
Hawkes Processes

Patrick J. Laub · Thomas Taimre · Philip K. Pollett

Last edited: July 10, 2015

Abstract Hawkes processes are a particularly interesting class of stochastic process that have been applied in diverse areas, from earthquake modelling to financial analysis. They are point processes whose defining characteristic is that they ‘self-excite’, meaning that each arrival increases the rate of future arrivals for some period of time. Hawkes processes are well established, particularly within the financial literature, yet many of the treatments are inaccessible to one not acquainted with the topic. This survey provides background, introduces the field and historical developments, and touches upon all major aspects of Hawkes processes.

1 Introduction

Events that are observed in time frequently cluster naturally. An earthquake typically increases the geological tension in the region where it occurs, and aftershocks likely follow [1]. A fight between rival gangs can ignite a spate of criminal retaliations [2]. Selling a significant quantity of a stock could precipitate a trading flurry or, on a larger scale, the collapse of a Wall Street investment bank could send shockwaves through the world’s financial centres [3].

The *Hawkes process* (HP) is a mathematical model for these ‘self-exciting’ processes, named after its creator Alan G. Hawkes [4]. The HP is a counting process that models a sequence of ‘arrivals’ of some type over time, for example, earthquakes, gang violence, trade orders, or bank defaults. Each arrival *excites* the process in the sense that the chance of a subsequent arrival is increased for some time period after the initial arrival. As such, it is a non-Markovian extension of the Poisson process.

Some datasets, such as the number of companies defaulting on loans each year [5], suggest that the underlying process is indeed self exciting. Furthermore, using the basic Poisson process to model say the arrival of trade orders of a stock is highly inappropriate, because participants in equity markets exhibit a herding behaviour, a standard example of economic reflexivity [6].

The process of generating, model fitting, and testing the goodness of fit of HPs is examined in this survey. As the HP literature in financial fields is particularly well developed, applications in these areas are considered chiefly here.

P. J. Laub

Department of Mathematics, The University of Queensland, Qld 4072, Australia, and
Department of Mathematics, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark.
E-mail: p.laub@uq.edu.au|math.au.dk]

T. Taimre

Department of Mathematics, The University of Queensland, Qld 4072, Australia
E-mail: t.taimre@uq.edu.au

P. K. Pollett

Department of Mathematics, The University of Queensland, Qld 4072, Australia
E-mail: pkp@maths.uq.edu.au

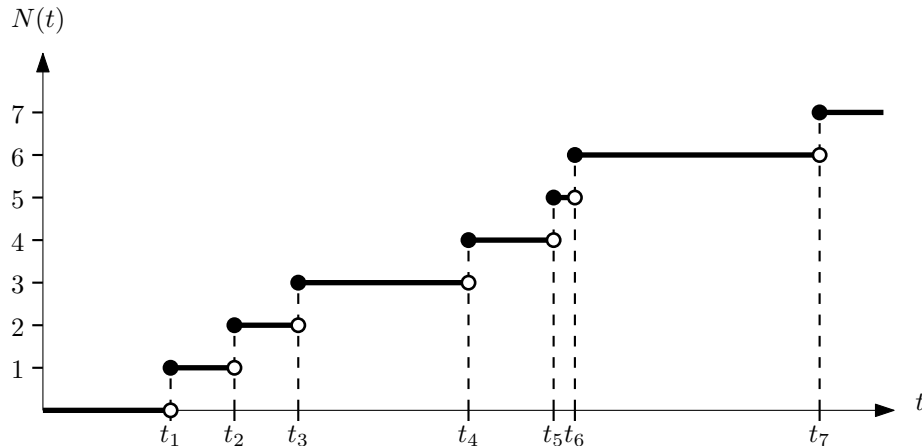


Fig. 1: An example point process realisation $\{t_1, t_2, \dots\}$ and corresponding counting process $N(t)$.

2 Background

Before discussing HPs, some key concepts must be elucidated. Firstly, we briefly give definitions for counting processes and point processes, thereby setting essential notation. Secondly, we discuss the lesser-known conditional intensity function and compensator, both core concepts for a clear understanding of HPs.

2.1 Counting and point processes

We begin with the definition of a counting process.

Definition 1 (Counting process) A *counting process* is a stochastic process $(N(t) : t \geq 0)$ taking values in \mathbb{N}_0 that satisfies $N(0) = 0$, is almost surely (a.s.) finite, and is a right-continuous step function with increments of size +1. Further, denote by $(\mathcal{H}(u) : u \geq 0)$ the *history* of the arrivals up to time u . (Strictly speaking $\mathcal{H}(\cdot)$ is a filtration, that is, an increasing sequence of σ -algebras.)

A counting process can be viewed as a cumulative count of the number of ‘arrivals’ into a system up to the current time. Another way to characterise such a process is to consider the sequence of random arrival times $\mathbf{T} = \{T_1, T_2, \dots\}$ at which the counting process $N(\cdot)$ has jumped. The process defined as these arrival times is called a point process, described in Definition 2 (adapted from [7]); see Fig. 1 for an example point process and its associated counting process.

Definition 2 (Point process) If a sequence of random variables $\mathbf{T} = \{T_1, T_2, \dots\}$, taking values in $[0, \infty)$, has $\mathbb{P}(0 \leq T_1 \leq T_2 \leq \dots) = 1$, and the number of points in a bounded region is a.s. finite, then \mathbf{T} is a (*simple*) *point process*.

The counting and point process terminology is often interchangeable. For example, if one refers to a Poisson process or a HP then the reader must infer from the context whether the counting process $N(\cdot)$ or the point process of times \mathbf{T} is being discussed.

One way to characterise a particular point process is to specify the distribution function of the next arrival time conditional on the past. Given the history up until the last arrival u , $\mathcal{H}(u)$, define (as per [8]) the conditional c.d.f. (and p.d.f.) of the next arrival time T_{k+1} as

$$F^*(t | \mathcal{H}(u)) = \int_u^t \mathbb{P}(T_{k+1} \in [s, s + ds] | \mathcal{H}(u)) ds = \int_u^t f^*(s | \mathcal{H}(u)) ds.$$

The joint p.d.f. for a realisation $\{t_1, t_2, \dots, t_k\}$ is then, by the chain rule,

$$f(t_1, t_2, \dots, t_k) = \prod_{i=1}^k f^*(t_i | \mathcal{H}(t_{i-1})). \quad (1)$$

In the literature the notation rarely specifies $\mathcal{H}(\cdot)$ explicitly, but rather a superscript asterisk is used (see for example [9]). We follow this convention and abbreviate $F^*(t | \mathcal{H}(u))$ and $f^*(t | \mathcal{H}(u))$ to $F^*(t)$ and $f^*(t)$, respectively.

Remark 1 The function $f^*(t)$ can be used to classify certain classes of point processes. For example, if a point process has an $f^*(t)$ which is independent of $\mathcal{H}(t)$ then the process is a *renewal process*.

2.2 Conditional intensity functions

Often it is difficult to work with the conditional arrival distribution $f^*(t)$. Instead, another characterisation of point processes is used: the conditional intensity function. Indeed if the conditional intensity function exists it uniquely characterises the finite-dimensional distributions of the point process (see Proposition 7.2.IV of [9]). Originally this function was called the hazard function [10] and was defined as

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}. \quad (2)$$

Although this definition is valid, we prefer an intuitive representation of the conditional intensity function as the expected rate of arrivals conditioned on $\mathcal{H}(t)$:

Definition 3 (Conditional intensity function) Consider a counting process $N(\cdot)$ with associated histories $\mathcal{H}(\cdot)$. If a (non-negative) function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(t+h) - N(t) | \mathcal{H}(t)]}{h}$$

which only relies on information of $N(\cdot)$ in the past (that is, $\lambda^*(t)$ is $\mathcal{H}(t)$ -measurable), then it is called the *conditional intensity function* of $N(\cdot)$.

The terms ‘self-exciting’ and ‘self-regulating’ can be made precise by using the conditional intensity function. If an arrival causes the conditional intensity function to increase then the process is said to be *self-exciting*. This behaviour causes temporal clustering of \mathbf{T} . In this setting $\lambda^*(t)$ must be chosen to avoid *explosion*, where we use the standard definition of explosion as the event that $N(t) - N(s) = \infty$ for $t - s < \infty$. See Fig. 2 for an example realisation of such a $\lambda^*(t)$.

Alternatively, if the conditional intensity function drops after an arrival the process is called *self-regulating* and the arrival times appear quite temporally regular. Such processes are not examined hereafter, though an illustrative example would be the arrival of speeding tickets to a driver over time (assuming each arrival causes a period of heightened caution when driving).

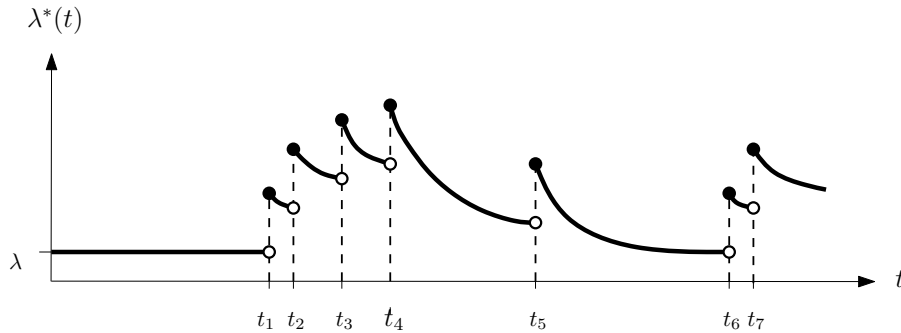


Fig. 2: An example conditional intensity function for a self-exciting process.

2.3 Compensators

Frequently the integrated conditional intensity function is needed (for example, in parameter estimation and goodness of fit testing); it is defined as follows.

Definition 4 (Compensator) For a counting process $N(\cdot)$ the non-decreasing function

$$\Lambda(t) = \int_0^t \lambda^*(s) ds$$

is called the *compensator* of the counting process.

In fact, a compensator is usually defined more generally and exists even when $\lambda^*(\cdot)$ does not exist. Technically $\Lambda(t)$ is the unique $\mathcal{H}(t)$ predictable function, with $\Lambda(0) = 0$, and is non-decreasing, such that $N(t) = M(t) + \Lambda(t)$ almost surely for $t \geq 0$ and where $M(t)$ is an $\mathcal{H}(t)$ local martingale, whose existence is guaranteed by the Doob–Meyer decomposition theorem. However, for HPs $\lambda^*(\cdot)$ always exists (in fact, as we shall see in Section 3, a HP is defined in terms of this function) and therefore Definition 4 is sufficient for our purposes.

3 Literature review

With essential background and core concepts outlined in Section 2, we now turn to discussing HPs, including their useful immigration–birth representation. We briefly touch on generalisations, before turning to an illustrative account of HPs for financial applications.

3.1 The Hawkes process

Point processes gained a significant amount of attention in the field of statistics during the 1950s and 1960s. First, Cox [10] introduced the notion of a doubly stochastic Poisson process (now called the Cox process) and Bartlett [11, 12, 13] investigated statistical methods for point processes based on their power spectral densities. At IBM Research Laboratories, Lewis [14] formulated a point process model (for computer failure patterns) which was a step in the direction of the HP. The activity culminated in the significant monograph by Cox and Lewis [15] on time series analysis; modern researchers appreciate

this text as an important development of point process theory since it canvassed their wide range of applications [9, p. 16].

It was in this context that Hawkes [4] set out to bring Bartlett’s spectral analysis approach to a new type of process: a self-exciting point process. The process Hawkes described was a one-dimensional point process (though originally specified for $t \in \mathbb{R}$ as opposed to $t \in [0, \infty)$), and is defined as follows.

Definition 5 (Hawkes process) Consider $(N(t) : t \geq 0)$ a counting process, with associated history $(\mathcal{H}(t) : t \geq 0)$, that satisfies

$$\mathbb{P}(N(t+h) - N(t) = m \mid \mathcal{H}(t)) = \begin{cases} \lambda^*(t)h + o(h), & m = 1 \\ o(h), & m > 1 \\ 1 - \lambda^*(t)h + o(h), & m = 0 \end{cases}.$$

Suppose the process’ conditional intensity function is of the form

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-u) dN(u) \quad (3)$$

for some $\lambda > 0$ and $\mu : (0, \infty) \rightarrow [0, \infty)$ which are called the *background intensity* and *excitation function* respectively. Assume that $\mu(\cdot) \neq 0$ to avoid the trivial case, that is, a homogeneous Poisson process. Such a process $N(\cdot)$ is a *Hawkes process*.

Remark 2 The definition above has t as non-negative, however an alternative form of the HP is to consider arrivals for $t \in \mathbb{R}$ and set $N(t)$ as the number of arrivals in $(0, t]$. Typically HP results hold for both definitions, though we will specify that this second $t \in \mathbb{R}$ definition is to be used when it is required.

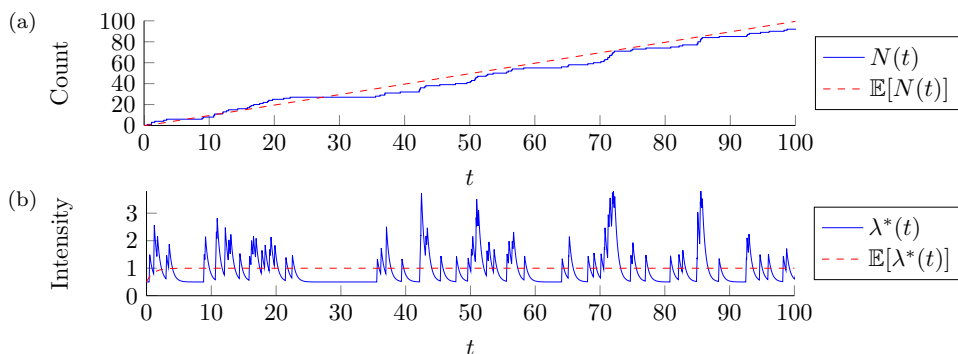


Fig. 3: (a) A typical Hawkes process realisation $N(t)$, and its associated $\lambda^*(t)$ in (b), both plotted against their expected values.

Remark 3 In modern terminology, Definition 5 describes a *linear* HP—the *nonlinear* version is given later in Definition 6. Unless otherwise qualified, the HPs in this paper will refer to this linear form.

A realisation of a HP is shown in Fig. 3 with the associated path of the conditional intensity process. Hawkes [16] soon extended this single point process into a collection of self- and mutually-exciting point processes, which we will turn to discussing after elaborating upon this one-dimensional process.

3.2 Hawkes conditional intensity function

The form of the Hawkes conditional intensity function in (3) is consistent with the literature though it somewhat obscures the intuition behind it. Using $\{t_1, t_2, \dots, t_k\}$ to denote the observed sequence of past arrival times of the point process up to time t , the Hawkes conditional intensity is

$$\lambda^*(t) = \lambda + \sum_{t_i < t} \mu(t - t_i).$$

The structure of this $\lambda^*(\cdot)$ is quite flexible and only requires specification of the background intensity $\lambda > 0$ and the excitation function $\mu(\cdot)$. A common choice for the excitation function is one of exponential decay; Hawkes [4] originally used this form as it simplified his theoretical derivations [17]. In this case $\mu(t) = \alpha e^{-\beta t}$, which is parameterised by constants $\alpha, \beta > 0$, and hence

$$\lambda^*(t) = \lambda + \int_{-\infty}^t \alpha e^{-\beta(t-s)} dN(s) = \lambda + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}. \quad (4)$$

The constants α and β have the following interpretation: each arrival in the system instantaneously increases the arrival intensity by α , then over time this arrival's influence decays at rate β .

Another frequent choice for $\mu(\cdot)$ is a power law function, giving

$$\lambda^*(t) = \lambda + \int_{-\infty}^t \frac{k}{(c + (t-s))^p} dN(s) = \lambda + \sum_{t_i < t} \frac{k}{(c + (t-t_i))^p}$$

with some positive scalars c, k , and p . The power law form was popularised by the geological model called Omori's law, used to predict the rate of aftershocks caused by an earthquake [18]. More computationally efficient than either of these excitation functions is a piecewise linear function as in [19]. However, the remaining discussion will focus on the exponential form of the excitation function, sometimes referred to as the HP with *exponentially decaying intensity*.

One can consider the impact of setting an initial condition $\lambda^*(0) = \lambda_0$, perhaps in order to model a process from some time after it is started. In this scenario the conditional intensity process (using the exponential form of $\mu(\cdot)$) satisfies the stochastic differential equation

$$d\lambda^*(t) = \beta(\lambda - \lambda^*(t)) dt + \alpha dN(t), \quad t \geq 0.$$

Applying stochastic calculus yields the general solution of

$$\lambda^*(t) = e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{\beta(t-s)} dN(s), \quad t \geq 0,$$

which is a natural extension of (4) [20].

3.3 Immigration–birth representation

Stability properties of the HP are often simpler to divine if it is viewed as a branching process. Imagine counting the population in a country where people arrive either via *immigration* or by *birth*. Say that the stream of immigrants to the country form a homogeneous Poisson process at rate λ . Each individual then produces zero or more children independently of one another, and the arrival of births form an inhomogeneous Poisson process.

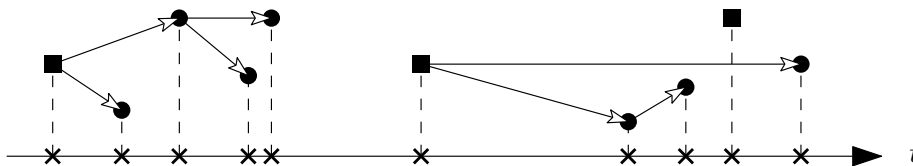


Fig. 4: Hawkes process represented as a collection of family trees (immigration–birth representation). Squares (■) indicate immigrants, circles (●) are offspring/descendants, and the crosses (×) denote the generated point process.

An illustration of this interpretation can be seen in Fig. 4. In branching theory terminology, this *immigration–birth representation* describes a Galton–Watson process with a modified time dimension. Hawkes [21] used the representation to derive asymptotic characteristics of the process, such as the following result.

Theorem 1 (Hawkes process asymptotic normality) *If*

$$0 < n := \int_0^\infty \mu(s) \, ds < 1 \text{ and } \int_0^\infty s\mu(s) \, ds < \infty$$

then the number of HP arrivals in $(0, t]$ is asymptotically ($t \rightarrow \infty$) normally distributed. More precisely, writing $N(0, t] = N(t) - N(0)$,

$$\mathbb{P} \left(\frac{N(0, t] - \lambda t / (1 - n)}{\sqrt{\lambda t / (1 - n)^3}} \leq y \right) \rightarrow \Phi(y),$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.

Remark 4 More modern work uses the immigration–birth representation for applying Bayesian techniques; see, for example, [22].

For an individual who enters the system at time $t_i \in \mathbb{R}$, the rate at which they produce offspring at future times $t > t_i$ is $\mu(t - t_i)$. Say that the direct offspring of this individual comprise the *first-generation*, and their offspring comprise the *second-generation*, and so on; members of the union of all these generations are called the *descendants* of this t_i arrival.

Using the notation from [23, Section 5.4], define Z_i to be the random number of offspring in the i th generation (with $Z_0 = 1$). As the first-generation offspring arrived from a Poisson process $Z_1 \sim \text{Poi}(n)$ where the mean n is known as the *branching ratio*. This branching ratio (which can take values in $(0, \infty)$) is defined in Theorem 1 and in the case of an exponentially decaying intensity is

$$n = \int_0^\infty \alpha e^{-\beta s} \, ds = \frac{\alpha}{\beta}. \quad (5)$$

Knowledge of the branching ratio can inform development of simulation algorithms. For each immigrant i , the times of the first-generation offspring arrivals—conditioned on knowing the total number of them Z_1 —are each i.i.d. with density $\mu(t - t_i)/n$. Section 6 explores HP simulation methods inspired by the immigration–birth representation in more detail.

The value of n also determines whether or not the HP explodes. To see this, let $g(t) = \mathbb{E}[\lambda^*(t)]$. A renewal-type equation will be constructed for g and then its limiting value will be determined. Conditioning on the time of the first jump,

$$g(t) = \mathbb{E}[\lambda^*(t)] = \mathbb{E}\left[\lambda + \int_0^t \mu(t-s) dN(s)\right] = \lambda + \int_0^t \mu(t-s) \mathbb{E}[dN(s)].$$

In order to calculate this expected value, start with

$$\lambda^*(s) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(s+h) - N(s) | \mathcal{H}(s)]}{h} = \frac{\mathbb{E}[dN(s) | \mathcal{H}(s)]}{ds}$$

and take expectations (and apply the tower property)

$$g(s) = \mathbb{E}[\lambda^*(s)] = \frac{\mathbb{E}[\mathbb{E}[dN(s) | \mathcal{H}(s)]]}{ds} = \frac{\mathbb{E}[dN(s)]}{ds}$$

to see that

$$\mathbb{E}[dN(s)] = g(s) ds.$$

Therefore

$$g(t) = \lambda + \int_0^t \mu(t-s) g(s) ds = \lambda + \int_0^t g(t-s) \mu(s) ds.$$

This renewal-type equation (in convolution notation is $g = \lambda + g \star \mu$) then has different solutions according to the value of n . Asmussen [24] splits the cases into: the *defective* case ($n < 1$), the *proper* case ($n = 1$), and the *excessive* case ($n > 1$). Asmussen's Proposition 7.4 states that for the defective case

$$g(t) = \mathbb{E}[\lambda^*(t)] \rightarrow \frac{\lambda}{1-n}, \quad \text{as } t \rightarrow \infty. \quad (6)$$

However in the excessive case, $\lambda^*(t) \rightarrow \infty$ exponentially quickly, and hence $N(\cdot)$ eventually explodes a.s.

Explosion for $n > 1$ is supported by viewing the arrivals as a branching process. Since $\mathbb{E}[Z_i] = n^i$ (see Section 5.4 Lemma 2 of [23]), the expected number of descendants for one individual is

$$\mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right] = \sum_{i=1}^{\infty} \mathbb{E}[Z_i] = \sum_{i=1}^{\infty} n^i = \begin{cases} \frac{n}{1-n}, & n < 1 \\ \infty, & n \geq 1 \end{cases}.$$

Therefore $n \geq 1$ means that one immigrant would generate infinitely many descendants on average.

When $n \in (0, 1)$ the branching ratio can be interpreted as a probability. It is the ratio of the number of descendants for one immigrant, to the size of their entire family (all descendants plus the original immigrant); that is

$$\frac{\mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right]}{1 + \mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right]} = \frac{\frac{n}{1-n}}{1 + \frac{n}{1-n}} = \frac{\frac{n}{1-n}}{\frac{1}{1-n}} = n.$$

Therefore, any HP arrival selected at random was generated *endogenously* (a child) with probability (w.p.) n or *exogenously* (an immigrant) w.p. $1 - n$. Most properties of the HP rely on the process being *stationary*, which is another way to insist that $n \in (0, 1)$ (a rigorous definition is given in Section 3.4), so this is assumed hereinafter.

3.4 Covariance and power spectral densities

HPs originated from the spectral analysis of general stationary point processes. The HP is stationary for finite values of t when it is defined as per Remark 2, so we will use this definition for the remainder of Subsection 3.4. Finding the power spectral density of the HP gives access to many techniques from the spectral analysis field; for example, model fitting can be achieved by using the observed periodogram of a realisation. The power spectral density is defined in terms of the covariance density. Once again the exposition is simplified by using the shorthand that

$$dN(t) = \lim_{h \downarrow 0} N(t+h) - N(t).$$

Unfortunately the term ‘stationary’ has many different meanings in probability theory. In this context the HP is stationary when the jump process $(dN(t) : t \geq 0)$ —which takes values in $\{0, 1\}$ —is *weakly stationary*. This means that $\mathbb{E}[dN(t)]$ and $\text{Cov}(dN(t), dN(t+s))$ do not depend on t . Stationarity in this sense does not imply stationarity of $N(\cdot)$ or stationarity of the inter-arrival times [25]. One consequence of stationarity is that $\lambda^*(\cdot)$ will have a long term mean (as given by (6))

$$\bar{\lambda}^* := \mathbb{E}[\lambda^*(t)] = \frac{\mathbb{E}[dN(t)]}{dt} = \frac{\lambda}{1-n}. \quad (7)$$

The *(auto)covariance density* is defined, for $\tau > 0$, to be

$$R(\tau) = \text{Cov} \left(\frac{dN(t)}{dt}, \frac{dN(t+\tau)}{d\tau} \right).$$

Due to the symmetry of covariance, $R(-\tau) = R(\tau)$, however $R(\cdot)$ cannot be extended to the whole of \mathbb{R} because there is an atom at 0. For simple point processes $\mathbb{E}[(dN(t))^2] = \mathbb{E}[dN(t)]$ (since $dN(t) \in \{0, 1\}$) therefore for $\tau = 0$

$$\mathbb{E}[(dN(t))^2] = \mathbb{E}[dN(t)] = \bar{\lambda}^* dt.$$

The *complete covariance density* (complete in that its domain is all of \mathbb{R}) is defined as

$$R^{(c)}(\tau) = \bar{\lambda}^* \delta(\tau) + R(\tau) \quad (8)$$

where $\delta(\cdot)$ is the Dirac delta function.

Remark 5 Typically $R(0)$ is defined such that $R^{(c)}(\cdot)$ is everywhere continuous. Lewis [25, p. 357] states that strictly speaking $R^{(c)}(\cdot)$ “does not have a ‘value’ at $\tau = 0$ ”. See [12, 15], and [4] for further details.

The corresponding *power spectral density function* is then

$$S(\omega) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\omega} R^{(c)}(\tau) d\tau = \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right]. \quad (9)$$

Up to now the discussion (excluding the final value of (7)) has considered general stationary point processes. To apply the theory specifically to HPs we need the following result.

Theorem 2 (Hawkes process power spectral density) *Consider a HP with an exponentially decaying intensity with $\alpha < \beta$. The intensity process then has covariance density, for $\tau > 0$,*

$$R(\tau) = \frac{\alpha\beta\lambda(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)\tau}.$$

Hence, its power spectral density is, $\forall \omega \in \mathbb{R}$,

$$S(\omega) = \frac{\lambda\beta}{2\pi(\beta - \alpha)} \left(1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right).$$

Proof (Adapted from [4].) Consider the covariance density for $\tau \in \mathbb{R} \setminus \{0\}$:

$$R(\tau) = \mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \right] - \bar{\lambda}^{*2}. \quad (10)$$

Firstly note that, via the tower property,

$$\begin{aligned} \mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \right] &= \mathbb{E} \left[\mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \mid \mathcal{H}(t + \tau) \right] \right] \\ &= \mathbb{E} \left[\frac{dN(t)}{dt} \mathbb{E} \left[\frac{dN(t + \tau)}{d\tau} \mid \mathcal{H}(t + \tau) \right] \right] \\ &= \mathbb{E} \left[\frac{dN(t)}{dt} \lambda^*(t + \tau) \right]. \end{aligned}$$

Hence (10) can be combined with (3) to see that $R(\tau)$ equals

$$\mathbb{E} \left[\frac{dN(t)}{dt} \left(\lambda + \int_{-\infty}^{t+\tau} \mu(t + \tau - s) dN(s) \right) \right] - \bar{\lambda}^{*2},$$

which yields

$$\begin{aligned} R(\tau) &= \bar{\lambda}^* \mu(\tau) + \int_{-\infty}^{\tau} \mu(\tau - v) R(v) dv \\ &= \bar{\lambda}^* \mu(\tau) + \int_0^{\infty} \mu(\tau + v) R(v) dv + \int_0^{\tau} \mu(\tau - v) R(v) dv. \end{aligned} \quad (11)$$

Refer to Appendix A.1 for details; this is a Wiener–Hopf-type integral equation. Taking the Laplace transform of (11) gives

$$\mathcal{L} \{R(\tau)\} (s) = \frac{\alpha \bar{\lambda}^* (2\beta - \alpha)}{2(\beta - \alpha)(s + \beta - \alpha)}.$$

Refer to Appendix A.2 for details. Note that (5) and (7) supply $\bar{\lambda}^* = \beta\lambda/(\beta - \alpha)$, which implies that

$$\mathcal{L} \{R(\tau)\} (s) = \frac{\alpha\beta\lambda(2\beta - \alpha)}{2(\beta - \alpha)^2(s + \beta - \alpha)}.$$

Therefore,

$$R(\tau) = \mathcal{L}^{-1} \left\{ \frac{\alpha\beta\lambda(2\beta - \alpha)}{2(\beta - \alpha)^2(s + \beta - \alpha)} \right\} = \frac{\alpha\beta\lambda(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)\tau}.$$

The values of $\bar{\lambda}^*$ and $\mathcal{L}\{R(\tau)\}(s)$ are then substituted into the definition given in (9):

$$\begin{aligned}
S(\omega) &= \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right] \\
&= \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_0^{\infty} e^{-i\tau\omega} R(\tau) d\tau + \int_0^{\infty} e^{i\tau\omega} R(\tau) d\tau \right] \\
&= \frac{1}{2\pi} \left[\bar{\lambda}^* + \mathcal{L}\{R(\tau)\}(i\omega) + \mathcal{L}\{R(\tau)\}(-i\omega) \right] \\
&= \frac{1}{2\pi} \left[\bar{\lambda}^* + \frac{\alpha\bar{\lambda}^*(2\beta - \alpha)}{2(\beta - \alpha)(i\omega + \beta - \alpha)} + \frac{\alpha\bar{\lambda}^*(2\beta - \alpha)}{2(\beta - \alpha)(-i\omega + \beta - \alpha)} \right] \\
&= \frac{\lambda\beta}{2\pi(\beta - \alpha)} \left[1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right].
\end{aligned}$$

Remark 6 The power spectral density appearing in Theorem 2 is a shifted scaled Cauchy p.d.f.

Remark 7 As $R(\cdot)$ is a real-valued symmetric function, its Fourier transform $S(\cdot)$ is also real-valued and symmetric, that is,

$$S(\omega) = \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right] = \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} \cos(\tau\omega) R(\tau) d\tau \right],$$

and

$$S_+(\omega) := S(-\omega) + S(\omega) = 2S(\omega).$$

It is common that $S_+(\cdot)$ is plotted instead of $S(\cdot)$, as in Section 4.5 of [15]; this is equivalent to wrapping the negative frequencies over to the positive half-line.

3.5 Generalisations

The immigration–birth representation is useful both theoretically and practically. However it can only be used to describe *linear* HPs. Brémaud and Massoulié [26] generalised the HP to its nonlinear form:

Definition 6 (Nonlinear Hawkes process) Consider a counting process with conditional intensity function of the form

$$\lambda^*(t) = \Psi \left(\int_{-\infty}^t \mu(t-s) N(ds) \right)$$

where $\Psi : \mathbb{R} \rightarrow [0, \infty)$, $\mu : (0, \infty) \rightarrow \mathbb{R}$. Then $N(\cdot)$ is a *nonlinear Hawkes process*. Selecting $\Psi(x) = \lambda + x$ reduces $N(\cdot)$ to the linear HP of Definition 5.

Modern work on nonlinear HPs is much rarer than the original linear case (for simulation see pp. 96–116 of [7], and associated theory in [27]). This is due to a combination of factors; firstly, the generalisation was introduced relatively recently, and secondly, the increased complexity frustrates even simple investigations.

Now to return to the extension mentioned earlier, that of a collection of self- and *mutually-exciting* HPs. The processes being examined are collections of one-dimensional HPs which ‘excite’ themselves and each other.

Definition 7 (Mutually exciting Hawkes process) Consider a collection of m counting processes $\{N_1(\cdot), \dots, N_m(\cdot)\}$ denoted \mathbf{N} . Say $\{T_{i,j} : i \in \{1, \dots, m\}, j \in \mathbb{N}\}$ are the random arrival times for each counting process (and $t_{i,j}$ for observed arrivals). If for each $i = 1, \dots, m$ then $N_i(\cdot)$ has conditional intensity of the form

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^m \int_{-\infty}^t \mu_j(t-u) dN_j(u) \quad (12)$$

for some $\lambda_i > 0$ and $\mu_i : (0, \infty) \rightarrow [0, \infty)$, then \mathbf{N} is called a *mutually exciting Hawkes process*.

When the excitation functions are set to be exponentially decaying, (12) can be written as

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^m \int_{-\infty}^t \alpha_{i,j} e^{-\beta_{i,j}(t-s)} dN_j(s) = \lambda_i + \sum_{j=1}^m \sum_{t_{j,k} < t} \alpha_{i,j} e^{-\beta_{i,j}(t-t_{j,k})} \quad (13)$$

for non-negative constants $\{\alpha_{i,j}, \beta_{i,j} : i, j = 1, \dots, m\}$.

Remark 8 There are models for HPs where the points themselves are multi-dimensional, for example, spatial HPs or temporo-spatial HPs [2]. One should not confuse mutually exciting HPs with these multi-dimensional HPs.

3.6 Financial applications

This section reviews primarily the work of Ait-Sahalia, et al. [28] and Filimonov and Sornette [6]. It assumes the reader is familiar with mathematical finance and the use of stochastic differential equations.

3.6.1 Financial contagion

We turn our attention to the moest recent applications of HPs. A major domain for self- and mutually-exciting processes is financial analysis. Frequently it is seen that large movements in a major stock market propagate in foreign markets as a process called *financial contagion*. Examples of this phenomenon are clearly visible in historical series of asset prices; Fig. 5 illustrates one such case.

The ‘Hawkes diffusion model’ introduced by [28] is an attempt to extend previous models of stock prices to include financial contagion. Modern models for stock prices are typically built upon the model popularised by [29] where the log returns on the stock follow geometric Brownian motion. Whilst this seminal paper was lauded by the economics community, the model inadequately captured the ‘fat tails’ of the return distribution and so was not commonly used by traders [30]. Merton [31] attempted to incorporate heavy tails by including a Poisson jump process to model booms and crashes in the stock returns; this model is often called Merton diffusion model. The Hawkes diffusion model extends this model by replacing the Poisson jump process with a mutually-exciting HP, so that crashes can self-excite and propagate in a market and between global markets.

The basic Hawkes diffusion model describes the log returns of m assets $\{X_1(\cdot), \dots, X_m(\cdot)\}$ where each asset $i = 1, \dots, m$ has associated expected return $\mu_i \in \mathbb{R}$, constant volatility $\sigma_i \in \mathbb{R}^+$, and standard Brownian motion $(W_i^X(t) : t \geq 0)$. The Brownian motions have constant correlation coefficients $\{\rho_{i,j} : i, j = 1, \dots, m\}$. Jumps are added by a self- and mutually-exciting HP (as per Definition 7 with some selection of constants $\alpha_{i,\cdot}$ and $\beta_{i,\cdot}$) with stochastic jump sizes $(Z_i(t) : t \geq 0)$. The asset dynamics are then assumed to satisfy

$$dX_i(t) = \mu_i dt + \sigma_i dW_i^X(t) + Z_i(t) dN_i(t).$$

The general Hawkes diffusion model replaces the constant volatilities with stochastic volatilities $\{V_1(\cdot), \dots, V_m(\cdot)\}$ specified by the Heston model. Each asset $i = 1, \dots, m$ has a: long-term mean volatility $\theta_i > 0$, rate of returning to this mean $\kappa_i > 0$, volatility of the volatility $\nu_i > 0$, and standard Brownian motion ($W_i^V(t) : t \geq 0$). Correlation between the $W_i^X(\cdot)$'s is optional, yet the effect would be dominated by the jump component. Then the full dynamics are captured by

$$\begin{aligned} dX_i(t) &= \mu_i dt + \sqrt{V_i(t)} dW_i^X(t) + Z_i(t) dN_i(t), \\ dV_i(t) &= \kappa_i(\theta_i - V_i(t)) dt + \nu_i \sqrt{V_i(t)} dW_i^V(t). \end{aligned}$$

However the added realism of the Hawkes diffusion model comes at a high price. The constant volatility model requires $5m + 3m^2$ parameters to be fit (assuming $Z_i(\cdot)$ is characterised by two parameters) and the stochastic volatility extension requires an extra $3m$ parameters (assuming $\forall i, j = 1, \dots, m$ that $\mathbb{E}[W_i(\cdot)^V W_j(\cdot)^V] = 0$). In [28] hypothesis tests reject the Merton diffusion model in favour of the Hawkes diffusion model, however there are no tests for overfitting the data (for example, Akaike or Bayesian information criterion comparisons). Remember that John Von Neumann (reputedly) claimed that “with four parameters I can fit an elephant” [32].

For computational necessity the authors made a number of simplifying assumptions to reduce the number of parameters to fit (such as the background intensity of crashes is the same for all markets). Even so, the Hawkes diffusion model was only able to be fitted for pairs of markets ($m = 2$) instead of for the globe as a whole. Since the model was calibrated to daily returns of market indices, historical data was easily available (for example, from Google or Yahoo! finance); care had to be taken to convert timezones and handle the different market opening and closing times. The parameter estimation method used by [28] was the generalised method of moments, however the theoretical moments derived satisfy long and convoluted equations.

3.6.2 Mid-price changes and high-frequency trading

A simpler system to model is a single stock's price over time, though there are many different prices to consider. For each stock one could use: the last transaction price, the best ask price, the best bid price, or the mid-price (defined as the average of best ask and best bid prices). The last transaction price includes inherent microstructure noise (for example, the bid–ask bounce), and the best ask and bid prices fail to represent the actions of both buyers and sellers in the market.

Filimonov and Sornette [6] model the mid-price changes over time as a HP. In particular they look at long-term trends of the (estimated) branching ratio. In this context, n represents the proportion of price moves that are not due to external market information but simply reactions to other market participants. This ratio can be seen as the quantification of the principle of economic reflexivity. The authors conclude that the branching ratio has increased dramatically from 30% in 1998 to 70% in 2007.

Later that year [33] critiqued the test procedure used in this analysis. Filimonov and Sornette [6] had worked with a dataset with timestamps accurate to a second, and this often led to multiple arrivals nominally at the same time (which is an impossible event for simple point processes). Fake precision was achieved by adding $\text{Unif}(0, 1)$ random fractions of seconds to all timestamps, a technique also used by [34]. Lorenzen found that this method added an element of smoothing to the data which gave it a better fit to the model than the actual millisecond precision data. The randomisation also introduced bias to the HP parameter estimates, particularly of α and β . Lorenzen formed a crude measure of high-frequency trading activity leading to an interesting correlation between this activity and n over the observed period.

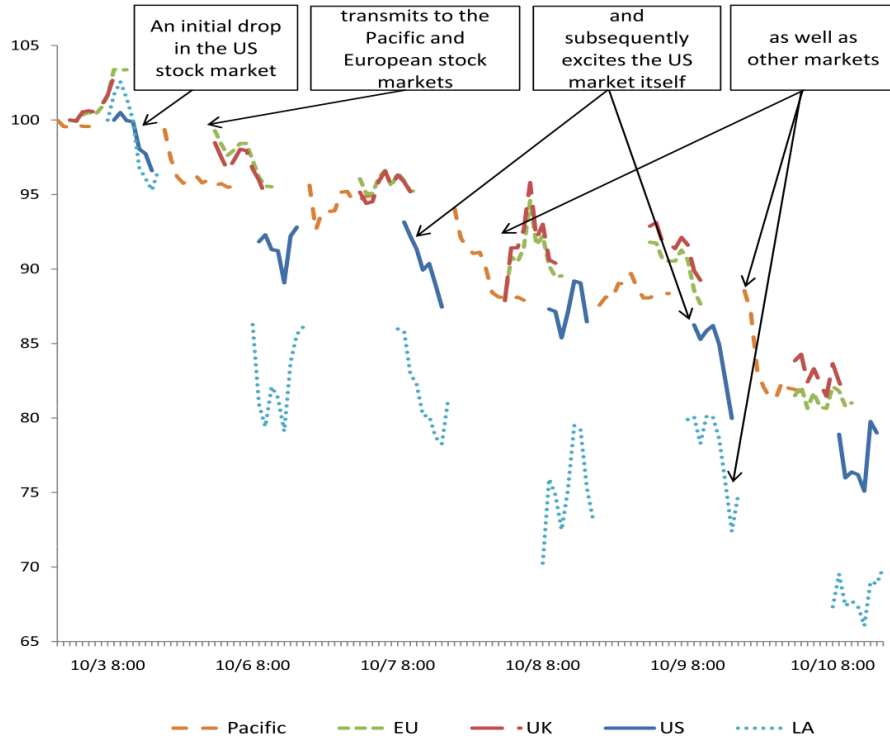


Fig. 5: Example of mutual excitation in global markets. This figure plots the cascade of declines in international equity markets experienced between October 3, 2008 and October 10, 2008 in the US; Latin America (LA); UK; Developed European countries (EU); and Developed countries in the Pacific. Data are hourly. The first observation of each price index series is normalised to 100 and the following observations are normalised by the same factor. Source: MSCI MXRT international equity indices on Bloomberg (reproduced from [28]).

Remark 9 Fortunately we have received comments from referees suggesting other very important works to consider, which we will briefly list here. The importance of Bowers [34] is highlighted, as is the series by Chavez-Demoulin et al. [35,36]. They point to the book by McNeil et al. [37] where a section is devoted to HP applications, and stress the relevance of the Parisian school in applying HPs to microstructure modelling, for example, the paper by Bacry et al. [38].

4 Parameter estimation

This section investigates the problem of generating parameters estimates $\hat{\theta} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ given some finite set of arrival times $\mathbf{t} = \{t_1, t_2, \dots, t_k\}$ presumed to be from a HP. For brevity, the notation here will omit the $\hat{\theta}$ and \mathbf{t} arguments from functions: $L = L(\hat{\theta}; \mathbf{t})$, $l = l(\hat{\theta}; \mathbf{t})$, $\lambda^*(t) = \lambda^*(t; \mathbf{t}, \hat{\theta})$, and $A(t) = A(t; \mathbf{t}, \hat{\theta})$. The estimators are tested over simulated data, for the sake of simplicity and lack of relevant data. Unfortunately this method bypasses the many significant challenges raised by real datasets, challenges that caused [39] to state that

“Our overall conclusion is that calibrating the Hawkes process is akin to an excursion within a minefield that requires expert and careful testing before any conclusive step can be taken.”

The method considered is *maximum likelihood estimation*, which begins by finding the *likelihood function*, and estimates the model parameters as the inputs which maximise this function.

4.1 Likelihood function derivation

Daley and Vere-Jones [9, Proposition 7.2.III] give the following result.

Theorem 3 (Hawkes process likelihood) *Let $N(\cdot)$ be a regular point process on $[0, T]$ for some finite positive T , and let t_1, \dots, t_k denote a realisation of $N(\cdot)$ over $[0, T]$. Then, the likelihood L of $N(\cdot)$ is expressible in the form*

$$L = \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp \left(- \int_0^T \lambda^*(u) \, du \right).$$

Proof First assume that the process is observed up to the time of the k th arrival. The joint density function from (1) is

$$L = f(t_1, t_2, \dots, t_k) = \prod_{i=1}^k f^*(t_i).$$

This function can be written in terms of the conditional intensity function. Rearrange (2) to find $f^*(t)$ in terms of $\lambda^*(t)$ (as per [40]):

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)} = \frac{\frac{d}{dt} F^*(t)}{1 - F^*(t)} = - \frac{d \log(1 - F^*(t))}{dt}.$$

Integrate both sides over the interval (t_k, t) :

$$- \int_{t_k}^t \lambda^*(u) \, du = \log(1 - F^*(t)) - \log(1 - F^*(t_k)).$$

The HP is a *simple* point process, meaning that multiple arrivals cannot occur at the same time. Hence $F^*(t_k) = 0$ as $T_{k+1} > t_k$, and so

$$- \int_{t_k}^t \lambda^*(u) \, du = \log(1 - F^*(t)). \quad (14)$$

Further rearranging yields

$$F^*(t) = 1 - \exp \left(- \int_{t_k}^t \lambda^*(u) \, du \right), \quad f^*(t) = \lambda^*(t) \exp \left(- \int_{t_k}^t \lambda^*(u) \, du \right). \quad (15)$$

Thus the likelihood becomes

$$L = \prod_{i=1}^k f^*(t_i) = \prod_{i=1}^k \lambda^*(t_i) \exp \left(- \int_{t_{i-1}}^{t_i} \lambda^*(u) \, du \right) = \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp \left(- \int_0^{t_k} \lambda^*(u) \, du \right). \quad (16)$$

Now suppose that the process is observed over some time period $[0, T] \supset [0, t_k]$. The likelihood will then include the probability of seeing no arrivals in the time interval $(t_k, T]$:

$$L = \left[\prod_{i=1}^k f^*(t_i) \right] (1 - F^*(T)).$$

Using the formulation of $F^*(t)$ from (15), then

$$L = \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp \left(- \int_0^T \lambda^*(u) du \right).$$

This completes the proof.

4.2 Simplifications for exponential decay

With the likelihood function from (16), the log-likelihood for the interval $[0, t_k]$ can be derived as

$$l = \sum_{i=1}^k \log(\lambda^*(t_i)) - \int_0^{t_k} \lambda^*(u) du = \sum_{i=1}^k \log(\lambda^*(t_i)) - \Lambda(t_k). \quad (17)$$

Note that the integral over $[0, t_k]$ can be broken up into the segments $[0, t_1]$, $(t_1, t_2]$, \dots , $(t_{k-1}, t_k]$, and therefore

$$\Lambda(t_k) = \int_0^{t_k} \lambda^*(u) du = \int_0^{t_1} \lambda^*(u) du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u) du.$$

This can be simplified in the case where $\lambda^*(\cdot)$ decays exponentially:

$$\begin{aligned} \Lambda(t_k) &= \int_0^{t_1} \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda + \sum_{t_j < u} \alpha e^{-\beta(u-t_j)} du \\ &= \lambda t_k + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^i e^{-\beta(u-t_j)} du \\ &= \lambda t_k + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^i \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)} du \\ &= \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \sum_{j=1}^i \left[e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)} \right]. \end{aligned}$$

Finally, many of the terms of this double summation cancel out leaving

$$\Lambda(t_k) = \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \left[e^{-\beta(t_k-t_i)} - e^{-\beta(t_i-t_i)} \right] = \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k-t_i)} - 1 \right]. \quad (18)$$

Note that here the final summand is unnecessary, though it is often included, see [33]. Substituting $\lambda^*(\cdot)$ and $\Lambda(\cdot)$ into (17) gives

$$l = \sum_{i=1}^k \log \left[\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} \right] - \lambda t_k + \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k-t_i)} - 1 \right]. \quad (19)$$

This direct approach is computationally infeasible as the first term's double summation entails $\mathcal{O}(k^2)$ complexity. Fortunately the similar structure of the inner summations allows l to be computed with $\mathcal{O}(k)$ complexity [41, 42]. For $i \in \{2, \dots, k\}$, let $A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)}$, so that

$$A(i) = e^{-\beta t_i + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_{i-1} + \beta t_j} = e^{-\beta(t_i - t_{i-1})} \left(1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1} - t_j)} \right) = e^{-\beta(t_i - t_{i-1})} (1 + A(i-1)). \quad (20)$$

With the added base case of $A(1) = 0$, l can be rewritten as

$$l = \sum_{i=1}^k \log(\lambda + \alpha A(i)) - \lambda t_k + \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k - t_i)} - 1 \right]. \quad (21)$$

Ozaki [8] also gives the partial derivatives and the Hessian for this log-likelihood function. Of particular note is that each derivative calculation can be achieved in order $\mathcal{O}(k)$ complexity when a recursive approach (similar to (20)) is taken [43].

Remark 10 The recursion implies that the joint process $(N(t), \lambda^*(t))$ is Markovian (see Remark 1.22 of [44]).

4.3 Discussion

Understanding of the maximum likelihood estimation method for the HP has changed significantly over time. The general form of the log-likelihood function (17) was known by Rubin [45]. It was applied to the HP by Ozaki [8] who derived (19) and the improved recursive form (21). Ozaki also found (as noted earlier) an efficient method for calculating the derivatives and the Hessian matrix. Consistency, asymptotic normality and efficiency of the estimator were proved by Ogata [41].

It is clear that the maximum likelihood estimation will usually be very effective for model fitting. However, [6] found that, for small samples, the estimator produces significant bias, encounters many local optima, and is highly sensitive to the selection of excitation function. Additionally, the $\mathcal{O}(k)$ complexity can render the method useless when samples become large; remember that any iterative optimisation routine would calculate the likelihood function perhaps thousands of times. The R 'hawkes' package thus implements this routine in C++ in an attempt to mitigate the performance issues.

This 'performance bottleneck' is largely the cause of the latest trend of using the generalised method of moments to perform parameter estimation. Da Fonseca and Zaatour [20] state that the procedure is "instantaneous" on their test sets. The method uses sample moments and the sample autocorrelation function which are smoothed via a (rather arbitrary) user-selected procedure.

5 Goodness of fit

This section outlines approaches to determining the appropriateness of a HPs model for point data, which is a critical link in their application.

5.1 Transformation to a Poisson process

Assessing the goodness of fit for some point data to a Hawkes model is an important practical consideration. In performing this assessment the point process' compensator is essential, as is the random time change theorem (here adapted from [46]):

Theorem 4 (Random time change theorem) *Say $\{t_1, t_2, \dots, t_k\}$ is a realisation over time $[0, T]$ from a point process with conditional intensity function $\lambda^*(\cdot)$. If $\lambda^*(\cdot)$ is positive over $[0, T]$ and $\Lambda(T) < \infty$ a.s. then the transformed points $\{\Lambda(t_1), \Lambda(t_2), \dots, \Lambda(t_k)\}$ form a Poisson process with unit rate.*

The random time change theorem is fundamental to the model fitting procedure called (*point process*) *residual analysis*. Original work [47] on residual analysis goes back to [48], [49], and [50]. Daley and Vere-Jones's Proposition 7.4.IV [9] rewords and extends the theorem as follows.

Theorem 5 (Residual analysis) *Consider an unbounded, increasing sequence of time points $\{t_1, t_2, \dots\}$ in the half-line $(0, \infty)$, and a monotonic, continuous compensator $\Lambda(\cdot)$ such that $\lim_{t \rightarrow \infty} \Lambda(t) = \infty$ a.s. The transformed sequence $\{t_1^*, t_2^*, \dots\} = \{\Lambda(t_1), \Lambda(t_2), \dots\}$, whose counting process is denoted $N^*(t)$, is a realisation of a unit rate Poisson process if and only if the original sequence $\{t_1, t_2, \dots\}$ is a realisation from the point process defined by $\Lambda(\cdot)$.*

Hence, equipped with a closed form of the compensator from (18), the quality of the statistical inference can be ascertained using standard fitness tests for Poisson processes. Fig. 6 shows a realisation of a HP and the corresponding transformed process. In Fig. 6 $\Lambda(t)$ appears identical to $N(t)$. They are actually slightly different ($\Lambda(\cdot)$ is continuous) however the similarity is expected due to Doob–Meyer decomposition of the compensator.

5.2 Tests for Poisson process

5.2.1 Basic tests

There are many procedures for testing whether a series of points form a Poisson process (see [15] for an extensive treatment). As a first test, one can run a hypothesis test to check $\sum_i \mathbb{1}_{\{t_i^* < t\}} \sim \text{Poi}(t)$. If this initial test succeeds then the interarrival times,

$$\{\tau_1, \tau_2, \tau_3, \dots\} = \{t_1^*, t_2^* - t_1^*, t_3^* - t_2^*, \dots\},$$

should be tested to ensure $\tau_i \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(1)$. A qualitative approach is to create a quantile–quantile (Q–Q) plot for τ_i using the exponential distribution (see for example Fig. 7a). Otherwise a quantitative alternative is to run Kolmogorov–Smirnov (or perhaps Anderson–Darling) tests.

5.2.2 Test for independence

The next test, after confirming there is reason to believe that the τ_i are exponentially distributed, is to check their independence. This can be done by looking for autocorrelation in the τ_i sequence. Obviously zero autocorrelation does not imply independence, but a non-zero amount would certainly imply a non-Poisson model. A visual examination can be conducted by plotting the points (U_{i+1}, U_i) . If there are noticeable patterns then the τ_i are autocorrelated. Otherwise the points should look evenly scattered; see for example Fig. 7b. Quantitative extensions exist; for example see Section 3.3.3 of [51], or serial correlation tests in [52].

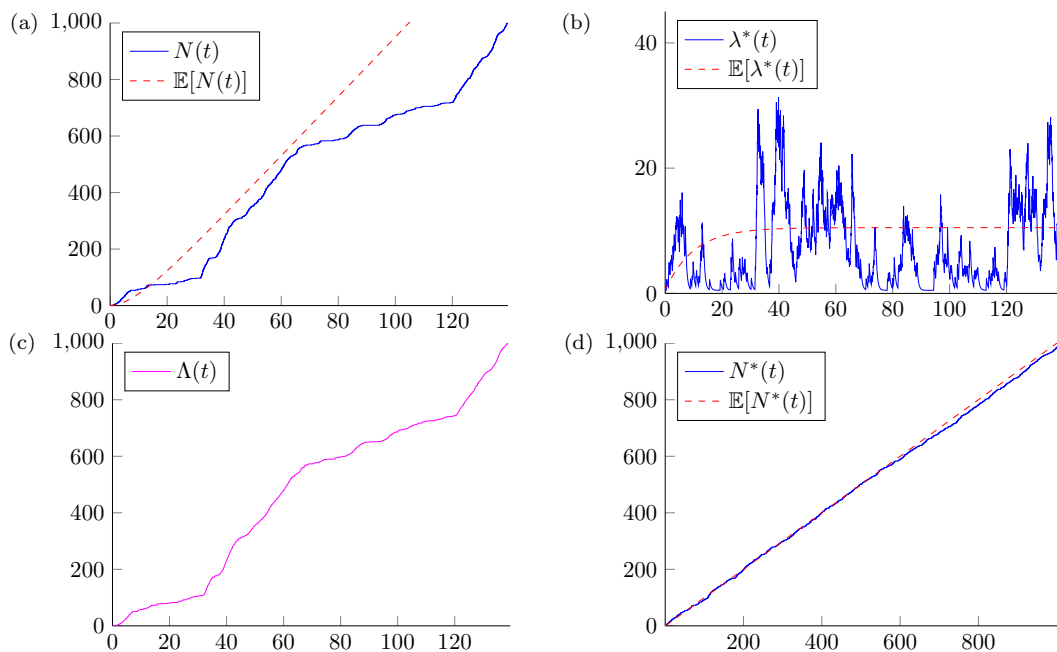


Fig. 6: An example of using the random time change theorem to transform a Hawkes process into a unit rate Poisson process. (a) A Hawkes process $N(t)$ with $(\lambda, \alpha, \beta) = (0.5, 2, 2.1)$, with the associated (b) conditional intensity function and (c) compensator. (d) The transformed process $N^*(t)$, where $t_i^* = \Lambda(t_i)$.

5.2.3 Lewis test

A statistical test with more power is the Lewis test as described by [53]. Firstly, it relies on the fact that if $\{t_1^*, t_2^*, \dots, t_N^*\}$ are arrival times for a unit rate Poisson process then $\{t_1^*/t_N^*, t_2^*/t_N^*, \dots, t_{N-1}^*/t_N^*\}$ are distributed as the order statistics of a uniform $[0, 1]$ random sample. This observation is called conditional uniformity, and forms the basis for a test itself. Lewis' test relies on applying Durbin's modification (introduced in [54] with a widely applicable treatment by [55]).

5.2.4 Brownian motion approximation test

An approximate test for Poissonity can be constructed by using the Brownian motion approximation to the Poisson process. This is to say, the observed times are transformed to be (approximately) Brownian motion, and then known properties of Brownian motion sample paths can be used to accept or reject the original sample.

The motivation for this line of enquiry comes from Algorithm 7.4.V of [9], which is described as an “approximate Kolmogorov–Smirnov-type test”. Unfortunately, a typographical error causes the algorithm (as printed) to produce incorrect answers for various significance levels. An alternative test based on the Brownian motion approximation is proposed here.

Say that $N(t)$ is a Poisson process of rate T . Define $M(t) = (N(t) - tT)/\sqrt{T}$ for $t \in [0, 1]$. Donsker's invariance principle implies that, as $T \rightarrow \infty$, $(M(t) : t \in [0, 1])$ converges in distribution to standard

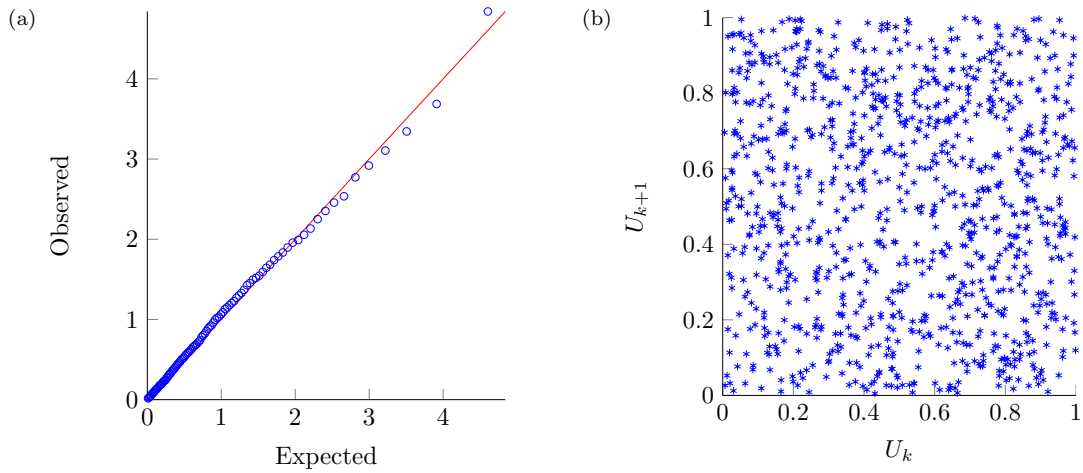


Fig. 7: (a) Q–Q testing for i.i.d. Exp(1) interarrival times. (b) A qualitative autocorrelation test. The U_k values are defined as $U_k = F(t_k^* - t_{k-1}^*) = 1 - e^{-(t_k^* - t_{k-1}^*)}$.

Brownian motion ($B(t) : t \in [0, 1]$). Fig. 8 shows example realisations of $M(t)$ for various T that, at least qualitatively, are reasonable approximations to standard Brownian motion.

An alternative test is to utilise the first arcsine law for Brownian motion, which states that the random time $M^* \in [0, 1]$, given by

$$M^* = \arg \max_{s \in [0, 1]} B(s),$$

is arcsine distributed (that is, $M^* \sim \text{Beta}(1/2, 1/2)$). Therefore the test takes a sequence of arrivals observed over $[0, T]$ and:

1. transforms the arrivals to $\{t_1^*/T, t_2^*/T, \dots, t_k^*/T\}$ which should be a Poisson process with rate T over $[0, 1]$,
2. constructs the Brownian motion approximation $M(t)$ as above, finds the maximiser M^* , and
3. accepts the ‘unit-rate Poisson process’ hypothesis if M^* lies within the $(\alpha/2, 1 - \alpha/2)$ quantiles of the Beta(1/2, 1/2) distribution; otherwise it is rejected.

As a final note, many other tests can be performed based on other properties of Brownian motion. For example, the test could be based simply on noting that $M(1) \sim N(0, 1)$, and thus accepts if $M(1) \in [Z_{\alpha/2}, Z_{1-\alpha/2}]$ and rejects otherwise.

6 Simulation methods

Simulation is an increasingly indispensable tool in probability modelling. Here we give details of three fundamental approaches to producing realisations of HPs.

6.1 Transformation methods

For general point processes a simulation algorithm is suggested by the converse of the random time change theorem (given in Section 5.1). In essence, a unit rate Poisson process $\{t_1^*, t_2^*, \dots\}$ is transformed

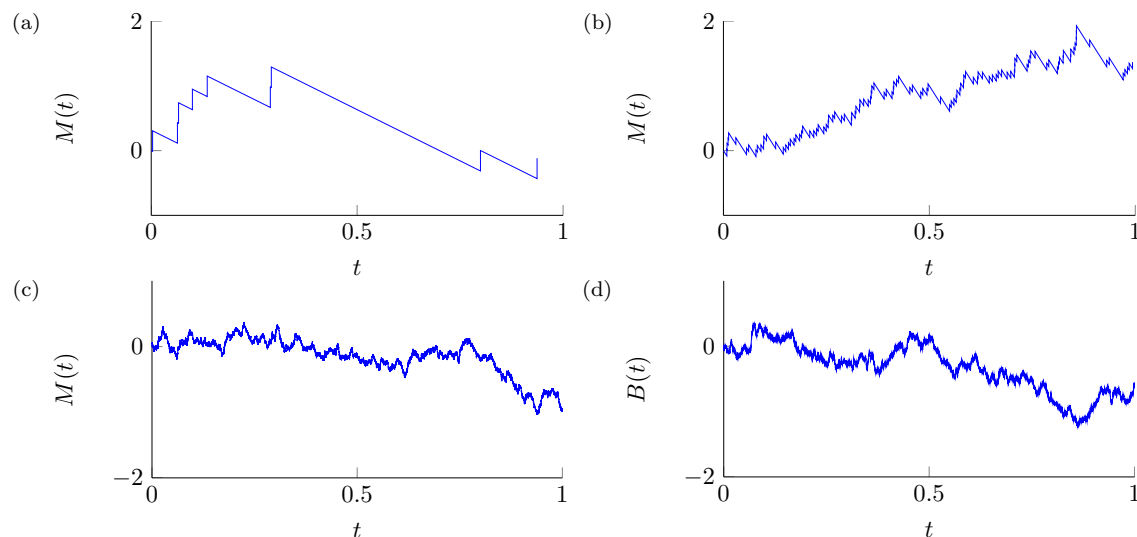


Fig. 8: Realisations of Poisson process approximations to Brownian motion. Plots (a)–(c) use the observed windows of $T = 10$, 100 , and $10,000$, respectively. Plot (d) is a direct simulation of Brownian motion for comparison.

by the inverse compensator $\Lambda(\cdot)^{-1}$ into any general point process defined by that compensator. The method, sometimes called the *inverse compensator method*, iteratively solves the equations

$$t_1^* = \int_0^{t_1} \lambda^*(s) ds, \quad t_{k+1}^* - t_k^* = \int_{t_k}^{t_{k+1}^*} \lambda^*(s) ds$$

for $\{t_1, t_2, \dots\}$, the desired point process (see [56] and Algorithm 7.4.III of [9]).

For HPs the algorithm was first suggested by Ozaki [8], but did not state explicitly any relation to time changes. It instead focused on (14),

$$\int_{t_k}^t \lambda^*(u) du = -\log(1 - F^*(t)),$$

which relates the conditional c.d.f. of the next arrival to the previous history of arrivals $\{t_1, t_2, \dots, t_k\}$ and the specified $\lambda^*(t)$. This relation means the next arrival time T_{k+1} can easily be generated by the inverse transform method, that is draw $U \sim \text{Unif}[0, 1]$ then t_{k+1} is found by solving

$$\int_{t_k}^{t_{k+1}} \lambda^*(u) du = -\log(U). \quad (22)$$

For an exponentially decaying intensity the equation becomes

$$\log(U) + \lambda(t_{k+1} - t_k) - \frac{\alpha}{\beta} \left(\sum_{i=1}^k e^{\beta(t_{k-1} - t_i)} - \sum_{i=1}^k e^{-\beta(t_k - t_i)} \right) = 0.$$

Solving for t_{k+1} can be achieved in linear time using the recursion of (20). However if a different excitation function is used then (22) must be solved numerically, for example using Newton's method [43], which entails a significant computational effort.

6.2 Ogata's modified thinning algorithm

HP generation is a similar problem to inhomogeneous Poisson process generation. The standard way to generate an inhomogeneous Poisson process driven by intensity function $\lambda(\cdot)$ is via thinning. Formally the process is described by Algorithm 1 [57]. The intuition is to generate a 'faster' homogeneous Poisson process, and remove points probabilistically so that the remaining points satisfy the time-varying intensity $\lambda(\cdot)$. The first process' rate M cannot be less than $\lambda(\cdot)$ over $[0, T]$.

A similar approach can be used for the HP, called *Ogata's modified thinning algorithm* [43,44]. The conditional intensity $\lambda^*(\cdot)$ does not have an a.s. asymptotic upper bound, however it is common for the intensity to be non-increasing in periods without any arrivals. This implies that for $t \in (T_i, T_{i+1}]$, $\lambda^*(t) \leq \lambda^*(T_i^+)$ (that is, the time just after T_i , when that arrival has been registered). So the M value can be updated during each simulation. Algorithm 2 describes the process and Fig. 9 shows an example of each thinning procedure.

Algorithm 1 Generate an inhomogeneous Poisson process by thinning.

```

1: procedure POISSONBYTHINNING( $T, \lambda(\cdot), M$ )
2:   require:  $\lambda(\cdot) \leq M$  on  $[0, T]$ 
3:    $P \leftarrow []$ ,  $t \leftarrow 0$ .
4:   while  $t < T$  do
5:      $E \leftarrow \text{Exp}(M)$ .
6:      $t \leftarrow t + E$ .
7:      $U \leftarrow \text{Unif}(0, M)$ .
8:     if  $t < T$  and  $U \leq \lambda(t)$  then
9:        $P \leftarrow [P, t]$ .
10:    end if
11:  end while
12:  return  $P$ 
13: end procedure

```

Algorithm 2 Generate a Hawkes process by thinning.

```

1: procedure HAWKESBYTHINNING( $T, \lambda^*(\cdot)$ )
2:   require:  $\lambda^*(\cdot)$  non-increasing in periods of no arrivals.
3:    $\varepsilon \leftarrow 10^{-10}$  (some tiny value  $> 0$ ).
4:    $P \leftarrow []$ ,  $t \leftarrow 0$ .
5:   while  $t < T$  do
6:     Find new upper bound:
7:      $M \leftarrow \lambda^*(t + \varepsilon)$ .
8:     Generate next candidate point:
9:      $E \leftarrow \text{Exp}(M)$ ,  $t \leftarrow t + E$ .
10:    Keep it with some probability:
11:     $U \leftarrow \text{Unif}(0, M)$ .
12:    if  $t < T$  and  $U \leq \lambda^*(t)$  then
13:       $P \leftarrow [P, t]$ .
14:    end if
15:  end while
16:  return  $P$ 
17: end procedure

```

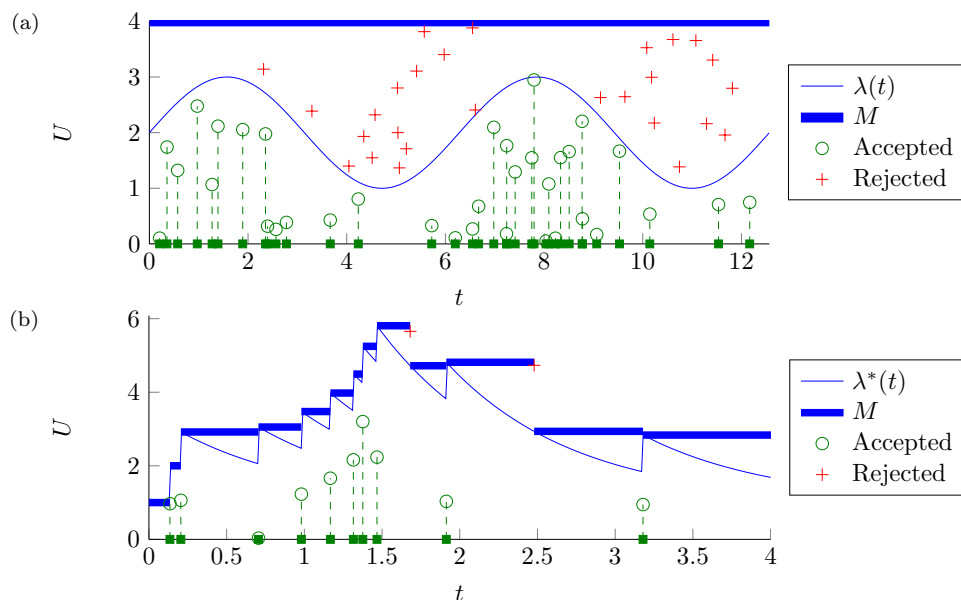


Fig. 9: Processes generated by thinning. (a) A Poisson process with intensity $\lambda(t) = 2 + \sin(t)$, bounded above by $M = 4$. (b) A Hawkes process with $(\lambda, \alpha, \beta) = (1, 1, 1.1)$. Each (t, U) point describes a suggested arrival at time t whose U value is given in Algorithm 1 and Algorithm 2. Plus signs indicate rejected points, circles accepted, and green squares the resulting point processes.

6.3 Superposition of Poisson processes

The immigration–birth representation gives rise to a simple simulation procedure: generate the immigrant arrivals, then generate the descendants for each immigrant. Algorithm 3 describes the procedure in full, with Fig. 10 showing an example realisation.

Immigrants form a homogeneous Poisson process of rate λ , so over an interval $[0, T]$ the number of immigrants is $\text{Poi}(\lambda T)$ distributed. Conditional on knowing that there are k immigrants, their arrival times C_1, C_2, \dots, C_k are distributed as the order statistics of i.i.d. $\text{Unif}[0, T]$ random variables.

Each immigrant’s descendants form an inhomogeneous Poisson process. The i th immigrant’s descendants arrive with intensity $\mu(t - C_i)$ for $t > C_i$. Denote D_i to be the number of descendants of immigrant i , then $\mathbb{E}[D_i] = \int_0^\infty \mu(s) ds = n$, and hence $D_i \stackrel{\text{i.i.d.}}{\sim} \text{Poi}(n)$. Say that the descendants of the i th immigrant arrive at times $(C_i + E_1, C_i + E_2, \dots, C_i + E_{D_i})$. Conditional on knowing D_i , the E_j are i.i.d. random variables distributed with p.d.f. $\mu(\cdot)/n$. For exponentially decaying intensities, this simplifies to $E_j \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\beta)$.

6.4 Other methods

This section’s contents are by no means a complete compilation of simulation techniques available for HPs. Dassios and Zhao [58] and Møller and Rasmussen [59] give alternatives to the methods listed above. Also not discussed is the problem of simulating mutually-exciting HPs, however there are many

Algorithm 3 Generate a Hawkes process by clusters.

```

1: procedure HAWKESBYCLUSTERS( $T, \lambda, \alpha, \beta$ )
2:    $P \leftarrow \{\}$ .
3:   Immigrants:
4:      $k \leftarrow \text{Poi}(\lambda T)$ 
5:      $C_1, C_2, \dots, C_k \stackrel{\text{i.i.d.}}{\leftarrow} \text{Unif}(0, T)$ .
6:   Descendants:
7:      $D_1, D_2, \dots, D_k \stackrel{\text{i.i.d.}}{\leftarrow} \text{Poi}(\alpha/\beta)$ .
8:   for  $i \leftarrow 1$  to  $k$  do
9:     if  $D_i > 0$  then
10:        $E_1, E_2, \dots, E_{D_i} \stackrel{\text{i.i.d.}}{\leftarrow} \text{Exp}(\beta)$ .
11:        $P \leftarrow P \cup \{C_i + E_1, \dots, C_i + E_{D_i}\}$ .
12:     end if
13:   end for
14:   Remove descendants outside  $[0, T]$ :
15:    $P \leftarrow \{P_i : P_i \in P, P_i \leq T\}$ .
16:   Add in immigrants and sort:
17:    $P \leftarrow \text{Sort}(P \cup \{C_1, C_2, \dots, C_k\})$ .
18:   return  $P$ 
19: end procedure

```

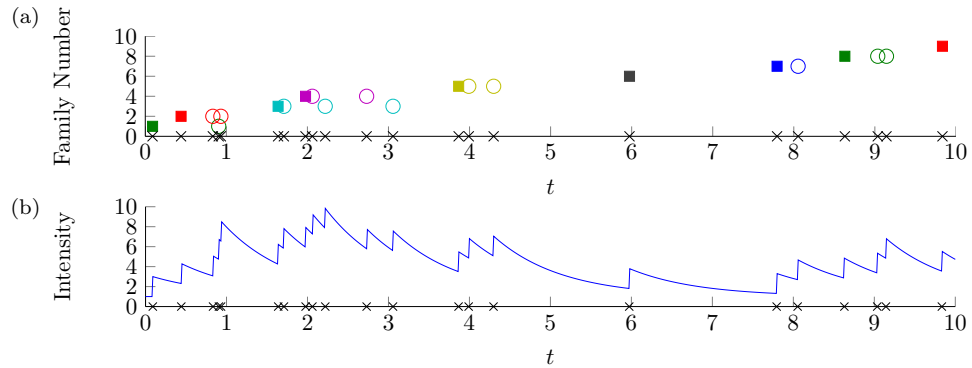


Fig. 10: A Hawkes Poisson process generated by clusters. Plot (a) shows the points generated by the immigrant–birth representation; it can be seen as a sequence of vertically stacked ‘family trees’. The immigrant points are plotted as squares, following circles of the same height and color are their offspring. The intensity function, with $(\lambda, \alpha, \beta) = (1, 2, 1.2)$, is plotted in (b). The resulting Hawkes process arrivals are drawn as crosses on the axis.

free software packages that provide this functionality. Fig. 11 shows an example realisation generated using the R package ‘hawkes’ (see also Roger D. Peng’s related R package ‘ptproc’).

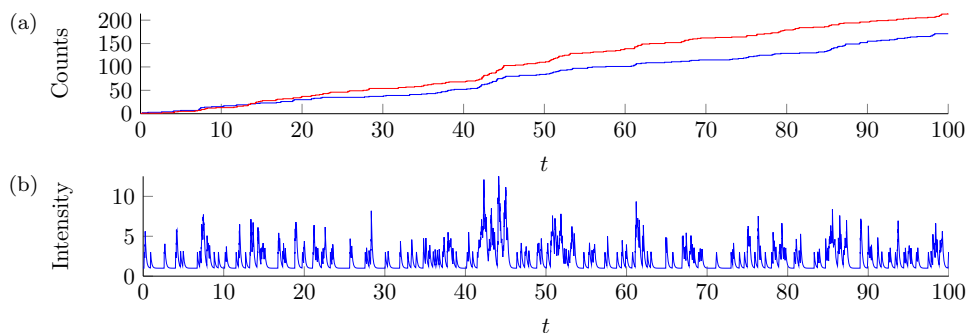


Fig. 11: A pair of mutually exciting Hawkes processes. (a) The two counting processes $N_1(t)$ and $N_2(t)$ with parameters: $\lambda_1 = \lambda_2 = 1$, $\alpha_{1,1} = \alpha_{1,2} = \alpha_{2,1} = \alpha_{2,2} = 2$, $\beta_{1,1} = \beta_{1,2} = \beta_{2,1} = \beta_{2,2} = 8$. (b) The processes' realised intensities (note that $\lambda_1^*(t) = \lambda_2^*(t)$ so only one is plotted).

7 Conclusion

HPs are fundamentally fascinating models of reality. Many of the standard probability models are Markovian and hence disregard the history of the process. The HP is structured around the premise that the history matters, which partly explains why they appear in such a broad range of applications.

If the exponentially decaying intensity can be utilised, then the joint process $(N(\cdot), \lambda^*(\cdot))$ satisfies the Markov condition, and both processes exhibit amazing analytical tractability. Explosion is avoided by ensuring that $\alpha < \beta$. The covariance density is a simple symmetric scaled exponential curve, and the power spectral density is a shifted scaled Cauchy p.d.f. The likelihood function and the compensator are elegant, and efficient to calculate using recursive structures. Exact simulation algorithms can generate this type of HP with optimal efficiency. Many aspects of the HP remain obtainable with any selection of excitation function; for example, the random time change theorem completely solves the problem of testing the goodness of a model's fit.

The use of HPs in finance appears itself to have been a self-exciting process. Aït-Sahalia et al. [28], Filimonov and Sornette [6], and Da Fonseca and Zaatour [20] formed the primary sources for the financial part of Section 3; these papers are surprisingly recent (given the fact that the model was introduced in 1971) and are representative of a current surge in HP research.

A Additional proof details

In this appendix, we collect additional detail elided from the proof of Theorem 2.

A.1 Supplementary to Theorem 2 (part one)

$$\begin{aligned}
R(\tau) &= \mathbb{E} \left[\frac{dN(t)}{dt} \left(\lambda + \int_{-\infty}^{t+\tau} \mu(t+\tau-s) dN(s) \right) \right] - \bar{\lambda}^{*2} \\
&= \lambda \mathbb{E} \left[\frac{dN(t)}{dt} \right] + \mathbb{E} \left[\frac{dN(t)}{dt} \left(\int_{-\infty}^{t+\tau} \mu(t+\tau-s) dN(s) \right) \right] - \bar{\lambda}^{*2} \\
&= \lambda \bar{\lambda}^* + \mathbb{E} \left[\frac{dN(t)}{dt} \int_{-\infty}^{t+\tau} \mu(t+\tau-s) dN(s) \right] - \bar{\lambda}^{*2}.
\end{aligned}$$

Introduce a change of variable $v = s - t$ and multiply by $\frac{dv}{dt}$:

$$R(\tau) = \lambda \bar{\lambda}^* + \mathbb{E} \left[\int_{-\infty}^{\tau} \mu(\tau-v) \frac{dN(t)}{dt} \frac{dN(t+v)}{dv} dv \right] - \bar{\lambda}^{*2} = \lambda \bar{\lambda}^* + \int_{-\infty}^{\tau} \mu(\tau-v) \mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t+v)}{dv} \right] dv - \bar{\lambda}^{*2}.$$

The expectation is (a shifted) $R^{(c)}(v)$. Substitute that and (8) in:

$$\begin{aligned}
R(\tau) &= \lambda \bar{\lambda}^* + \int_{-\infty}^{\tau} \mu(\tau-v) (R^{(c)}(v) + \bar{\lambda}^{*2}) dv - \bar{\lambda}^{*2} \\
&= \lambda \bar{\lambda}^* + \int_{-\infty}^{\tau} \mu(\tau-v) (\bar{\lambda}^* \delta(v) + R(v)) dv + \bar{\lambda}^{*2} \int_{-\infty}^{\tau} \mu(\tau-v) dv - \bar{\lambda}^{*2} \\
&= \lambda \bar{\lambda}^* + \bar{\lambda}^* \mu(\tau) + \int_{-\infty}^{\tau} \mu(\tau-v) R(v) dv + n \bar{\lambda}^{*2} - \bar{\lambda}^{*2} \\
&= \bar{\lambda}^* \mu(\tau) + \int_{-\infty}^{\tau} \mu(\tau-v) R(v) dv + \bar{\lambda}^* (\lambda - (1-n) \bar{\lambda}^*).
\end{aligned}$$

Using (7) yields

$$\begin{aligned}
\lambda - (1-n) \bar{\lambda}^* &= \lambda - (1-n) \frac{\lambda}{1-n} = 0. \\
\therefore R(\tau) &= \bar{\lambda}^* \mu(\tau) + \int_{-\infty}^{\tau} \mu(\tau-v) R(v) dv.
\end{aligned}$$

A.2 Supplementary to Theorem 2 (part two)

Split the right-hand side of the equation into three functions g_1, g_2 , and g_3 :

$$R(\tau) = \underbrace{\bar{\lambda}^* \mu(\tau)}_{g_1(\tau)} + \underbrace{\int_0^{\infty} \mu(\tau+v) R(v) dv}_{g_2(\tau)} + \underbrace{\int_0^{\tau} \mu(\tau-v) R(v) dv}_{g_3(\tau)}. \quad (23)$$

Taking the Laplace transform of each term gives

$$\begin{aligned}
\mathcal{L} \{g_1(\tau)\} (s) &= \int_0^s e^{-s\tau} \bar{\lambda}^* \alpha e^{-\beta\tau} d\tau = \frac{\alpha}{s+\beta} \bar{\lambda}^*, \\
\mathcal{L} \{g_2(\tau)\} (s) &= \int_0^{\infty} e^{-s\tau} \int_0^{\infty} \alpha e^{-\beta(\tau+v)} R(v) dv d\tau \\
&= \alpha \int_0^{\infty} e^{-\beta v} R(v) \int_0^{\infty} e^{-\tau(s+\beta)} d\tau dv \\
&= \frac{\alpha}{s+\beta} \int_0^{\infty} e^{-\beta v} R(v) dv \\
&= \frac{\alpha}{s+\beta} \mathcal{L} \{R\} (\beta),
\end{aligned}$$

and

$$\mathcal{L}\{g_3(\tau)\}(s) = \mathcal{L}\{\mu(\tau)\}(s)\mathcal{L}\{R(\tau)\}(s) = \frac{\alpha}{s+\beta}\mathcal{L}\{R(\tau)\}(s).$$

Therefore the Laplace transform of (23)

$$\mathcal{L}\{R(\tau)\}(s) = \frac{\alpha}{s+\beta}\left(\bar{\lambda}^* + \mathcal{L}\{R(\tau)\}(\beta) + \mathcal{L}\{R(\tau)\}(s)\right). \quad (24)$$

Substituting $s = \beta$ and rearranging gives that

$$\mathcal{L}\{R(\tau)\}(\beta) = \frac{\alpha\bar{\lambda}^*}{2(\beta-\alpha)}. \quad (25)$$

So substituting the value of $\mathcal{L}\{R(\tau)\}(\beta)$ into (24) means

$$\begin{aligned} \mathcal{L}\{R(\tau)\}(s) &= \frac{\alpha}{s+\beta}\left(\bar{\lambda}^* + \frac{\alpha\bar{\lambda}^*}{2(\beta-\alpha)} + \mathcal{L}\{R(\tau)\}(s)\right) \\ \Rightarrow \mathcal{L}\{R(\tau)\}(s) &= \frac{\frac{\alpha}{s+\beta}\left(\bar{\lambda}^* + \frac{\alpha\bar{\lambda}^*}{2(\beta-\alpha)}\right)}{1 - \frac{\alpha}{s+\beta}} = \frac{\alpha\bar{\lambda}^*(2\beta-\alpha)}{2(\beta-\alpha)(s+\beta-\alpha)}. \end{aligned}$$

References

1. Y. Ogata, *Journal of the American Statistical Association* **83**(401), 9 (1988)
2. G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, G.E. Tita, *Journal of the American Statistical Association* **106**(493), 100 (2011)
3. S. Azizpour, K. Giesecke, G. Schwenkler. Exploring the sources of default clustering. <http://web.stanford.edu/dept/MSandE/cgi-bin/people/faculty/giesecke/pdfs/exploring.pdf> (2010). Working paper, retrieved on 10 Feb 2015
4. A.G. Hawkes, *Biometrika* **58**(1), 83 (1971)
5. D. Lando, M.S. Nielsen, *Journal of Financial Intermediation* **19**(3), 355 (2010)
6. V. Filimonov, D. Sornette, *Physical Review E* **85**(5), 056108 (2012)
7. L. Carstensen, *Hawkes processes and combinatorial transcriptional regulation*. Ph.D. thesis, University of Copenhagen (2010)
8. T. Ozaki, *Annals of the Institute of Statistical Mathematics* **31**(1), 145 (1979)
9. D. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods* (Springer, 2003)
10. D.R. Cox, *Journal of the Royal Statistical Society. Series B (Methodological)* **17**(2), 129 (1955)
11. M.S. Bartlett, *Journal of the Royal Statistical Society. Series B (Methodological)* **25**(2), 264 (1963)
12. M.S. Bartlett, *Sankhyā: The Indian Journal of Statistics, Series A* **25**(3), 245 (1963)
13. M.S. Bartlett, *Biometrika* **51**(3/4), 299 (1964)
14. P.A. Lewis, *Journal of the Royal Statistical Society. Series B (Methodological)* **26**(3), 398 (1964)
15. D.R. Cox, P.A. Lewis, *The Statistical Analysis of Series of Events* (Monographs on Applied Probability and Statistics, London: Chapman and Hall, 1966)
16. A.G. Hawkes, *Journal of the Royal Statistical Society. Series B (Methodological)* **33**(3), 438 (1971)
17. N. Hautsch, *Econometrics of Financial High-Frequency Data* (Springer, 2011)
18. Y. Ogata, *Pure and Applied Geophysics* **155**(2/4), 471 (1999)
19. V. Chatalbashev, Y. Liang, A. Officer, N. Trichakis. Exciting times for trade arrivals. <http://users.iems.northwestern.edu/~armbruster/2007msande444/report1a.pdf> (2007). Stanford University MS&E 444 group project submission, retrieved on 10 Feb 2015
20. J. Da Fonseca, R. Zaatour, *Journal of Futures Markets* **34**(6), 548 (2014)
21. A.G. Hawkes, D. Oakes, *Journal of Applied Probability* **11**(3), 493 (1974)
22. J.G. Rasmussen, *Methodology and Computing in Applied Probability* **15**(3), 623 (2013)
23. G. Grimmett, D. Stirzaker, *Probability and Random Processes* (Oxford University Press, 2001)
24. S. Asmussen, *Applied Probability and Queues*, 2nd edn. Applications of Mathematics: Stochastic Modelling and Applied Probability (Springer, 2003)
25. P.A. Lewis, *Journal of Sound and Vibration* **12**(3), 353 (1970)
26. P. Brémaud, L. Massoulié, *The Annals of Probability* **24**(3), 1563 (1996)

27. L. Zhu, *Journal of Applied Probability* **50**(3), 760 (2013)
28. Y. Ait-Sahalia, J. Cacho-Diaz, R.J. Laeven, Modeling financial contagion using mutually exciting jump processes. Tech. Rep. 15850, National Bureau of Economic Research, USA (2010)
29. F. Black, M. Scholes, *The Journal of Political Economy* **81**(3), 637 (1973)
30. E.G. Haug, N.N. Taleb, *Wilmott Magazine* **71** (2014)
31. R.C. Merton, *Journal of Financial Economics* **3**(1), 125 (1976)
32. F. Dyson, *Nature* **427**(6972), 297 (2004)
33. F. Lorenzen, Analysis of order clustering using high frequency data: A point process approach. Ph.D. thesis, Swiss Federal Institute of Technology Zurich (ETH Zurich) (2012)
34. C.G. Bowsher, *Journal of Econometrics* **141**(2), 876 (2007)
35. V. Chavez-Demoulin, A.C. Davison, A.J. McNeil, *Quantitative Finance* **5**(2), 227 (2005)
36. V. Chavez-Demoulin, J. McGill, *Journal of Banking & Finance* **36**(12), 3415 (2012)
37. A.J. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools: Concepts, Techniques and Tools* (Princeton university press, 2015)
38. E. Bacry, S. Delattre, M. Hoffmann, J.F. Muzy, *Quantitative Finance* **13**(1), 65 (2013)
39. V. Filimonov, D. Sornette, Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. Tech. Rep. 13-60, Swiss Finance Institute Research Paper (2013)
40. J.G. Rasmussen. Temporal point processes: the conditional intensity function. <http://people.math.aau.dk/~jgr/teaching/punktproc11/tpp.pdf> (2009). Course notes for 'rumlige punktprocesser' (spatial point processes), retrieved on 10 Feb 2015
41. Y. Ogata, *Annals of the Institute of Statistical Mathematics* **30**(1), 243 (1978)
42. S. Crowley. Point process models for multivariate high-frequency irregularly spaced data. <http://vixra.org/pdf/1211.0094v6.pdf> (2013). Working paper, retrieved on 10 Feb 2015
43. Y. Ogata, *Information Theory, IEEE Transactions on* **27**(1), 23 (1981)
44. T.J. Liniger, Multivariate Hawkes processes. Ph.D. thesis, Swiss Federal Institute of Technology Zurich (ETH Zurich) (2009)
45. I. Rubin, *Information Theory, IEEE Transactions on* **18**(5), 547 (1972)
46. E. Brown, R. Barbieri, V. Ventura, R. Kass, L. Frank, *Neural computation* **14**(2), 325 (2002)
47. P. Embrechts, T. Liniger, L. Lin, *Journal of Applied Probability* **48A**, 367 (2011). Special volume: a Festschrift for Søren Asmussen
48. P.A. Meyer, in *Séminaire de Probabilités V Université de Strasbourg* (Springer, 1971), pp. 191–195
49. F. Papangelou, *Transactions of the American Mathematical Society* **165**, 483 (1972)
50. S. Watanabe, *Japan. J. Math* **34**(53-70), 82 (1964)
51. D.E. Knuth, *Art of Computer Programming, Volume 2: Seminumerical Algorithms, The* (Addison-Wesley Professional, 2014)
52. D. Kroese, T. Taimre, Z.I. Botev, *Handbook of Monte Carlo methods* (Wiley, 2011)
53. S.H. Kim, W. Whitt, The power of alternative Kolmogorov–Smirnov tests based on transformations of the data (2013). Submitted to *ACM Transactions on Modeling and Computer Simulation*, Special Issue in Honor of Don Iglehart (Issue 25.4)
54. J. Durbin, *Biometrika* **53**(3/4), 41 (1961)
55. P.A. Lewis, *Biometrika* **52**(1/2), 67 (1965)
56. K. Giesecke, P. Tomecek. Dependent events and changes of time. <http://web.stanford.edu/dept/MSandE/cgi-bin/people/faculty/giesecke/pdfs/dect.pdf> (2005). Working paper, retrieved on 10 Feb 2015
57. P.A. Lewis, G.S. Shedler, *Naval Research Logistics Quarterly* **26**(3), 403 (1979)
58. A. Dassios, H. Zhao, *Electronic Communications in Probability* **18**(62) (2013)
59. J. Møller, J.G. Rasmussen, *Advances in Applied Probability* **37**(3), 629 (2005)