only the case of a non-explosive process, i.e. $\alpha < \beta$. Theoretically, for a fixed $\frac{\alpha}{\beta}$, $\mathcal{M}_\Lambda$ is not affected by λ_0. In fact, if we could assume stationarity, $\mathcal{M}_\Lambda = \frac{\beta}{\beta-\alpha} - 1$, regardless of the realized path of the process.

This fact may result counterintuitive, since one could expect that the larger the ratio $\frac{\alpha}{\lambda_0}$ (representing the jump impact on the intensity relatively to the baseline part λ_0), the larger the distance of the Hawkes process from the Poisson one. However, the closed formula for $\mathcal{M}_\Lambda$ in terms of α, β is not accurate when stationarity is not guaranteed, as it happens for processes on the half line. Also this measure, when computed on a Poisson process, would be equal to 0.

4.3 Measures Based on Inter-Arrival Times Frequency Distribution

It is known that inter-arrival times of a Poisson process are i.i.d. exponential r.v.'s. Inter-arrival times of a Hawkes process are not independent nor identically distributed (see Foschi

et al. 2019, Proposition 2). We are interested however in studying how the values of the inter-arrival times are distributed with respect to the data generated by an exponential distribution with a given parameter.

We would like to provide a synthetic and more precise quantification than the information given by a quantile-quantile plot, by measuring the intersection between the hypographs of the frequency distribution of the inter-arrival times and of the exponential density of parameter λ_0.

Definition 4 Given a realization H of the process $N^{(\lambda_0, \alpha, \beta)}$,

$$\mathcal{M}^0_\cap(H) = 1 - \int_0^{+\infty} \min(f(x), \lambda_0 e^{-\lambda_0 x}) dx,$$

where f is the empirical frequency distribution of the inter-arrival times of H.

Remark 6 $\int_0^{+\infty} \min(f(x), \lambda_0 e^{-\lambda_0 x}) dx$ is the area of the intersection between the hypographs of the two functions $f(x)$ and $\lambda_0 e^{-\lambda_0 x}$ (see Fig. 1). Since the integral of a probability density function as well as the one of a relative frequency distribution is 1, $\mathcal{M}^0_\cap \in [0, 1]$.

Remark 7 When N is a Poisson process with intensity λ_0, theoretically, for any realization H (that has an infinite number of jumps) $\mathcal{M}^0_\cap(H) = 0$, and we fix this value as our benchmark. In practice, simulated paths or observed realizations of a process necessarily have a finite number of jumps, and the equality $\mathcal{M}^0_\cap(H) = 0$ holds only as the limit when the number of jumps j tends to infinity. The convergence to 0 is due to the fact that, when $j \to +\infty$, the empirical density converges to the true theoretical density $\lambda_0 e^{-\lambda_0 x}$.

In order to implement the computation of $\mathcal{M}^0_\cap$, an operational definition is needed. We denote by a, b respectively the minimum and the maximum observed inter-arrival time. It is not restrictive considering $a = 0$. We fix $\epsilon > 0$, suitably small with respect to the length $b - a$, e.g. $\epsilon = 10^{-4}$.

$$\mathcal{M}^0_\cap(H) = 1 - \sum_{i=1}^{\lceil b\epsilon^{-1} \rceil} \min \left(\frac{1}{j} \sum_{h=1}^{j} \mathbf{1}_{(\epsilon(i-1), \epsilon i]}(x_h), \, e^{-\lambda_0 \epsilon i} (e^{\lambda_0 \epsilon} - 1) \right),$$

where $\{x_1, \ldots, x_j\}$ are the inter-arrival times of the path H and $e^{-\lambda_0 \epsilon i} (e^{\lambda_0 \epsilon} - 1) = \int_{\epsilon(i-1)}^{\epsilon i} \lambda_0 e^{-\lambda_0 x} dx$.

We expect that, the stronger the self-excitation effect, the higher the number of inter-arrival times not following an exponential law, and therefore the smaller the intersection of the hypograph of their distribution with the one of an exponential distribution. Since $\mathcal{M}^0_\cap$ compares the frequency distribution of the inter-arrival times with the exponential density with parameter λ_0, it quantifies the distance of the Hawkes process from its "baseline" Poisson process.

This kind of measure is also suitable to be extended to the case when we don't know λ_0 or we aim at comparing the Hawkes process with a general Poisson process. In other words, $\mathcal{M}^0_\cap$ can be adapted to measure how much the Poisson best fitting with data is actually well fitting. In this case, we estimate the parameter $\hat{\lambda}$ of the exponential best fitting with the inter-arrival times and, as above, we measure the intersection between the hypograph of the frequency distribution of the inter-arrival times and the one of the exponential density of parameter $\hat{\lambda}$. We define $\mathcal{M}_\cap$ analogously to $\mathcal{M}^0_\cap$:

Definition 5 Given a realization H of the process $N^{(\lambda_0, \alpha, \beta)}$,

$$\mathcal{M}_\cap(H) = 1 - \int_0^{+\infty} \min(f(x), \hat{\lambda} e^{-\hat{\lambda} x}) dx,$$

where f is the empirical frequency distribution of the inter-arrival times of H.

Remark 8 If H is a realization of a Poisson process with intensity λ_0, $\hat{\lambda}$ is very close to λ_0 and therefore $\mathcal{M}_\cap(H)$ is approximatively equal to $\mathcal{M}_\cap^0(H)$.

If H is a realization of a Poisson process with intensity $\lambda \neq \lambda_0$ and $\hat{\lambda}$ is the maximum likelihood estimator (MLE) of λ, the exponential density with parameter $\hat{\lambda}$ is the closest one to the empirical density of the inter-arrival times; thus $\mathcal{M}_\cap^0(H) \geq \mathcal{M}_\cap(H)$. In particular, as $j \to +\infty$, $\mathcal{M}_\cap(H)$ tends to 0, while $\mathcal{M}_\cap^0(H)$ does not.

If, finally, H is a realization of a Hawkes process with parameters $(\lambda_0, \alpha, \beta)$, since $\forall\, t > 0$, $\lambda_t > \lambda_0$, the jumps of H are more frequent than the ones of a Poisson process with intensity λ_0 and therefore $\hat{\lambda} > \lambda_0$. Even if the inter-arrival times have no more an exponential distribution, among the exponential densities, the one with parameter $\hat{\lambda}$ is the closest one to the empirical density of the inter-arrival times, and therefore $\mathcal{M}_\cap^0(H) \geq \mathcal{M}_\cap(H)$.

## 5 Simulations Results and Comparisons

In this section, we illustrate by means of some examples the impact of different sets of parameters $(\lambda_0, \alpha, \beta)$ on the four measures $\mathcal{M}_{\text{disp}}$, $\mathcal{M}_\Lambda$, $\mathcal{M}_\cap^0$, $\mathcal{M}_\cap(H)$ defined above.

Intuitively a small α and a large β make the distances small; still we are interested in quantify it, in order to establish whether a Poisson process can be used to fit the data generated by $N^{(\lambda_0, \alpha, \beta)}$ and possibly the error due to the approximation.

When α and β are both small or both large, we have no more general results to quantify the discrepancy of the Hawkes process from a Poisson one.

We provide examples of processes whose different behaviours may be explained by our measures.
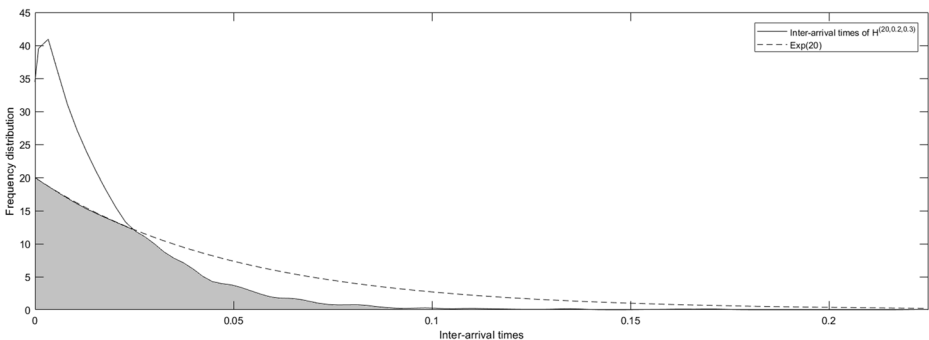
Fig. 1 Intersection between the hypographs of an exponential density of parameter λ_0 (*dashed line*) and the frequency distribution graph of inter-arrival times of the simulated path $H_1^{(20, 0.2, 0.3)}$ (*solid line*). On the x axis are reported the values of the inter-arrival times, on the y axis the relative frequencies

Since the difference from a Poisson process concerns several aspects, we expect that any trait has repercussions on a different measure and therefore that the four measures are not necessarily concordant.

For each process $N^{(\lambda_0,\alpha,\beta)}$ with given parameters, we simulate $d = 2500$ paths $H^{(\lambda_0,\alpha,\beta)}$, on a time interval $[0, T]$, $T = \bar{m}\delta$. $\delta > 0$ is a scale parameter, allowing us, in the applications, to express the time horizon T in the desired unit of measurement. We fix $\delta = 4.96 \cdot 10^{-5}$. For different processes, instead, we choose different $\bar{m}$, in order to avoid realizations $H^{(\lambda_0,\alpha,\beta)}$ with too few jumps. For each $H^{(\lambda_0,\alpha,\beta)}$, we compute the mean values of the four measures and their standard deviations, as reported in Table 1.

As a first result, applying each measure $\mathcal{M}$ to simulated paths provides us with an estimate $\mathcal{M}(H^{(\lambda_0,\alpha,\beta)}) = \frac{1}{d}\sum_{h=1}^{d} \mathcal{M}(H_h^{(\lambda_0,\alpha,\beta)})$ of the theoretical discrepancy $\mathcal{M}(N^{(\lambda_0,\alpha,\beta)})$ between $N^{(\lambda_0,\alpha,\beta)}$ and a Poisson process.

Description of Table 1 In Table 1, we highlight some groups of processes.

In a first group, we fix $\lambda_0 = 3$, $\beta = 15$ and vary α. As we can expect, as α increases, all the measures increase. In particular, the increase of $\mathcal{M}_{\text{disp}}$ is a consequence both of wider intervals for the values of $\bar{\lambda}_{T_n}$'s and of the fact that $\bar{\lambda}_{T_n}$'s are less concentrated around the modal value, that is very close to λ_0. This means that, for small α's the sequence $\{\bar{\lambda}_{T_n}\}_{n=1,...,j}$ exhibits small and dense fluctuations, that become larger and less dense as α increases; furthermore for small α's, $\{\bar{\lambda}_{T_n}\}_{n=1,...,j}$ comes back very often very close to λ_0. This behaviour has also an impact on $\mathcal{M}_\Lambda$: a $\{\bar{\lambda}_{T_n}\}_{n=1,...,j}$ with small and dense fluctuations, often coming back to λ_0 has a hypograph with a smaller area, leading to a small $\mathcal{M}_\Lambda$. Also the measures $\mathcal{M}_\cap^0$, $\mathcal{M}_\cap$ show that the processes are far from Poisson processes. We notice that, in the first two cases, $\hat{\lambda}$ is not significantly different from λ_0 and this implies that also $\mathcal{M}_\cap^0$ and $\mathcal{M}_\cap$ are not significantly different. In the last two cases, $\hat{\lambda} > \lambda_0$ significantly and this implies that $\mathcal{M}_\cap^0 > \mathcal{M}_\cap$; but still $\mathcal{M}_\cap$ is significantly greater than 0.

A wider case record can be observed in the second group of simulated processes, where we fix $\lambda_0 = 20$, $\alpha = 0.2$ and vary β. As β increases, $\mathcal{M}_{\text{disp}}$, $\mathcal{M}_\Lambda$ and $\mathcal{M}_\cap^0$ decrease. However, we notice that the decreasing trend of $\mathcal{M}_\cap^0$ is significant only for small values of β.

For $\beta = 0.3$, $\{\bar{\lambda}_{T_n}\}_{n=1,...,j}$ increases in a first period and then has fluctuations of a small amplitude with respect to the range of its values. Such range narrows as β increases, but, in the meanwhile, the relative amplitude of fluctuations and their frequency increases, until $\{\bar{\lambda}_{T_n}\}_{n=1,...,\bar{n}}$ often takes values very close to λ_0 (see also Fig. 2).

As concerns $\mathcal{M}_\cap$, when β is large enough to cause a quick decay of the intensity towards λ_0, $\hat{\lambda}$ is not significantly different from λ_0 and therefore $\mathcal{M}_\cap$ is not significantly different from $\mathcal{M}_\cap^0$. When β is small (we refer here to the case $\beta = 0.3$), we obtain $\mathcal{M}_\cap^0 > \mathcal{M}_\cap$; furthermore $\mathcal{M}_\cap(H^{(20,0.2,0.3)}) < \mathcal{M}_\cap(H^{(20,0.2,\beta)})$, $\beta = 3, 20, 100$.

The same trend can be observed in the other groups of processes, $H^{(20,1,\beta)}$, $H^{(300,0.2,\beta)}$, $H^{(300,1,\beta)}$ respectively.

Finally, the last two groups in the table show how, even with a very small α (or α small with respect to λ_0), the measures are able to reveal that the processes are not Poisson.

Description of Table 2 Within each group, the values of $\mathcal{M}_\Lambda$ are not significantly different. This feature is consistent with the theoretical formula $\mathcal{M}_\Lambda = (\frac{\beta}{\alpha} - 1)^{-1}$, where $\mathcal{M}_\Lambda$ only depends on the ratio $\frac{\alpha}{\beta}$. The values also are close to the theoretical ones: $\mathcal{M}_\Lambda(N^{(\lambda_0,\alpha,100\alpha)}) = 0.0101$, while $\mathcal{M}_\Lambda(N^{(\lambda_0,\alpha,3\alpha)}) = 0.5$, that is in the confidence intervals of $\mathcal{M}_\Lambda(H^{(\lambda_0,\alpha,3\alpha)})$ for the most part of the simulated processes, namely the ones with

Table 1 The table summarizes the results obtained from the simulation of $d = 2500$ paths for each Hawkes process $N^{(\lambda_0, \alpha, \beta)}$ with parameters λ_0, α, β (reported in the first column)

λ_0, α, β	$\mathcal{M}_{\text{disp}}$	$\mathcal{M}_\Lambda$	$\mathcal{M}_\cap^0$	$\hat{\lambda}$	$\mathcal{M}_\cap$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\lambda_0}$	$\frac{\alpha}{\hat{\lambda}}$	$\bar{m}$
3, 0.2, 15	0.0642	0.0134	0.8051	3.02	0.8044	0.0133	0.0667	0.0663	10^7
	(0.0029)	(0.0003)	(0.0067)	(0.08)	(0.009)				
3, 1, 15	0.3413	0.0686	0.799	3.09	0.7961	0.0667	0.3333	0.3238	10^7
	(0.0168)	(0.0018)	(0.007)	(0.08)	(0.0096)				
3, 5, 15	2.86	0.4537	0.7279	4.09	0.6783	0.3333	1.667	0.2479	10^7
	(0.22)	(0.0172)	(0.0095)	(0.16)	(0.0179)				
20, 0.2, 0.3	8.4914	1.6	0.4104	52.13	0.2085	0.6667	0.01	0.0038	10^6
	(0.797)	(0.083)	(0.0092)	(2.66)	(0.0111)				
20, 0.2, 3	0.3928	0.71	0.2874	21.36	0.2825	0.0667	0.01	0.0094	$3 \cdot 10^6$
	(0.0161)	(0.0013)	(0.006)	(0.4)	(0.0069)				
20, 0.2, 20	0.1427	0.0101	0.2982	20.17	0.2976	0.01	0.01	0.0099	10^6
	(0.0039)	(0.0002)	(0.0062)	(0.36)	(0.0069)				
20, 0.2, 100	0.0633	0.002	0.2998	20.02	0.2997	0.002	0.01	0.001	$3 \cdot 10^6$
	(0.0019)	($3.7 \cdot 10^{-5}$)	(0.0064)	(0.37)	(0.0071)				
20, 1, 3	2.7667	0.4985	0.2653	29.98	0.2075	0.3333	0.05	0.0333	$3 \cdot 10^6$
	(0.143)	(0.0113)	(0.0042)	(0.68)	(0.0061)				
20, 1, 20	0.1427	0.052	0.2927	21.00	0.2898	0.05	0.05	0.04806	$3 \cdot 10^6$
	(0.0039)	(0.001)	(0.0061)	(0.39)	(0.0071)				
20, 1, 100	0.3194	0.01	0.2993	20.08	0.299	0.01	0.05	0.0498	$3 \cdot 10^6$
	(0.0097)	(0.0002)	(0.0064)	(0.36)	(0.0071)				
300, 0.2, 0.3	120.5367	0.9602	0.2665	588.27	0.0757	0.6667	0.0007	0.0003	$3 \cdot 10^5$
	(3.0703)	(0.0197)	(0.0072)	(11.59)	(0.0027)				
300, 0.2, 3	2.6757	0.0696	0.0854	320.89	0.0798	0.0667	0.0007	0.0006	$3 \cdot 10^5$
	(0.1149)	(0.0011)	(0.0042)	(4.98)	(0.0042)				
300, 0.2, 20	0.5635	0.01	0.0831	303.06	0.0825	0.01	0.0007	0.0007	$3 \cdot 10^5$
	(0.0241)	(0.0002)	(0.0044)	(4.6)	(0.0045)				
300, 0.2, 100	0.2454	0.002	0.0831	300.64	0.0829	0.002	0.0007	0.0007	$3 \cdot 10^5$
	(0.0064)	($2.9 \cdot 10^{-5}$)	(0.0044)	(4.48)	(0.0044)				
300, 1, 3	19.7375	0.4834	0.1672	445.24	0.0714	0.3333	0.0033	0.0022	$3 \cdot 10^5$
	(0.985)	(0.0089)	(0.0066)	(8.14)	(0.0034)				
300, 1, 20	2.9438	0.0524	0.0843	315.54	0.0807	0.05	0.0033	0.0032	$3 \cdot 10^5$
	(0.1292)	(0.0008)	(0.0042)	(4.87)	(0.0043)				
300, 1, 100	1.236	0.0101	0.0831	302.87	0.0826	0.01	0.0033	0.0033	$3 \cdot 10^5$
	(0.0317)	(0.0002)	(0.0044)	(4.55)	(0.0044)				
20, 0.002, 0.3	0.0252	0.0063	0.4817	20.13	0.481	0.0067	0.0001	$9.9 \cdot 10^{-5}$	10^6
	(0.0015)	(0.0002)	(0.0123)	(0.65)	(0.0152)				
20, 0.002, 3	0.0037	0.0007	0.4833	20.02	0.4832	0.0007	0.0001	$9.9 \cdot 10^{-5}$	10^6
	(0.0002)	($2 \cdot 10^{-5}$)	(0.012)	(0.63)	(0.0148)				
20, 0.02, 0.3	0.2705	0.066	0.4648	21.36	0.4578	0.0667	0.001	0.0009	10^6
	(0.0163)	(0.0022)	(0.0117)	(0.7)	(0.0145)				

Table 1 (continued)

λ_0, α, β	$\mathcal{M}_{\text{disp}}$	$\mathcal{M}_\Lambda$	$\mathcal{M}_\cap^0$	$\hat{\lambda}$	$\mathcal{M}_\cap$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\lambda_0}$	$\frac{\alpha}{\hat{\lambda}}$	$\bar{m}$
20, 0.02, 3	0.0373	0.0067	0.4814	20.15	0.4806	0.0067	0.001	0.001	10^6
	(0.0025)	(0.0002)	(0.0121)	(0.64)	(0.0149)				

In the others columns are reported the mean of the measures $\mathcal{M}_{\text{disp}}$, $\mathcal{M}_\Lambda$, $\mathcal{M}_\cap^0$, $\mathcal{M}_\cap$ and, before this last one, the mean of the $\hat{\lambda}$. The standard errors of each measure are reported on the line below within brackets. $\bar{m}$ is the number of trials in each path for generating jumps

$\alpha = 1, 5, 100$. Again within the group with $\frac{\alpha}{\beta} = \frac{1}{3}$, for the ones with $\lambda_0 = 300$, we obtain not significantly different values of $\mathcal{M}_\cap^0$. The values of $\mathcal{M}_\cap$ are quite close each other, but display an increasing trend wrt α or β, while the values of $\mathcal{M}_{\text{disp}}$ are strongly different and increasing wrt α or β. This means that, as α increases, the amplitude of the fluctuations of $\{\bar{\lambda}_{T_n}\}_{n=1,\dots,j}$ increases too. However the larger amplitude is balanced by a faster decay, i.e. by a higher frequency of the fluctuations, making the area of the hypograph of $\{\bar{\lambda}_{T_n}\}_{n=1,\dots,j}$ almost constant.

The use of the simulated paths has however a further application, to check whether, in the cases when a parameters' estimation is needed, the estimated values are correct.

In fact, apart from $\mathcal{M}_\cap$, whose computation is completely based on data and does not involve the parameters, the measures depend on the values of λ_0, α, β.

As an example, in Table 3 we show how the measures may be biased, when a wrong value is assigned to a parameter.

Description of Table 3 The corresponding values of the measures are significantly different from the values of the same measures computed on $H^{(\hat{\lambda}_0,\hat{\alpha},\hat{\beta})}$ (for some of them, see Table 1). This inconsistency informs us that $(\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$ is not a correct estimate of the parameters of the generating process.

We notice that, even if the wrong value of α, $\hat{\alpha}$, is very small (close to 0), the values of the measures are still significantly different from 0, meaning that they lead us to the correct conclusion that the process is not Poisson, when instead the parameters' estimate does not give us such an evidence. We also remark that $\mathcal{M}_{\text{disp}}$ is not affected by a misspecification of
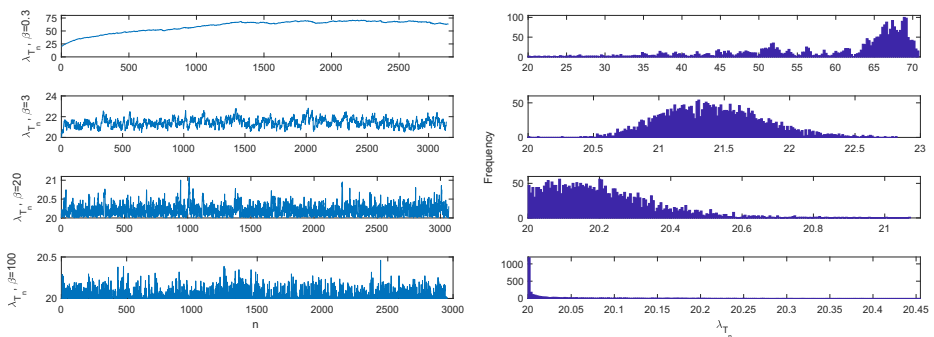
Fig. 2 The figure summarizes the behaviour of $\bar{\lambda}_{T_n}$ for paths respectively generated by $N^{(20,0.2,\beta)}$, $\beta = 0.3, 3, 20, 100$. The figures on the *left column* are the plots of $\bar{\lambda}_{T_n}$ with respect to the jump index n. The figures on the *right column* represent for any $\beta = 0.3, 3, 20, 100$, the frequencies of values of $\bar{\lambda}_{T_n}$

Table 2 In this table, we report the results for two groups of processes; within each group, we vary the parameters keeping the ratio $\frac{\alpha}{\beta}$ constant

λ_0, α, β	$\mathcal{M}_{\text{disp}}$	$\mathcal{M}_\Lambda$	$\mathcal{M}_\cap^0$	$\hat{\lambda}$	$\mathcal{M}_\cap$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\lambda_0}$	$\frac{\alpha}{\hat{\lambda}}$	$\bar{n}$
20, 0.2, 20	0.1427	0.0101	0.2982	20.17	0.2976	0.01	0.01	0.0099	10^6
	(0.0039)	(0.0002)	(0.0062)	(0.36)	(0.0069)				
20, 1, 100	0.3194	0.01	0.2993	20.08	0.299	0.01	0.05	0.0498	$3 \cdot 10^6$
	(0.0097)	(0.0002)	(0.0064)	(0.36)	(0.0071)				
300, 0.2, 20	0.5635	0.01	0.0831	303.06	0.0825	0.01	0.0007	0.0007	$3 \cdot 10^5$
	(0.0241)	(0.0002)	(0.0044)	(4.6)	(0.0045)				
300, 1, 100	1.236	0.0101	0.0831	302.87	0.0826	0.01	0.0033	0.0033	$3 \cdot 10^5$
	(0.0317)	(0.0002)	(0.0044)	(4.55)	(0.0044)				
3, 5, 15	2.86	0.4537	0.7279	4.09	0.6783	0.3333	1.667	0.2479	10^7
	(0.22)	(0.0172)	(0.0095)	(0.16)	(0.0179)				
20, 1, 3	2.7667	0.4985	0.2653	29.98	0.2075	0.3333	0.05	0.0333	$3 \cdot 10^6$
	(0.143)	(0.0113)	(0.0042)	(0.68)	(0.0061)				
300, 1, 100	1.236	0.0101	0.0831	302.87	0.0826	0.01	0.0033	0.0033	$3 \cdot 10^5$
	(0.0317)	(0.0002)	(0.0044)	(4.55)	(0.0044)				
300, 100, 300	111.72	0.4898	0.1768	441.01	0.086	0.3333	0.3333	0.2268	$3 \cdot 10^5$
	(3.67)	(0.0097)	(0.0072)	(8.72)	(0.0043)				
300, 500, 1500	287.52	0.4547	0.1815	409.32	0.1199	0.3333	1.6667	1.2215	$3 \cdot 10^5$
	(13.35)	(0.0093)	(0.0074)	(8.39)	(0.0063)				
300, 1000, 3000	453.55	0.4356	0.1706	382.09	0.1273	0.3333	3.3333	2.6171	$3 \cdot 10^5$
	(27.26)	(0.0101)	(0.0081)	(9.13)	(0.0076)				

λ_0, while $\mathcal{M}_\cap^0$ is not affected by a misspecification of α. The value $\widetilde{\mathcal{M}}_{\text{disp}}$ deriving from the misspecified $\hat{\alpha}$ is such that $\frac{\widetilde{\mathcal{M}}_{\text{disp}}}{\mathcal{M}_{\text{disp}}} = \frac{\hat{\alpha}}{\alpha}$. As to $\widetilde{\mathcal{M}}_\cap^0$, it attains its minimum value, coinciding with $\mathcal{M}_\cap$, for $\hat{\lambda}_0 = \hat{\lambda}$.

5.1 An Application to Data

We consider the sequence of jump times filtered out (see Foschi et al. 2019) from the dataset of the five minutes prices of the assets JPM, from 3/1/2006 to 31/7/2013. On a time horizon $T = 7.53$, corresponding to $\bar{m} = 151791$, we obtain a record of 816 jump times, that can be described by an exponential Hawkes process. In Foschi et al. (2019) a procedure for parameters' estimation is also developed, providing us with the values $\hat{\lambda}_0 = 53.27$, $\hat{\alpha} = 4.72$, $\hat{\beta} = 9.14$. As mentioned, our first goal in computing $\mathcal{M}_{\text{disp}}$, $\mathcal{M}_\Lambda$, $\mathcal{M}_\cap^0$, $\mathcal{M}_\cap$ is quantifying the discrepancy between a Poisson process and the sample H generated by the process $N^{(53.27, 4.72, 9.14)}$ with respect to different aspects. We obtain

$$\mathcal{M}_{\text{disp}}(H) = 21.3905, \ \mathcal{M}_\Lambda(H) = 1.0347,$$

$$\mathcal{M}_\cap^0(H) = 0.3823, \ \hat{\lambda} = 108.6039, \ \mathcal{M}_\cap(H) = 0.3295.$$

However, such measures allow us to achieve another important conclusion. We simulate, with the same $\bar{m} = 151791$, $d = 2500$ paths $H^{(53.27, 4.72, 9.14)}$ of the Hawkes process

Table 3 On the 2500 simulated paths $H^{(20,0.2,0.3)}$ of the Hawkes process $N^{(20,0.2,0.3)}$, we compute the measures as before, but supposing the parameters assume different values $\hat{\lambda}_0, \hat{\alpha}, \hat{\beta}$, reported in the first column

Misspecified parameter(s)	$\widetilde{\mathcal{M}}_{\text{disp}}$	$\widetilde{\mathcal{M}}_\Lambda$	$\widetilde{\mathcal{M}}^0_\cap$
$\hat{\lambda}_0 = 25$	8.4914	1.2824	0.3493
	(0.7973)	(0.0667)	(0.0083)
$\hat{\lambda}_0 = 50$	8.4914	0.64	0.2119
	(0.7973)	(0.033)	(0.008)
$\hat{\lambda}_0 = 52.127$	8.4914	0.615	0.2091
	(0.7973)	(0.032)	(0.0088)
$\hat{\lambda}_0 = 70$	8.4914	0.458	0.2168
	(0.7973)	(0.023)	(0.0148)
$\hat{\alpha} = 0.02$	0.8491	0.1603	0.4104
	(0.0797)	(0.0083)	(0.0092)
$\hat{\alpha} = 0.002$	0.0849	0.016	0.4104
	(0.0079)	(0.0008)	(0.0092)
$\hat{\alpha} = 1$	42.4571	8.015	0.4104
	(3.9865)	(0.4167)	(0.0092)
$\hat{\beta} = 0.45$	5.2996	1.0984	0.4104
	(0.5477)	(0.0568)	(0.0092)
$\hat{\beta} = 3$	0.86	0.1724	0.4104
	(0.0832)	(0.0088)	(0.0092)
$\hat{\lambda}_0 = 25$	4.2457	0.4612	0.3493
$\hat{\alpha} = 0.1$	(0.3987)	(0.033)	(0.0083)

$N^{(53.27,4.72,9.14)}$ and compute

$$\mathcal{M}_{\text{disp}}(H^{(53.27,4.72,9.14)}) = 17.0054\ (2.4148), \quad \mathcal{M}_\Lambda(H^{(53.27,4.72,9.14)}) = 1.0459\ (0.0755),$$

$$\mathcal{M}^0_\cap(H^{(53.27,4.72,9.14)}) = 0.3813\ (0.0098),$$

$$\hat{\lambda} = 111.5729\ (3.1366), \quad \mathcal{M}_\cap(H^{(53.27,4.72,9.14)}) = 0.2853\ (0.0159).$$

We can now check whether the discrepancies obtained from the data are consistent with the reference values obtained from simulated paths. In this case, all the $\mathcal{M}(H)$'s are not significantly different from the $\mathcal{M}(H^{(53.27,4.72,9.14)})$'s and therefore we find a further confirmation of the fact that an exponential Hawkes model is well describing the data in H and that the parameters $\hat{\lambda}_0 = 53.27$, $\hat{\alpha} = 4.72$, $\hat{\beta} = 9.14$ are correctly estimated.

6 Concluding Remarks

After having theoretically studied the effect of jumps and their configurations on conditional or stochastic intensity and integrated intensity, we defined different measures, quantifying the distance between a Hawkes and a Poisson process. Since the difference between a Hawkes and a Poisson process is a complex matter and concerns several aspects, we need different measures, each one quantifying a different trait and having its own advantages and disadvantages. All our measures are designed to be applied to a dataset or to a single path generated by a stochastic process. The only exception is $\mathcal{M}_\Lambda$, that, under some hypotheses, allows us to measure the theoretical discrepancy between a Hawkes data generating process $N^{(\lambda_0,\alpha,\beta)}$ and a Poisson process with intensity λ_0. On the other hand, $\mathcal{M}_\Lambda$ allows us to

compare $N^{(\lambda_0,\alpha,\beta)}$ only to the Poisson process with intensity λ_0 and not to a general Poisson process or to the best fitting Poisson process, like $\mathcal{M}_\cap^0$.

$\mathcal{M}_{\text{disp}}$ instead quantifies the discrepancy between $H^{(\lambda_0,\alpha,\beta)}$ and a Poisson process in general; however, its computation requires the knowledge of (λ_0,α,β) and availability of data.

Unlike the other measures, $\mathcal{M}_\cap$ is computed only from the data: it does not require the knowledge of (λ_0,α,β) and it quantifies the discrepancy of $H^{(\lambda_0,\alpha,\beta)}$ from the best fitting Poisson process. This also means that, for a given path H generated by $N^{(\lambda_0,\alpha,\beta)}$ (with f the frequency distribution of the inter-arrival times of H), in the case when we don't know λ_0 and its estimate is $\hat{\lambda}_0 \geq \lambda_0$,

$$\mathcal{M}_\cap(H) = \min_{\hat{\lambda}_0 \geq \lambda_0} \left(1 - \int_0^{+\infty} \min(f(x), \hat{\lambda}_0 e^{-\hat{\lambda}_0 x}) dx\right) = \min_{\hat{\lambda}_0 \geq \lambda_0} \widetilde{\mathcal{M}}_\cap^0(H).$$

The fact that $\mathcal{M}_\cap$ can be computed only from data has both advantages and disadvantages. A negative aspect of $\mathcal{M}_\cap$ is that its accuracy depends on the sample size and on the precision ϵ. However, on the other hand, the main advantage of $\mathcal{M}_\cap$ is that $\mathcal{M}_\cap$ is insensitive to parameters' misspecification and therefore it provides the most reliable criterion, when both $\mathcal{M}_{\text{disp}}$ and $\mathcal{M}_\Lambda$ are small, for deciding whether H is generated by a Hawkes process or by a Poisson.

Furthermore, since $\mathcal{M}_\cap$ is not affected by the values of (λ_0,α,β), it can be used to detect a misspecification of the parameters: in fact, it may happen, e.g. when the estimated $\hat{\alpha}$ is very small, that the misspecified values of $\mathcal{M}_{\text{disp}}(H)$ and $\mathcal{M}_\Lambda(H)$, with H, generated by $N^{(\lambda_0,\alpha,\beta)}$, are not significantly different from $\mathcal{M}_{\text{disp}}(H^{(\lambda_0,\hat{\alpha},\beta)})$ and $\mathcal{M}_\Lambda(H^{(\lambda_0,\hat{\alpha},\beta)})$ respectively. $\mathcal{M}_\cap(H)$, instead, remains significantly different from $\mathcal{M}_\cap(H^{(\lambda_0,\hat{\alpha},\beta)})$.

A further step along this line is to extend these measures in order to deal with distances between unspecified processes and Poisson or Hawkes processes: a comparison with a Poisson process may be useful for a general analysis of the process; a comparison with a Hawkes process may be more application-oriented. We leave this investigation to further research.

References

Bacry E, Mastromatteo I, Muzy JF (2015) Hawkes processes in finance. Market Microstructure and Liquidity 1(1):59
Cha JH, Finkelstein M (2012) Information-based thinning of point processes and its application to shock models. J Stat Plan Inference 142:2345–2350

Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes, vol II: general theory and structure. Springer, New York. 2nd revised and extended ed edition

Delattre S, Fournier N, Hoffmannm M (2016) Hawkes processes on large networks. Ann Appl Probab 26(1):216–261

Foschi R, Lilla F, Mancini C (2019) Warnings about future jumps: properties of the exponential hawkes model. Available at SSRN: https://ssrn.com/abstract=3459443 or https://doi.org/10.2139/ssrn.3459443

Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. Biometrika 58(1):83–90

Hawkes AG (2018) Hawkes processes and their applications to finance: a review. Quant Finan 18(2):193–198

Karr A (1991) Point processes and their statistical inference. CRC Press

Rangan A, Grace RE (1988) A non-Markov model for the optimum replacement of self-repairing systems subject to shocks. J Appl Probab 25(2):375–382

Rasmussen JG (2011) Lecture notes: temporal point processes and the conditional intensity function. arXiv:1806.00221

Reynaud-Bouret P, Rivoirard V, Tuleau-Malot C (2013) Inference of functional connectivity in neurosciences via Hawkes processes. In: 2013 IEEE global conference on signal and information processing, pp 317–320

Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380