Check for updates

*Adv. Appl. Probab.* **58,** 1276–1317 (2026)
*doi:*10.1017/apr.2025.10049

# THE FUNCTIONAL DISCRETE-TIME APPROXIMATION OF MARKED HAWKES RISK PROCESSES

MAHMOUD KHABOU,* *Imperial College London*
LAURE COUTIN,** *Institut de Mathématiques de Toulouse*

**Abstract**

The marked Hawkes risk process is a compound point process where the occurrence and amplitude of past events impact the future. Since data in real life are acquired over a discrete time grid, we propose a strong discrete-time approximation of the continuous-time risk process obtained by embedding from the same Poisson measure. We then prove trajectorial convergence results in both fractional Sobolev spaces and the Skorokhod space, hence extending the theorems proven in Huang and Khabou ((2023). *Stoch. Process. Appl.* **161**, 201–241) and Kirchner ((2016). *Stoch. Process. Appl.* **126**(8), 2494–2525). We also provide upper bounds on the convergence speed with explicit dependence on the size of the discretization step, the time horizon, and the regularity of the kernel.

*Keywords:* Hawkes process; Euler scheme; Sobolev space; Skorokhod space; Poisson autoregression

2020 Mathematics Subject Classification: Primary 60G55; 65C30
Secondary 62L20; 62M10

## 1. Introduction

Initially introduced as a model for contagious events [19], continuous-time linear Hawkes processes found applications in many fields of study involving self-excitation or cross-excitation, such as portfolio credit risk [13], microstructure price dynamics [31], queuing networks [11], social media networks [32], and earthquakes [34]. The Hawkes process was then extended to the nonlinear setting in a seminal article by Brémaud and Massoulié [6], thus allowing a more general dependence on the past including self-inhibition, in opposition to the mere affine dependence allowed by the linear process. This generalization comes at a price: the nonlinear Hawkes process lacks the Galton–Watson framework, as well as closed formulae for its expected value and covariance [20]. Nonetheless, nonlinear Hawkes processes have found applications in fields where self-inhibition and cross-inhibition are crucial, such as neuroscience [30]. Because of the flexibility they confer, nonlinear Hawkes processes and their estimation recently became a popular field of research. For instance, Sulem *et al.* [37] provided

Received 17 September 2024; accepted 5 December 2025.
* Postal address: Imperial College London, 180 Queen's Gate, South Kensington, London, SW7 2AZ, UK. Email: m.khabou@imperial.ac.uk
** Postal address: Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse CEDEX, France. Email: coutin@math.univ-toulouse.fr

© The Author(s), 2026. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

guaranties for the Bayesian estimation of such processes. Similarly, in the Markov setting, we refer to [12] for the nonparametric estimation on neuronal networks.

As the amount of data involving bin counts (or count series) has been increasing, appropriate discrete-time count models have recently become a relevant object of research. In the context of autoregressive processes, a number of count series models have been proposed to capture the self-exciting (or self-inhibiting) aspects of a given dynamics. One of them is the integer autoregressive process of order $p$ (INAR($p$)) [1]. The generalization of this model to the infinity order (INAR($\infty$)) has been proven to be a discrete-time version of linear Hawkes processes (see [29] and the convergence results therein). Because of the way they are constructed, INAR time series are linear models capable of capturing only self-excitation.

Another prominent family of count series that has been extensively studied in the literature is Poisson autoregressions [14, 35]. Statistical guaranties such as strong consistency and asymptotic normality of maximum likelihood estimators have been proven for Poisson autoregressions, initially for the univariate case [15] and then generalized to the multivariate [17] and network [2] settings.

Beyond their simpler definition, Poisson autoregressions present another advantage compared with INAR series: they can be nonlinear and thus allow the modeling of self-inhibition in counts. Such nonlinear models have been studied in [16] in the Markovian setting and have been generalized to the case of infinite memory in [25].

The Markovian nonlinear Poisson autoregressions have been proven to be a discrete-time version of nonlinear Hawkes processes with Erlang kernels (that is, the product of a polynomial and an exponential) in the Skorokhod topology in [23]. This constituted a bridge between the literature on count series and the literature on Hawkes processes in the Markovian case.

In this article, we suggest a straightforward approach based on the intuitive discretization of a stochastic integral, further bridging the literature on Hawkes processes with the literature on count series in the general setting. This can be used, for instance, to derive more data-oriented estimation procedures for nonlinear Hawkes processes. The results proven in this article also constitute a building block for quantitative Euler schemes for the simulation of state-dependent Hawkes jump-diffusions [27].

To give an explicit illustration, we start by recalling the definition of a continuous-time Hawkes process. Let $N = (N_t)_{t\in[0,T]}$ be a point process observed on a time interval $[0, T]$ and measurable with respect to its canonical filtration $\mathcal{F}^N$. Assume that $N$ has an intensity; that is, a predictable process measuring the propensity of $N$ to jump in the near future, or informally

$$\lambda_t \mathrm{d}t = \mathbb{E}\left[\mathrm{d}N_t | \mathcal{F}^N_{t-}\right], \tag{1}$$

where $\mathrm{d}N_t = N_{t+\mathrm{d}t} - N_t$ takes the value 1 or 0, depending on the presence of a jump at time $t$. Obviously, the larger $\lambda_t$ is at a given time $t$, the more likely $N_t$ is to jump, and vice versa.

We say that $N$ is a Hawkes process of kernel $h$ and jump rate $\psi$ if the intensity takes the form

$$\lambda_t = \psi\left(\mu + \int_0^{t-} h(t-s)\mathrm{d}N_s\right), \quad t \in [0, T], \tag{2}$$

where the integral is taken in the Stieltjes sense. We take $\psi$ to be a positive Lipschitz function and $h$ to be integrable. Up to some standard stability hypothesis that will be stated in the next section, it is possible to build the Hawkes process on $\mathbb{R}_+$ and show that, like for the Poisson process, the random variable $N_T$ is of order $O(T)$ on average.

In this article, we work with a discrete-time model entirely based on the Riemann sum approximation of the integral (2). Indeed, given a discretization time step $\Delta > 0$, the infinitesimal increment $\mathrm{d}N_t$ can be seen as the number of events observed in the 'small' time interval

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$(n\Delta, (n+1)\Delta]$, where $n = \lfloor \frac{t}{\Delta} \rfloor$. Giving a discrete approximation of $N$ is then equivalent to giving a sequence of bin counts $(X_0, X_1, \dots, X_M)$, where $M = \lfloor \frac{T}{\Delta} \rfloor$. Knowing all of the past count values, we simulate the count series $X_n$ according to a Poisson distribution of parameter $\Delta \cdot l_n^\Delta$, where the expression for the (discrete) intensity is

$$l_n^\Delta = \psi \left( \sum_{k=1}^{n-1} h\left(\Delta \cdot (n-k)\right) X_k \right),$$

ensuring that it is predictable with respect to the filtration $(\sigma(X_0, \dots, X_k)_{k\in\mathbb{N}})$. We point out that given $l_n^\Delta$, $X_n$ is independent of $(X_0, \dots, X_{n-1})$.

The choice of the Poisson distribution $\mathcal{P}(\Delta \cdot l_n^\Delta)$ guarantees that $\Delta \cdot l_n^\Delta = \mathbb{E}[X_n | X_0, \dots, X_{n-1}]$ (which is the discrete equivalent of (1)) and is preferred to the more trivial choice of the uniform distribution $\mathcal{U}(\Delta \cdot l_n^\Delta)$ used in [36] for the following reasons:

(i) A priori, we do not have a guarantee that $\Delta \cdot l_n^\Delta < 1$.

(ii) Even for reasonably small time steps $\Delta$, we can always expect to see two or more events in a given time bin, especially if events are clustered due to self-excitation.

We then embed the discrete-time process back into the continuous-time setting by means of càdlàg (right continuous with left limits) embedding, thus obtaining a new process $N^\Delta$ defined as

$$N_t^\Delta = \sum_{k=0}^{\lfloor t/\Delta \rfloor} X_k.$$

Since the new process has a discontinuous trajectory, the uniform distance fails to capture its proximity to the original process $N$, as two very similar trajectories will have a large distance in the uniform metric as soon as they have one discontinuity that does not occur at the exact same time. This is why we provide approximation results in two different spaces that are more adapted to càdlàg processes: the fractional Sobolev space and the Skorokhod space, both of which are defined in Section 3.

To the best of our knowledge, this is the first work that provides strong approximation results for Hawkes processes, the other two studies [23, 29] proving weak convergence in the Skorokhod metric. Furthermore, the bounds are provided in this article with explicit dependence on the time step $\Delta$ and the time horizon $T$, making them useful for numerical applications. We also point out that unlike [23], in which Markov process techniques are used, we prove convergence results for any kernel of finite $p$-variation and not only of the Erlang type. Moreover, since we use Poisson autoregressions instead of INAR time series as used in [29], we give convergence results that work for nonlinear Hawkes processes.

It is also worthwhile highlighting that our framework is different from the classical approximation results of Poisson jump-diffusions, because interarrival times cannot be explicitly simulated for Hawkes processes (unless the process is linear with an exponential kernel [10]).

This article is organized as follows: In Section 2 we give the rigorous definitions of the continuous-time and discrete-time Hawkes processes, as results of thinning from the same underlying Poisson randomness. In Section 3, explicit bounds on the Sobolev and Skorokhod distances between the continuous-time Hawkes process and its discrete-time approximation are given. Our findings are summarized in Section 4. Appendix A contains the different proofs, and Appendix B contains the technical lemmata.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

## 2. Definitions

In the two following subsections, we define the compound marked Hawkes process, both in continuous time and in discrete time using the Poisson embedding idea from [6]. The main source of randomness for these two processes is a tridimensional Poisson measure $P$ that takes values in the configuration space

$$\Omega := \left\{ \omega = \sum_{i=1}^{n} \delta_{(\tau_i', \theta_i, y_i)},\ 0 < \tau_1' < \cdots < \tau_n',\ (\theta_i, y_i) \in \mathbb{R}_+ \times \mathbb{R} \text{ and } n \in \mathbb{N} \cup \{+\infty\} \right\}.$$

Given a nonnegative Borel measure $\nu$ on $\mathbb{R}$ such that $\nu(\mathbb{R}) = 1$, we take $\mathbb{P}$ to be the probability measure under which the point measure $P$ defined as

$$P((0, t], (0, \theta], (-\infty, y])(\omega) = \omega((0, t], (0, \theta], (-\infty, y]), \quad (t, \theta, y) \in \mathbb{R}_+^2 \times \mathbb{R},$$

is a Poisson measure with intensity $\mathrm{d}t \times \mathrm{d}\theta \times \nu(\mathrm{d}y)$. We also let

$$\mathcal{F}_t = \sigma\ (P(\mathcal{T} \times S), \quad \mathcal{T} \subset \mathcal{B}((0, t]) \text{ and } S \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R})),$$

be the filtration associated with $P$. Throughout this article, the conditional expectation knowing $\mathcal{F}_t$ is denoted by $\mathbb{E}_t$.

### 2.1. The continuous-time setting

For a fixed $T > 0$, let $h$ be a function in $L^1([0, T])$. Let $\psi$ be a positive $L$-Lipschitz function on $\mathbb{R}$ and $b$ be a positive Borel function on $\mathbb{R}$. We now state the stability assumption on the aforementioned elements.

**Assumption 1.** We have

$$\rho_h := L\|h\|_1 \mathbb{E}[b(Y)] < 1,$$

where $Y$ is a random variable of distribution $\nu$ and $\|h\|_1 = \int_0^T |h(t)|\mathrm{d}t$. We also assume that $\nu$ has a finite first moment; that is, $\mathbb{E}[|Y|] < +\infty$.

We now give the definition of the marked compound Hawkes process (or risk) as a result of thinning from the underlying tridimensional Poisson measure $P$. The procedure is now standard (see, for example, [6] or [33]) and gives the Hawkes process as the unique pathwise solution of a stochastic differential equation (SDE) as examined in more detail in [21].

**Definition 1.** Let P be a tridimensional Poisson measure on $\mathbb{R}_+^2 \times \mathbb{R}$ of intensity $\mathrm{d}t \times \mathrm{d}\theta \times \nu(\mathrm{d}y)$. Fix the time horizon $T > 0$ and let $h \in L^1([0, T])$, $\psi : \mathbb{R} \to \mathbb{R}_+$ $L$-Lipschitz, and $b : \mathbb{R} \to \mathbb{R}_+$ such that Assumption 1 is in force. The SDE on [0,T]

$$\begin{cases} R_t &= \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} y \mathbf{1}_{\theta \le \lambda_s} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y), \\ \lambda_t &= \psi\left( \int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}} h(t-s) \mathbf{1}_{\theta \le \lambda_s} b(y) P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y) \right) \end{cases} \tag{3}$$

has a unique pathwise solution $(R, \lambda)$ such that R is $\mathcal{F}$ measurable and $\lambda \in L^1(\mathrm{d}\mathbb{P} \times \mathrm{d}t)$ is $\mathcal{F}$ predictable.

We say that $R$ is a marked Hawkes risk of kernel $h$, jump rate $\psi$, and claim size distribution $\nu$. We call $\lambda$ the *intensity* of $R$, and $b$ the marks the modulation function.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

Furthermore, we define the marked simple Hawkes process $N$ as

$$N_t := \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\theta\le\lambda_s} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y), \quad t\in[0,T],$$

and the auxiliary process as

$$\xi_t := \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} b(y)\mathbf{1}_{\theta\le\lambda_s} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y).$$

*Proof.* The proof is an adaptation of a classical contraction argument to the $L^1$ setting, following the proof of Theorem 1.2 in [18].

First, we start by mentioning that the system (3) is triangular. Therefore, once $\lambda$ is constructed, $R$ is simply obtained by thinning from the underlying Poisson measure $P$. We thus focus on the construction of the intensity $\lambda$.

Let $f$ be the function that maps a predictable, integrable process on $[0, T]$ of finite expectation to the integrable predictable process $\kappa := f(\lambda)$ defined by

$$\kappa_t = \psi\left(\int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}} h(t-s)\mathbf{1}_{\theta\le\lambda_s} b(y) P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)\right).$$

Given two predictable processes $\lambda$ and $\lambda'$ and their images $\kappa$ and $\kappa'$, we have

$$\begin{aligned}\mathbb{E}\left[\int_0^T \left|\kappa'_t - \kappa_t\right| \mathrm{d}t\right] &\le L\mathbb{E}\left[\int_0^T \left|\int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}} h(t-s)(\mathbf{1}_{\theta\le\lambda'_s} - \mathbf{1}_{\theta\le\lambda_s}) b(y) P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)\right| \mathrm{d}t\right] \\ &\le L\mathbb{E}\left[\int_0^T \int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}} |h(t-s)| \left|\mathbf{1}_{\theta\le\lambda'_s} - \mathbf{1}_{\theta\le\lambda_s}\right| b(y) P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)\mathrm{d}t\right],\end{aligned}$$

and using the fact that $\lambda$ and $\lambda'$ are predictable, we have

$$\begin{aligned}\mathbb{E}\left[\int_0^T \left|\kappa'_t - \kappa_t\right| \mathrm{d}t\right] &\le L\mathbb{E}\left[\int_0^T \int_{[0,t)} |h(t-s)| \left|\lambda'_s - \lambda_s\right| \mathbb{E}[b(Y)]\mathrm{d}s\mathrm{d}t\right] \\ &= L\mathbb{E}[b(Y)]\mathbb{E}\left[\int_0^T \int_s^T |h(t-s)| \left|\lambda'_s - \lambda_s\right| \mathrm{d}t\mathrm{d}s\right] \\ &= L\mathbb{E}[b(Y)]\mathbb{E}\left[\int_0^T \left|\lambda'_s - \lambda_s\right| \int_s^T |h(t-s)|\mathrm{d}t\mathrm{d}s\right] \\ &\le L\mathbb{E}[b(Y)]\|h\|_1 \mathbb{E}\left[\int_0^T \left|\lambda'_s - \lambda_s\right| \mathrm{d}t\right].\end{aligned}$$

By introducing the norm

$$\|X\| = \mathbb{E}\left[\int_0^T |X_t| \mathrm{d}t\right],$$

we can write the last inequality in the form

$$\|f(\lambda') - f(\lambda)\| \le L\mathbb{E}[b(Y)]\|h\|_1 \|\lambda' - \lambda\|,$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

where $L\mathbb{E}[b(Y)]\|h\|_1 < 1$ thanks to Assumption 1. Using the Banach fixed-point theorem, we conclude that $f$ has a unique fixed point $\lambda$. □

**Remark 1.** The assumption that $b$ is positive is superfluous from a mathematical point of view and can be omitted (up to the introduction of absolute values). However, we chose to keep it here because we want the excitation/inhibition to be determined by the sign of the kernel $h$, assuming that $\psi$ is monotonous.

If $\psi(x) = \mu + x$, $x \in \mathbb{R}_+$ for a positive constant $\mu$ under the constraint $h \geq 0$, we say that the Hawkes process/risk is *linear*. In this particular case, the Hawkes dynamics have a branching process representation. Although the first two moments of the process are explicitly known (up to the computation of an infinite sum of convolutions of $h$; see [3, 21]), linear Hawkes processes do not allow for self-inhibition.

The marked Hawkes risk can also be defined in a more elementary (yet informal) way without explicit use of the thinning procedure. Let $N$ be a simple point process on $[0, T]$ and $(Y_k)_{k\in\mathbb{N}}$ be a family of independent and identically distributed (i.i.d) random variables of common distribution $\nu$. Note that we do not assume the variables $(Y_k)_{k\in\mathbb{N}}$ and the process $N$ to be independent. $N$ is said to be a marked simple Hawkes process if its intensity $\lambda$ follows the dynamics

$$\lambda_t = \psi\left(\sum_{\tau_i<t} h(t-\tau_i)b(Y_i)\right),$$

where $(\tau_i)_{i\in\mathbb{N}}$ are the arrival times of the points of $N$. Note that the intensity can also be expressed in the integral form

$$\lambda_t = \psi\left(\int_0^{t-} h(t-s)\mathrm{d}\xi_s\right),$$

where $\xi$ is the auxiliary process $\xi_t = \sum_{k=1}^{N_t} b(Y_k)$, $t \in [0, T]$. The risk process is simply the aggregation of all the marks

$$R_t = \sum_{k=1}^{N_t} Y_k, \quad t \in [0, T].$$

The marked Hawkes process is useful in modeling phenomena where the intensity is impacted not only by the realization of an event $\tau_i$ but also by its 'severity' $Y_i$. For instance, the choice $b(y) = \mathbf{1}_{y\geq a}$ means that only claims of a size larger than a given threshold $a$ have an impact on the intensity. Karabash and Zhu [24] provided limit theorems for a general class of marked Hawkes processes, albeit in the linear setting.

This constitutes a generalization of the compound Hawkes process usually studied in the literature (see [13, 26, 28]) where the marks $Y$ do not impact the intensity.

If the modulation function $b$ is chosen to be equal to the constant 1, we can retrieve the usual unmarked intensity $\lambda_t = \psi\left(\int_0^{t-} h(t-s)\mathrm{d}N_s\right)$. Similarly, the choice $Y \equiv 1$ ensures that $R \equiv N$; hence we will focus exclusively on $R$.

The goal is to suggest an intuitive discretization scheme on $[0, T]$ and to obtain a bound on the distance between this scheme and the continuous-time process in a convenient functional space.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

### 2.2. The discrete-time setting

Throughout this article, the bounded interval $[0, T]$ is discretized into $M \in \mathbb{N}^*$ equidistant intervals $(t_i, t_{i+1}]$ of length $\Delta$, where $\Delta = T/M$. For a given $t \in [0, T]$, we define $(t)_\Delta = \lfloor t/\Delta \rfloor \Delta$ to be its projection on the time grid. We also define $n_t = \lfloor t/\Delta \rfloor$.

For a $k \leq M$, we set $h_k = h(k\Delta)$. We omit the dependence on $\Delta$ to avoid cumbersome notation. Before defining the discrete-time marked Hawkes risk, we give the following stability assumption.

**Assumption 2.** Assume that

$$\rho_{h,\Delta} := L \sum_{k=1}^{M} |h_k| \Delta \mathbb{E}[b(Y)] < 1,$$

where $L$ is the Lipschitz coefficient of $\psi$ and $Y$ is a random variable of distribution $\nu$. We also assume that $\nu$ has a finite first moment; that is, $\mathbb{E}[|Y|] < +\infty$.

Just like the continuous-time case, we build the discrete-time marked Hawkes risk using the same tridimensional Poisson measure $P$.

**Definition 2.** Let P be a tridimensional Poisson measure on $\mathbb{R}_+^2 \times \mathbb{R}$ of intensity $\mathrm{d}t \times \mathrm{d}\theta \times \nu(\mathrm{d}y)$. Fix the time horizon $T > 0$ and let $h \in L^1([0, T])$, $\psi : \mathbb{R} \to \mathbb{R}_+$ $L$-Lipschitz, and $b : \mathbb{R} \to \mathbb{R}_+$ such that Assumption 2 is in force. Fix $M \in \mathbb{N}$ and let $\Delta = T/M$.

Define the measurable (with respect to the filtration $(\mathcal{F}_{n\Delta})_{n=0,\dots,M}$) sequence $(X_k^\Delta)_{k=0,\dots,M}$ and the predictable sequence $(l_k^\Delta)_{k=0,\dots,M}$ recursively:

$$\begin{cases} X_0^\Delta &= 0, \;\; l_0^\Delta = l_1^\Delta = \psi(0), \;\; D_0^\Delta = 0, \\ X_n^\Delta &= \int_{((n-1)\Delta, n\Delta] \times \mathbb{R}_+ \times \mathbb{R}} b(y) \mathbf{1}_{\theta \leq l_n^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y), \\ l_n^\Delta &= \psi\left(\sum_{k=1}^{n-1} h_{n-k} X_k^\Delta\right), \\ D_n^\Delta &= \int_{((n-1)\Delta, n\Delta] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\theta \leq l_n^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y), \end{cases} \quad \text{for } n \geq 1. \tag{4}$$

The discrete-time Hawkes risk $R^\Delta$ and intensity $\lambda^\Delta$ are the càdlàg piecewise constant processes defined as

$$\begin{cases} \lambda_t^\Delta &= \lambda_{(t)_\Delta}^\Delta = l_{n_t}^\Delta, \\ R_t^\Delta &= R_{(t)_\Delta}^\Delta = \sum_{k=1}^{n_t} \int_{((k-1)\Delta, k\Delta] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\theta \leq \lambda_{k\Delta}^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y). \end{cases}$$

Furthermore, the discrete-time auxiliary process can be defined as

$$\xi_t^\Delta = \xi_{(t)_\Delta}^\Delta = \sum_{k=1}^{n_t} X_k,$$

and the discrete Hawkes process can be defined as

$$N_t^\Delta = N_{(t)_\Delta}^\Delta = \sum_{k=1}^{n_t} D_k.$$

Note that despite their names, the discrete-time processes defined above are continuous-time embeddings of the times series $l^\Delta$ and $X^\Delta$; however, their values are allowed to change exclusively on the points of the discretization grid.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

**Algorithm 1** Nonlinear Poisson autoregression with general kernel.

**Require:** $(X_k^\Delta)_{k\in\{1,\dots,n\}}$, $(R_{k\Delta}^\Delta)_{k\in\{1,\dots,n\}}$
  $S \leftarrow 0$
  **for** $i = 1, \dots, n$ **do**
    $S \leftarrow S + h_{n+1-i} X_i^\Delta$
  **end for**
  $l_{n+1}^\Delta \leftarrow \psi(S)$ ▷The stochastic intensity built from previous counts
  $D_{n+1}^\Delta \leftarrow \mathcal{P}\left(\Delta \cdot l_{n+1}^\Delta\right)$
  Sample $Y_1^{n+1}, \dots, Y_{D_{n+1}^\Delta}^{n+1} \sim \nu$ i.i.d.
  $X_{n+1}^\Delta \leftarrow 0$
  $R_{(n+1)\Delta}^\Delta \leftarrow R_{n\Delta}^\Delta$
  **for** $i = 1, \dots, D_{n+1}^\Delta$ **do**
    $X_{n+1}^\Delta \leftarrow X_{n+1}^\Delta + b\left(Y_i^{n+1}\right)$
    $R_{(n+1)\Delta}^\Delta \leftarrow R_{(n+1)\Delta}^\Delta + Y_i^{n+1}$
  **end for**
  **return** $(X_k^\Delta)_{k\in\{1,\dots,n+1\}}$, $(R_{k\Delta}^\Delta)_{k\in\{1,\dots,n+1\}}$

### 2.3. Simulation of nonlinear Poisson autoregressions

For simulation purposes, it is possible to build the sequences $X^\Delta$ and $l^\Delta$ without simulating the underlying Poisson measure $P$. This is based on the following observation: knowing $X_1^\Delta, \dots, X_n^\Delta$ (and thus $l_{n+1}^\Delta$ according to (4)), the variable

$$X_{n+1}^\Delta = \int_{(n\Delta,(n+1)\Delta]\times\mathbb{R}_+\times\mathbb{R}} b(y)\mathbf{1}_{\theta\le l_{n+1}^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)$$

is a compound Poisson variable; that is,

$$X_{n+1}^\Delta = \sum_{k=1}^{D_{n+1}^\Delta} b(Y_k^{n+1}),$$

where $D_{n+1}^\Delta | (X_1^\Delta, \dots, X_n^\Delta) \sim \mathcal{P}(\Delta \cdot l_{n+1}^\Delta)$ and $Y_1^{n+1}, \dots, Y_{D_{n+1}^\Delta}^{n+1}$ are i.i.d variables of distribution $\nu$ and independent of $D_{n+1}^\Delta$.

Algorithm 1 yields the simulation of $(X_{n+1}^\Delta, R_{(n+1)\Delta}^\Delta)$ given $(X_k^\Delta, R_{k\Delta}^\Delta)_{k=1,\dots,n}$.

Intuitively, the variable $X_{n+1}^\Delta = \sum_{k=1}^{D_{n+1}^\Delta} b(Y_k^{n+1})$ is the discrete equivalent of the increment $\mathrm{d}\xi_t$ given in Definition 1. This is why we see $R^\Delta$ as a good approximation of $R$, which is what we will prove in the rest of this article.

We conclude this subsection by discussing the numerical cost of simulating the discrete-time process. Generally, the computation of the recursion (4) is of order $O(M^2)$. This cost can be reduced in two cases:

(i) If the kernel is an Erlang function, that is, $h(t) = Q(t)e^{-\beta t}$, where $\beta > 0$ and $Q$ is a polynomial of degree $q$, then the intensity is a Markov chain (up to the introduction of auxiliary processes; see [23]) and the computation cost is of order $O(qM)$.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

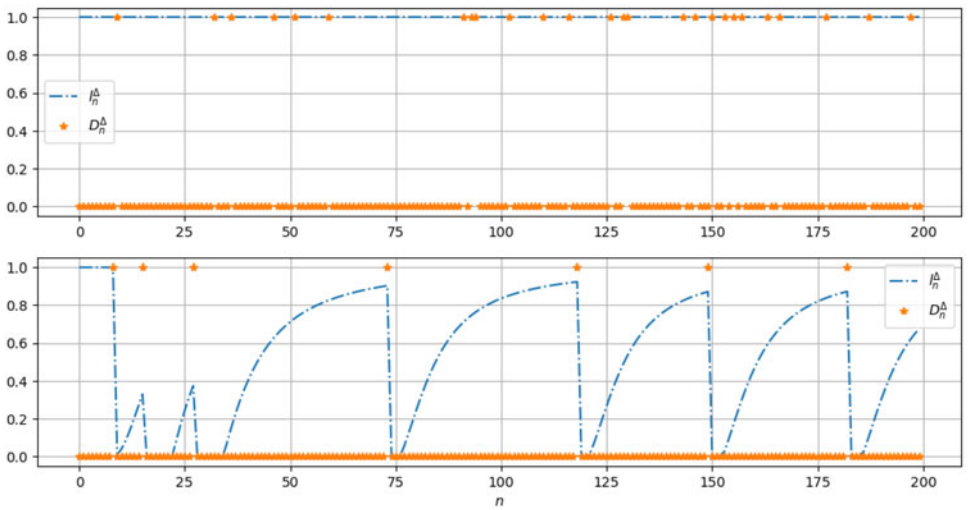



FIGURE 1. Top: Discrete Poisson process of a constant deterministic intensity. Because of the independence, some nonzero counts are close to other nonzero counts. Bottom: Self-inhibiting Poisson autoregression. The realization of a nonzero count decreases the likelihood of observing counts in the near future.

(ii) If the kernel $h$ is of compact support $S$, that is, $h(t) = 0$ for any $t \geq S$, then the cost is of order $O(rM)$, where $r = S/\Delta$.

As an illustration, we simulate a nonlinear self-inhibiting Poisson autoregression (see Figure 1). For simplicity, we let $\nu(\mathrm{d}y) = \delta_1(\mathrm{d}y)$. In this case, $X_n^\Delta = D_n^\Delta$ for all $n \geq 0$. We choose the kernel to be

$$h(t) = -\frac{1}{1+t^2},$$

and the jump rate $\psi(x) = (1+x)_+$. We also simulate a discrete Poisson process of constant discrete intensity $\Delta$ (see Figure 1).

**Remark 2.** We seek to approximate $h$ by a piecewise constant function. We choose to approximate $h$ on $[t_k, t_{k+1}[$ by $h(t_k)$. We can also approximate $h$ on $[t_k, t_{k+1}[$ by $\Delta^{-1}\int_{(t_k,t_{k+1}]} h(s)\mathrm{d}s$. This latter choice is more convenient from the mathematical point of view. Nevertheless, the exact computation of the coefficients $(\Delta^{-1}\int_{(t_k,t_{k+1}]} h(s)\mathrm{d}s)_k$ can be cumbersome for general kernel $h$.

Before we show that $R^\Delta$ converges to $R$ as the time step $\Delta$ goes to zero, we discuss the choice of the spaces in which the convergence occurs.

### 2.4. Discussion of functional spaces

The quality of the strong approximation of continuous stochastic processes $(Z_t)_{t\in[0,T]}$ (and, in particular, diffusion SDEs) is evaluated by controlling the average uniform error

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |Z_t - Z_t^\Delta|\right].$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

While the supremum norm is adapted to processes that have continuous trajectories, it can also be extended to jump-diffusion SDEs driven by a homogeneous Poisson process [7]. This is the case because the arrival times of a Poisson process $(\tau_i)$ are explicitly known and can be simulated exactly (unlike for general Hawkes processes), therefore rendering the problem equivalent to approximating a diffusion on $(\tau_i, \tau_{i+1}]$.

The supremum norm, however, should not be used to evaluate the proximity of càdlàg trajectories.

By construction, the Hawkes process and the discrete Hawkes process are càdlàg piecewise constant functions. The Hawkes process jump times are contained in the set of the jump times of the underlying Poisson random measure. On the other hand, the jumps of the discrete Hawkes process are contained in $\{t_k,\ k = 1, ..., N\}$, the set of the points of the subdivision. Thus, almost surely $\sup_{s\in[0,T]} |N_s - N_s^\Delta| \geq 1$ on a set with non-null probability. The processes $(N^\Delta)_{\Delta\in(0,1)}$ will not converge to $N$ in the uniform norm on compact sets of $\mathbb{R}^+$ almost surely. This leads us to compute the distance between the Hawkes process and the discrete Hawkes process in $\mathbb{D}([0, T], \mathbb{R})$ equipped with the Skorokhod metric

$$d_S(f, g) = \inf_{\mu\in\Lambda} \left\{ \sup_{0\leq t\leq T} |t - \mu(t)| \vee \sup_{0\leq t\leq T} |f(t) - g(\mu(t))| \right\}, \tag{5}$$

where $\Lambda$ is the set of strictly increasing continuous functions from $[0, T]$ to itself such that $\mu(0) = 0$ and $\mu(T) = T$. Note that since this distance allows some flexibility in the time at which the jump occurs (the role of the time change $\mu$), the problem that occurs with the uniform distance disappears.

By working in $\mathbb{D}([0, T], \mathbb{R})$, we do not pay attention to the properties or the potential regularity of the paths of the processes. As increasing processes, the sample path of the Hawkes process and the discrete Hawkes process belong to a set of functions with finite bounded variation starting from 0 endowed with the distance

$$d_{FV,T}(f, g) = \sup_{(s_i)} \sum_i |f(s_i) - g(s_i)|,$$

where the supremum runs over all finite partitions $(s_i)$ of $[0, T]$. It should be noticed that $\sup_{s\in[0,T]} |f(s) - g(s)| \leq d_{FV,T}(f, g)$ if $f(0) = g(0) = 0$. Thus, $N^\Delta$ will not converge to $N$ when $\Delta$ converges to 0 for the $d_{FV,T}$ distance.

As pointed out in [8] or [9], fractional Sobolev spaces provide a good framework for the study of the rate of convergence of càdlàg piecewise constant processes. According to Definition 2.1 in [4], for $p \geq 1$ and $0 < \eta < 1$

$$W_T^{\eta,p} = \{u \in L^p([0, T]),\ \frac{u(s) - u(t)}{(t - s)^{\eta+\frac{1}{p}}} \in L^p([0, T]^2)\}. \tag{6}$$

They are Banach spaces endowed with the norm $\|\cdot\|_{W_T^{\eta,p}}$, where

$$\|u\|^p_{W_T^{\eta,p}} := \int_0^T |u(s)|^p \mathrm{d}s + \int\int_{[0,T]^2} \frac{|u(s) - u(t)|^p}{(t - s)^{\eta p+1}} \mathrm{d}s\mathrm{d}t.$$

The classical Sobolev space is $W^{1,1}([0, T]) = \{u \in L^1([0, T]),\ u(\cdot) = u(0) + \int_0^\cdot u'(s)\mathrm{d}s \text{ with } u' \in L^1([0, T])\}$ with $\|u\|_{W^{1,1}([0,T])} = \int_0^T [|u(s)| + |u'(s)|]\mathrm{d}s$. For $p = 1$,

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

fractional Sobolev spaces are interpolated spaces between $L^1([0,T])$ and the classical Sobolev space:

$$L^1([0,T]) \subset W_T^{\eta,1} \subset W^{1,1}([0,T]) \quad \text{for all } \eta \in (0,1]$$

with continuous injections; see [4] for more details. One can go further. The sample paths of $R - R^\Delta$ are not absolutely continuous with respect to the Lebesgue measure, but they can be seen as fractional integral in the following senses. Since $R - R^\Delta$ is a linear combination of indicator functions, according to Example 4.1 in [4], there exists a process $g^\Delta$ with sample paths in $L^1([0,T])$ almost surely such that $R - R^\Delta = I_{0^+}^\eta(g^\Delta)$. Here, for $u \in L^1([0,T])$, the Liouville fractional of $u$ of order $\eta$ is defined by $I_{0^+}^\eta(u)(t) = \frac{1}{\Gamma(\eta)} \int_0^t (t-s)^{\eta-1} u(s)\mathrm{d}s, \quad t \in [0,T]$. The Riemann–Liouville space

$$I_{0^+}^\eta(L^1([0,T])) = \{u = I_{0^+}^\eta(u'), \quad u' \in L^1([0,T])\}$$

endowed with

$$\|u\|_{RL,\eta,T} = \|u'\|_{L^1([0,T])}$$

is a Banach space. According to Theorem 4.1 in [4], since $R - R^\Delta$ is piecewise constant, $\lim_{\eta \to 1} \|R - R^\Delta\|_{RL,\eta,T} = d_{FV,T}(R, R^\Delta)$. Moreover, according to Theorem 3.2 in [4], for $\eta', \eta \in (0,1)$ $\eta > \eta'$,

$$W_T^{\eta,1} \cap I_{0^+}^{\eta'}(L^1([0,T])) \subset W_T^{\eta',1},$$

with a continuous injection. Thus, we compute the distance between the Hawkes process and the discrete Hawkes process in $W^{\eta,1}([0,T])$ for all $0 < \eta < 1$ and in Riemann–Liouville fractional Sobolev spaces. See [4] for some definitions. We expect that the rate of convergence in $W_T^{\eta,1}$ will vanish when $\eta$ goes to 1.

## 3. Main Results

In this section, we show the strong convergence of the discrete-time Hawkes risk process to the continuous-time counterpart and give convergence rates both in the time step $\Delta$ and in the time horizon $T$. In this section $K$ is a constant that may depend on $L$, $\nu$, $h$, and $b$ but is independent of $T$ and $\Delta$. We emphasize that while $K$ is finite, if the stability assumption given by Assumption 1 holds, it can diverge to infinity as $\rho_h$ approaches 1.

We also emphasize that our continuous-time and discrete-time processes are thinned from the same Poisson measure $P$. As an illustration, Figure 2 shows both processes and their underlying common randomness for two values of the time step $\Delta$.

### 3.1. Preliminary results

First, we give a regularity result on the limit process; that is, the continuous-time process. This is motivated by the fact that if $\lambda$ is too irregular, a piecewise approximation of it would accept (or miss) many points that are not accepted (or are rejected) in the continuous-time process.

**Lemma 1.** *Let $h$, $\psi$, and $b$ be three functions satisfying Assumption 1. There exists a constant $K$ such that for any $v \in [0,T]$ we have*

$$\mathbb{E}\left[|\lambda_v - \lambda_{(v)_\Delta}|\right] \le K \left( \int_0^\Delta |h(y)| \mathrm{d}y + \sup_{\epsilon \in [0,\Delta]} \int_0^{T-\Delta} |h(y+\epsilon) - h(y)| \, \mathrm{d}y \right).$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

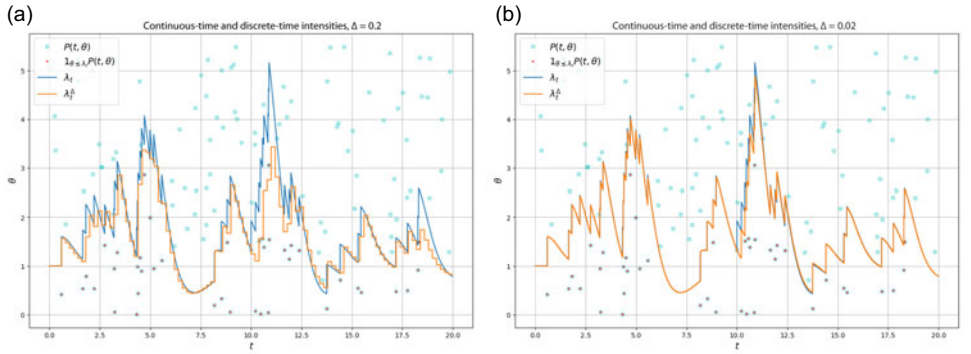



FIGURE 2. A realization of the discrete-time and continuous-time intensities as thinning from the same underlying Poisson measure $P$. The jump rate $\psi(x) = (1+x)_+$ and the kernel function $h(t) = \frac{0.6 \cdot \cos(t)}{1+t^2}$. (*a*) When the discretization step $\Delta$ is relatively large, the discrete intensity is more susceptible to miss points that are accepted by the continuous-time trajectory; (*b*) As the discretization step $\Delta$ becomes smaller, the two trajectories become closer and tend to accept the exact same points.

*Proof.* For $v = T$, $\lambda_v - \lambda_{(v)_\Delta} = 0$; thus we choose $v \in [0, T)$. Since $\psi$ is $L$-Lipschitz and using a linear change of variables, we have

$$\begin{aligned}
&\mathbb{E}\left[|\lambda_v - \lambda_{(v)_\Delta}|\right] \\
&\quad \le L\mathbb{E}\left[\left|\int_0^v h(v-s)\mathrm{d}\xi_s - \int_0^{(v)_\Delta} h((v)_\Delta - s)\mathrm{d}\xi_s\right|\right] \\
&\quad \le L\left(\mathbb{E}\left[\left|\int_{(v)_\Delta}^v h(v-s)\mathrm{d}\xi_s\right|\right] + \mathbb{E}\left[\left|\int_0^{(v)_\Delta} h(v-s) - h((v)_\Delta - s)\mathrm{d}\xi_s\right|\right]\right) \\
&\quad \le L\left(\int_{(v)_\Delta}^v |h(v-s)|\,\mathbb{E}[b(Y)]\mathbb{E}\left[\lambda_s\right]\mathrm{d}s + \int_0^{(v)_\Delta} |h(v-s) - h((v)_\Delta - s)|\,\mathbb{E}[b(Y)]\mathbb{E}[\lambda_s]\mathrm{d}s\right),
\end{aligned}$$

because $\mathrm{d}\mathbb{E}[\xi_s] = \mathbb{E}[b(Y)]\mathbb{E}[\lambda_s]\mathrm{d}s$. Using the bound on the expected value proved in Lemma 5, we have

$$\mathbb{E}\left[|\lambda_v - \lambda_{(v)_\Delta}|\right] \le \frac{\mathbb{E}[b(Y)]L\psi(0)}{1 - \mathbb{E}[b(Y)]L\|h\|_1}\left(\int_0^\Delta |h(y)|\mathrm{d}y + \int_0^{(v)_\Delta} |h(y+v-(v)_\Delta) - h(y)|\,\mathrm{d}y\right),$$

and the result follows immediately. □

This lemma shows that the kernel's regularity in the sense of the shift operator in the $L^1$ norm yields the intensity's regularity.

Since the risk process is the result of accepting the points of the underlying Poisson measure under the intensity's curve, the approximation of the intensity is an important step in proving the convergence of $R^\Delta$ to $R$. We now give the first bound on the distance between intensities on the points of the discretization grid.

The following constants will be useful for the sequel.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

**Definition 3.** We recall that $\rho_h = L\|h\|_1 \mathbb{E}[b(Y)]$ and $\rho_{h,\Delta} = L\sum_{k=1}^{M} |h_k|\Delta\mathbb{E}[b(Y)]$. We set

$$C_S(h, \Delta) = \frac{1}{(1-\rho_h)} + \frac{1}{(1-\rho_{h,\Delta})},$$

$$C_R(h, \Delta) = \int_0^{\Delta} |h(y)|\mathrm{d}y + \sup_{\epsilon\in[0,\Delta]} \int_0^{T-\Delta} |h(y+\epsilon) - h(y)|\mathrm{d}y + \int_0^{T-\Delta} |h(y) - h((y)_\Delta + \Delta)|\mathrm{d}y.$$

The constant $C_S(h, \Delta)$ is related to the stability assumptions given by Assumptions 1 and 2, while the constant $C_R(h, \Delta)$ depends on the regularity of the kernel.

**Lemma 2.** *Let $\lambda$ (or $\lambda^\Delta$) be the intensities defined by thinning from the Poisson measure $P$ in Definition 1 (or Definition 2). Assume that Assumptions 1 and 2 hold. There exists a constant $K$ such that for all $u \in [0, T]$ we have*

$$\mathbb{E}\left[\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|\right] \le KC_R(h, \Delta)C_S(h, \Delta). \tag{7}$$

The proof of this Lemma is given in Appendix A.

Before proving the convergence of the discrete-time Hawkes risk, we point out that if $h$ satisfies Assumption 1 and is sufficiently regular, that is, it satisfies Assumption 3, then Assumption 2 becomes superfluous.

**Assumption 3.**

$$\lim_{\Delta\to 0} \int_0^{T-\Delta} |h(t) - h((t)_\Delta + \Delta)|\mathrm{d}t = 0.$$

**Lemma 3.** *If Assumptions 1 and 3 are satisfied by a kernel $h$, then for any $\epsilon > 0$ small enough, there exists a threshold $\Delta_1 > 0$ such that*

$$\frac{1}{1-\rho_{h,\Delta}} \le \frac{1}{1-\rho_h} + \epsilon$$

*for any $\Delta \le \Delta_1$. In particular, if there exists $\Delta_0 > 0$, Assumption 2 is verified by $h$ for any $\Delta \le \Delta_0$.*

*Proof.* For a given kernel $h \in L^1([0, T])$ we have

$$\begin{aligned}
\Delta \sum_{i=1}^{M-1} |h(i\Delta)| &\le \sum_{i=1}^{M-1} \Delta \left| h(i\Delta) - \Delta^{-1} \int_{(i-1)\Delta}^{i\Delta} h(s)\mathrm{d}s \right| + \sum_{i=1}^{M-1} \left| \int_{(i-1)\Delta}^{i\Delta} h(s)\mathrm{d}s \right| \\
&\le \sum_{i=1}^{M-1} \int_{(i-1)\Delta}^{i\Delta} |h(i\Delta) - h(s)|\mathrm{d}s + \int_0^T |h(s)|\mathrm{d}s \\
&= \int_0^{T-\Delta} |h(s) - h((s)_\Delta + \Delta)|\mathrm{d}s + \|h\|_1.
\end{aligned} \tag{8}$$

Since the first time on the right-hand side of the last inequality tends to zero, we conclude that Assumption 2 is satisfied below a certain threshold $\Delta_1 > 0$. □

To conclude this subsection, we comment on the role of the stability assumption given by Assumption 1. First, we point out that in the linear setting (that is, $\psi(x) = \mu + x$ for some $\mu > 0$), the violation of that assumption means that $\mathbb{E}[\lambda_t]$ diverges. In particular, if $h(t) = \alpha \mathrm{e}^{-\beta t}$ with $\alpha > \beta$, then $\mathbb{E}[\lambda_t] \sim C\mathrm{e}^{(\alpha-\beta)t}$; see [13]. We expect that the bounds proven in

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

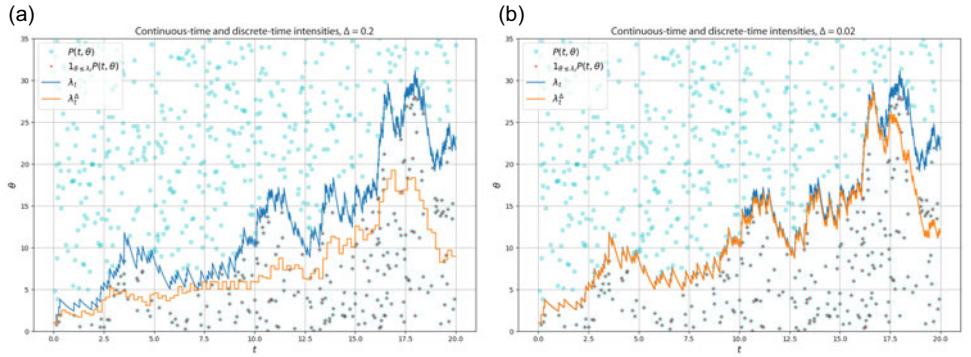



FIGURE 3. A realization of the discrete-time and continuous-time intensities as thinning from the same underlying Poisson measure $P$. The jump rate $\psi(x) = 1 + x$ and the kernel function $h(t) = 1.01\mathrm{e}^{-t}$. This figure should be contrasted with Figure 2. (*a*) When the discretization step $\Delta$ is relatively large, instability means that the continuous-time intensity and its discrete-time approximation diverge greatly; (*b*) As the discretization step $\Delta$ becomes smaller, the two trajectories become closer, at least for short times. As time increases, this becomes less true as instability amplifies small differences.

Lemma 1 no longer hold uniformly in time, but instead hold with a factor $\mathrm{e}^{cT}$ for some $c > 0$, where $T$ is the time horizon. Such results can be proven with use of the Grönwall lemma, but this is beyond the scope of this article.

As an illustration, we simulate a slightly unstable linear Hawkes process with an exponential kernel (not marked) and its discrete-time approximation in Figure 3.

With these results stated, we are now ready to give the first approximation result for the risk process. We will skip the dependence on $C_S(h, \Delta)$ in the convergence results.

### 3.2. Convergence in the fractional Sobolev space

The goal of this subsection is to give a bound on the distance between the risk process on the compact interval $[0, T]$ and its discrete counterpart in the fractional Sobolev space defined in (6). Recall that for $\eta \in (0, 1)$, $W_T^{\eta,q} := \left\{u \in L^q([0, T]) : \|u\|_{W_T^{\eta,q}} < +\infty\right\}$. Here we restrict ourselves to $q = 1$ and omit it from the notation. We also recall the fractional integral. For $u \in L^1([0, T])$ and $0 < \eta < 1$,

$$I_{0+}^{\eta}(u)(t) = \frac{1}{\Gamma(\eta)} \int_0^t (t - s)^{\eta - 1} u(s) \mathrm{d}s, \quad t \in [0, T],$$

and the Riemann-Liouville space

$$I_{0+}^{\eta}(L^1([0, T])) = \{u \in L^1([0, T]) \mid \text{there exists } v \in L^1([0, T]), \quad u = I_{0+}^{\eta}(v)\}.$$

**Proposition 1.** *Let $T > 0$. Let $(h, \psi, \nu, b)$ fulfill Assumption 1. Let $R$ (or $R^{\Delta}$) be the continuous-time (or discrete-time) risk process. We have*

$$(R_t)_{t \in [0,T]} \in W_T^{\eta} \cap I_{0+}^{\eta}(L^1([0, T])) \text{ and } (R_t^{\Delta})_{t \in [0,T]} \in W_T^{\eta} \cap I_{0+}^{\eta}(L^1([0, T]))$$

*almost surely.*

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

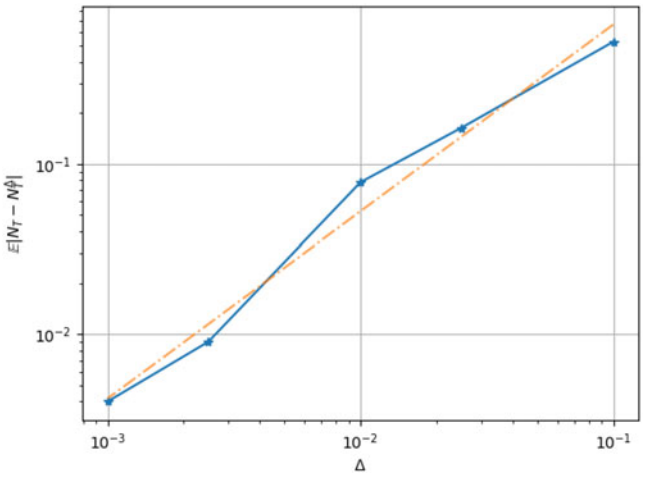



FIGURE 4. A Monte Carlo approximation of $\mathbb{E}\left[|N_T - N_T^\Delta|\right]$ for $T = 5$ (blue), and the least-squares linear approximation (orange), with the equation being $y = 8.4 \cdot \Delta^{1.1}$.

*Proof.* Note that $R_t = \sum_{k=1}^{N_t} Y_k$ and $R_t^\Delta = \sum_{k=1}^{N_t^\Delta} Y_k$ for all $t \in [0, T]$. Moreover, $N$ and $N^\Delta$ are counting processes with finite expectation. Indeed,

$$\mathbb{E}[N_T] + \mathbb{E}[N_T^\Delta] = \int_0^T \left(\mathbb{E}[\lambda_s] + \mathbb{E}[\lambda_s^\Delta]\right) \mathrm{d}s < \infty,$$

according to Lemmas 5 and 4. Thus, the processes $R$ and $R^\Delta$ are linear combination of indicators of finite subinterval of the form $[a, T]$ with $a > 0$ of $[0, T]$ which belongs to $W_T^\eta \cap I_{0^+}^\eta(L^1([0, T]))$ according to Example 4.1 in [4]. □

We now give a bound on the difference between the aggregate risk observed for the continuous-time risk process and its discrete-time counterpart, which will be crucial in proving convergence in the fractional Sobolev space.

**Proposition 2.** *Let $T > \Delta > 0$. Let $(h, \psi, \nu, b)$ fulfill Assumptions 1 and 2. Let $R$ (or $R^\Delta$) be a continuous-time (or discrete-time) Hawkes risk process, defined by thinning from the same underlying Poisson measure $P$.*

*There exists a constant $K$ such that for all $0 \le s \le t \le T$*

$$\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right] \le K\left(C_R(h, \Delta)\left[(t)_\Delta - (s)_\Delta\right] + \Delta\right),$$

*where $C_R(h, \Delta)$ is given in Definition 3.*

The proof of this proposition is given in Appendix A. Before we move to strong functional convergence results, we give in Figure 4 a numerical illustration of Proposition 2 for the simple Hawkes process of kernel $h(t) = \frac{0.6 \cdot \cos(t)}{1+t^2}$.

We are now ready to give the first general strong convergence result for the discrete-time Hawkes risk processes.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

**Theorem 1.** *Let $T > 0$. Let $(h, \psi, \nu, b)$ fulfill Assumptions 1 and 2. Let $R$ (or $R^\Delta$) be a continuous-time (or discrete-time) Hawkes risk process. We have*

$$\mathbb{E}\left[\|R - R^\Delta\|_{W_T^\eta}\right] \leq K\left[T^2 C_R(h, \Delta) + T\Delta^{1-\eta}\right],$$

*where $C_R(h, \Delta)$ is defined in Definition 3 and $K$ is a positive multiplicative constant depending on $\eta$ that does not depend on $T$ or $\Delta$.*

**Remark 3.** The rate of convergence depends on the functional space in which we are envisioning the processes under study. The rate of convergence in $\Delta$ decreases with the parameter $\eta$ and vanishes when $\eta$ goes to 1.

The strong approximation result in Theorem 1 is not useful for numerical approximations, because the dependence of $C_R(h, \Delta)$ on $\Delta$ is not explicit a priori. It turns out that if the kernel is of finite $p$-variations, we can give the order of convergence in $\Delta$.

For $p \geq 1$, the $p$-variation of a function $f:[0, T] \to \mathbb{R}$ is defined as

$$\|f\|_{p\text{-var}} := \left(\sup_{\mathcal{D}} \sum_{t_i \in \mathcal{D}} |f(t_{i+1}) - f(t_i)|^p\right)^{1/p},$$

where $\mathcal{D}$ is the set of finite subdivisions of $[0, T]$.

The set of functions of finite $p$-variation contains piecewise Hölder functions of a Hölder index $p^{-1} \in (0, 1]$. As we will see in Corollary 1, these functions are regular enough to have more explicit rates of convergence. Using Lemma 9, we derive Corollary 1.

**Corollary 1.** *Let $T > 0$ and $p \geq 1$ and assume that Assumption 1 holds. Moreover, assume that the kernel $h$ is of a finite $p$-variation $\|h\|_{p\text{-var}}$. Then, for $\Delta$ small enough, Assumption 3 is fulfilled and*

$$C_R(h, \Delta) \leq K\left[\|h\|_{p\text{-var}} T^{\frac{p-1}{p}} \Delta^{\frac{1}{p}} + \Delta\|h\|_\infty\right].$$

*Moreover,*

$$\mathbb{E}\left[\|R - R^\Delta\|_{W_T^\eta}\right] \leq K_\eta\left[\|h\|_{p\text{-var}} T^{\frac{3p-1}{p}} \Delta^{\frac{1}{p}} + T\Delta^{1-\eta} + \Delta\|h\|_\infty\right]$$

*for some positive constant $K_\eta$ that does not depend on $T$ or $\Delta$.*

**Remark 4.** The parameter $p$ measures the regularity of the kernel $h$. The rate of convergence in $\Delta$ is a decreasing function of $p$, and in $T$ is an increasing function of $p$.

We now turn to the proof of Theorem 1.

*Proof.* First, we recall the result of Proposition 2:

$$\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right] \leq K\left[C_R(h, \Delta)\left((t)_\Delta - (s)_\Delta\right) + \Delta\right].$$

The norm of the difference between the continuous-time risk process and its discrete-time counterpart in the fractional Sobolev space is written as

$$\mathbb{E}\left[\|R - R^\Delta\|_{W_T^\eta}\right] = \int_0^T \mathbb{E}\left[|R_t - R_t^\Delta|\right] \mathrm{d}t + \int_0^T \int_0^T \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t - s|^{1+\eta}} \mathrm{d}t\mathrm{d}s.$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

We bound each of the three terms individually. We start with the second term. Thanks to the symmetry of the integrals with respect to $s$ and $t$,

$$\begin{aligned}
&\int_0^T \int_0^T \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&= \sum_{i,j=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} \int_{(j-1)\Delta}^{j\Delta\wedge T} \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&= \sum_{i=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} \int_{(i-1)\Delta}^{i\Delta\wedge T} \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&\quad + 2\sum_{i=1}^{M-1} \int_{(i-1)\Delta}^{i\Delta} \int_{i\Delta}^{(i+1)\Delta\wedge T} \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&\quad + 2\sum_{i=1}^{M-2} \sum_{j=i+2}^{M} \int_{(i-1)\Delta}^{i\Delta} \int_{(j-1)\Delta}^{j\Delta\wedge T} \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s.
\end{aligned}$$

For the first term, since $s$ and $t$ are in the same interval $[(i-1)\Delta, i\Delta)$, we have $R_t^\Delta - R_s^\Delta = 0$. Moreover,

$$\mathbb{E}\left[|R_t - R_s|\right] \le \mathbb{E}\left[\sum_{i=N_s+1}^{N_t} |Y_i|\right] = \mathbb{E}[|Y|]\mathbb{E}[N_t - N_s] = K\int_s^t \mathbb{E}[\lambda_u]\mathrm{d}u.$$

Using Lemma 5 and the fact that $\rho_h < 1$, we have

$$\begin{aligned}
\mathbb{E}\left[|R_t - R_s|\right] &\le \frac{K(t-s)}{1-\rho_h} \\
&\le K(t-s).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} \int_{(i-1)\Delta}^{i\Delta\wedge T} \frac{\mathbb{E}\left[|(R_t - R_s) - (R_t^\Delta - R_s^\Delta)|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s &= \sum_{i=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} \int_{(i-1)\Delta}^{i\Delta\wedge T} \frac{\mathbb{E}\left[|R_t - R_s|\right]}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&\le K\sum_{i=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} \int_{(i-1)\Delta}^{i\Delta\wedge T} \frac{|t-s|}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s \\
&\le K\sum_{i=1}^{M} \int_{(i-1)\Delta\wedge T}^{i\Delta\wedge T} \int_{(i-1)\Delta}^{t} \frac{\mathrm{d}s}{(t-s)^{\eta}} \mathrm{d}t \\
&\le K\sum_{i=1}^{M} \int_{(i-1)\Delta}^{i\Delta\wedge T} (t-(i-1)\Delta)^{1-\eta}\, \mathrm{d}t \\
&\le K_\eta T\Delta^{1-\eta}.
\end{aligned}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

When $s$ and $t$ are in adjacent bins, $R_t^\Delta - R_s^\Delta = \sum_{k=1}^{D_{n_t}^\Delta} Y_j$ in distribution, where $D_{n_t}^\Delta = \int_{((n_t-1)\Delta, n_t\Delta]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\theta\le\lambda_u^\Delta} P(\mathrm{d}u, \mathrm{d}\theta, \mathrm{d}y)$. Thus,

$$\begin{aligned}
&\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\int_{i\Delta}^{(i+1)\Delta\wedge T}\frac{\mathbb{E}\left[|(R_t-R_s)-(R_t^\Delta-R_s^\Delta)|\right]}{|t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s\\
&\qquad=\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\int_{i\Delta}^{(i+1)\Delta\wedge T}\frac{\mathbb{E}\left[\left|R_t-R_s-\sum_{j=1}^{D_{n_t}^\Delta-1}Y_j\right|\right]}{|t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s\\
&\qquad\le\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\int_{i\Delta}^{(i+1)\Delta\wedge T}\frac{\mathbb{E}\left[|R_t-R_s|\right]+\mathbb{E}[|Y|]\mathbb{E}\left[D_{n_t}^\Delta\right]}{|t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s.
\end{aligned}$$

And since $\mathbb{E}\left[D_{n_t}^\Delta\right]=\int_{(n_t-1)\Delta}^{n_t\Delta}\mathbb{E}\left[\lambda_u^\Delta\right]\mathrm{d}u$, we bound $\lambda_u^\Delta$ using Lemma 4 and the definition of $C_S(h,\Delta)$ given in Definition 3:

$$\begin{aligned}
&\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\int_{i\Delta}^{(i+1)\Delta\wedge T}\frac{\mathbb{E}\left[|(R_t-R_s)-(R_t^\Delta-R_s^\Delta)|\right]}{|t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s\\
&\qquad\le K\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\int_{i\Delta}^{(i+1)\Delta\wedge T}\frac{1}{(t-s)^\eta}+\frac{\Delta}{(t-s)^{1+\eta}}\mathrm{d}t\mathrm{d}s\\
&\qquad\le K\sum_{i=1}^{M-1}\int_{(i-1)\Delta}^{i\Delta}\Big(((i+1)\Delta-s)^{1-\eta}-(i\Delta-s)^{1-\eta}\\
&\qquad\quad-\Delta((i+1)\Delta-s)^{-\eta}+\Delta(i\Delta-s)^{-\eta}\Big)\mathrm{d}s\\
&\qquad=K\sum_{i=1}^{M-1}\int_0^\Delta(x+\Delta)^{1-\eta}-x^{1-\eta}-\Delta(x+\Delta)^{-\eta}+\Delta x^{-\eta}\mathrm{d}x\\
&\qquad\le K\sum_{i=1}^{M-1}\Delta^{2-\eta}\\
&\qquad\le K_\eta T\Delta^{1-\eta}.
\end{aligned}$$

We now treat the third-case scenario, when $s$ and $t$ are separated by more than one bin. Keeping in mind that $\Delta < t-s$ implies $((t)_\Delta-(s)_\Delta)\le(t-s+\Delta)\le 2(t-s)$ and using the result of Proposition 2 and Remark 3, we have

$$\begin{aligned}
&\sum_{i=1}^{M-2}\sum_{j=i+2}^{M}\int_{(i-1)\Delta}^{i\Delta}\int_{(j-1)\Delta}^{j\Delta\wedge T}\frac{\mathbb{E}\left[|(R_t-R_s)-(R_t^\Delta-R_s^\Delta)|\right]}{\mathbb{E}|Y_1||t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s\\
&\qquad\le K\sum_{i=1}^{M-2}\sum_{j=i+2}^{M}\int_{(i-1)\Delta}^{i\Delta}\int_{(j-1)\Delta}^{j\Delta\wedge T}\frac{C_R(h,\Delta)(t-s)+\Delta}{|t-s|^{1+\eta}}\mathrm{d}t\mathrm{d}s
\end{aligned}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\leq K \sum_{i=1}^{M-2} \int_{(i-1)\Delta}^{i\Delta} \int_{(i+1)\Delta}^{T} \frac{C_R(h,\,\Delta)(t-s)+\Delta}{|t-s|^{1+\eta}} \mathrm{d}t\mathrm{d}s$$

$$\leq K \sum_{i=1}^{M-2} \int_{(i-1)\Delta}^{i\Delta} C_R(h,\,\Delta)(t-s)^{1-\eta} + \Delta((i+1)\Delta - s)^{-\eta} \mathrm{d}s$$

$$\leq K_\eta \left[ C_R(h,\,\Delta) T^{2-\eta} + T\Delta^{1-\eta} \right].$$

For the first term, using the fact that $t_\Delta \leq t$, we have

$$\int_0^T \mathbb{E}\left[|R_t - R_t^\Delta|\right] \mathrm{d}t \leq K \int_0^T (C_R(h,\,\Delta) t_\Delta + \Delta) \mathrm{d}t$$

$$\leq K \int_0^T (C_R(\Delta) t + \Delta) \mathrm{d}t$$

$$\leq K \left[ C_R(\Delta) T^2 + \Delta T \right].$$

□

This yields the first explicit speed of convergence of the discrete-time marked Hawkes risk for a rich family of kernels. We point out that convergence rates can also be obtained for different kernels that do not lie in the set of functions with finite $p$-variations. A particularly interesting example is the Hawkes process driven by the kernel

$$h(t) = \frac{C}{\sqrt{t}} \mathbf{1}_{t \in (0,T]},$$

where $C > 0$ is a constant that ensures that Assumption 1 is in force.

Since this kernel is not of finite $p$-variation, it should be verified that Assumption 2 holds. This is possible because the sum of the inverses of the square root can be bounded by the integral of $\frac{1}{\sqrt{x}}$. Once this is done, we can apply Theorem 1. Using the fact that $h$ is decreasing on $(0,\, T]$, we can bound the modulus of continuity of the shift operator in an elementary fashion:

$$\int_0^{T-\Delta} |h(y+\epsilon) - h(y)| \mathrm{d}y = \int_0^{T-\Delta} h(y) - h(y+\epsilon) \mathrm{d}y$$

$$= \int_0^{T-\Delta} h(y) \mathrm{d}y - \int_0^{T-\Delta} h(y+\epsilon) \mathrm{d}y$$

$$= O(\Delta^{\frac{1}{2}}).$$

Hence,

$$\mathbb{E}\left[ \|R^\Delta - R\|_{W_T^\eta} \right] \leq K(T\Delta^{\frac{1}{2}} + T\Delta^{1-\eta}).$$

This rate is slower than the rate of convergence for Hawkes risks whose kernels are of bounded variations $K(T^2\Delta + T\Delta^{1-\eta})$, which is natural because of the singularity of the inverse square root near zero.

We now prove the convergence in the space of càdlàg functions equipped with the Skorokhod metric for a class of Hawkes processes.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

### 3.3 Convergence in the Skorokhod space

We start this section by recalling the definition of the Skorokhod space. We call $\mathbb{D}([0, T], \mathbb{R})$ the space of right continuous functions, with left limit (càdlàg). The canonical metric over this space is the Skorokhod metric defined by (5). The fact that this distance allows for small uncertainties in time, unlike the uniform distance, ensures that it is well adapted for Hawkes risk processes whether in continuous time or in discrete time. The different properties of the Skorokhod space can be found in [5]. Let $\Lambda$ denote the class of strictly increasing continuous mapping of $|0, T]$ into itself such that $\lambda(0) = 0$, $\lambda(T) = T$. For $x$ and $y$ in $\mathbb{D}([0, T])$,

$$d_S(x, y) = \inf_{\lambda \in \Lambda} \{\|\lambda - \mathrm{Id}\|_\infty \vee \|x - y\lambda\|_\infty\},$$

where Id is the identity map from $[0, T]$.

Since the jump times of the continuous-time Hawkes risk $R$ occur almost surely outside the time grid $\sigma = (k\Delta)_{k=1,\dots,M}$, the projection $(A_\sigma R)_{t\in[0,T]} := (R_{(t)_\Delta})_{t\in[0,T]}$ represents an intermediate process between $R$ and $R^\Delta$; hence,

$$d_S(R, R^\Delta) \le d_S(R, A_\sigma R) + d_S(A_\sigma R, R^\Delta).$$

The first term on the left-hand side simply evaluates the distance between the path of $R$ and its projection on the grid, and is bounded (see [5], Lemma 3, p. 127) by

$$d_S(R, A_\sigma R) \le \Delta \vee \omega'_R(\Delta),$$

where $\omega'_R(\Delta)$ is the $\Delta$ modulus of continuity for the càdlàg process

$$\omega'_X(\Delta) = \inf_{\Delta\text{-sparse}} \max_{1\le i\le M} \sup_{u,v\in[t_{i-1},t_i)} |X_u - X_v|,$$

the infimum being taken on the set of all partitions $\{t_i\}$ of $[0, T]$ such that $\min t_i - t_{i-1} > \Delta$. Therefore, the Skorokhod distance between the continuous-time Hawkes risk and its discrete-time counterpart is controlled by

$$d_S(R, R^\Delta) \le \Delta + \omega'_R(\Delta) + d_S(A_\sigma R, R^\Delta).$$

It is then enough to control the regularity of $R$ (using $(\omega'_R(\Delta))$ and its distance from $R^\Delta$ on the points of the grid $(d_S(A_\sigma R, R^\Delta))$. Indeed, noticing that both $A_\sigma R$ and $R^\Delta$ are constant on $[k\Delta, (k + 1)\Delta)$ for $k = 1, \dots, M$, we immediately see that

$$\begin{aligned} d_S(A_\sigma R, R^\Delta) &\le \|A_\sigma R - R^\Delta\|_\infty \\ &= \max_{k=1,\dots,M} |R_{k\Delta} - R^\Delta_{k\Delta}|, \end{aligned}$$

yielding

$$d_S(R, R^\Delta) \le \Delta + \omega'_R(\Delta) + \max_{k=1,\dots,M} |R_{k\Delta} - R^\Delta_{k\Delta}|. \tag{9}$$

Before giving an upper bound on the distance between the two processes evaluated on the points of the grid, we remind the reader that $\mathcal{F}$ is the filtration associated with the common underlying Poisson measure $P$.

**Proposition 3.** *Assume that Assumptions 1 and 3 or Assumptions 1 and 2 are in force. Assume also that $\nu$ has a finite second moment. Let*

$$\Xi_k := R_{k\Delta} - \mathbb{E}[Y] \int_0^{k\Delta} \lambda_s \mathrm{d}s$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

*and*

$$\Xi_k^\Delta := R_{k\Delta}^\Delta - \mathbb{E}[Y] \sum_{i=1}^{k} \lambda_{i\Delta}^\Delta \Delta.$$

*Then* $(\Xi_k)_{k=0,\dots,M}$ *and* $(\Xi_k^\Delta)_{k=0,\dots,M}$ *are* $(\mathcal{F}_{k\Delta})_{k=1,\dots,M}$*-martingales. Moreover,*

$$\mathbb{E}\Big[\max_{k=1,\dots,M} |\Xi_k - \Xi_k^\Delta|\Big] \le K\sqrt{TC_R(h,\Delta)},$$

*where K is a positive multiplicative constant that does not depend on T or* $\Delta$ *and* $C_R(h,\Delta)$ *is defined in Definition 3.*

The proof of this Proposition can be found in Appendix A.

We finally give an upper bound on the Skorokhod distance between the continuous-time Hawkes risk and its discrete-time counterpart in the case that the jump rate $\psi$ is bounded.

**Theorem 2.** *Suppose that Assumptions 1 and 3 or Assumptions 1 and 2 are in force. Assume furthermore that* $\psi$ *is bounded and that* $\nu$ *has a finite second moment.There exists a positive constant K that does not depend on T and* $\Delta$ *such that*

$$\mathbb{E}\left[d_S(R, R^\Delta)\right] \le K\left(\Delta\left(1 + \|\psi\|_\infty[1 + \|\psi\|_\infty^2 T^2]\right) + \sqrt{TC_R(h,\Delta)} + TC_R(h,\Delta)\right).$$

*Moreover, if h is of finite p-variations for* $p \ge 1$*, then Assumption 3 automatically holds and*

$$\mathbb{E}\left[d_S\left(R, R^\Delta\right)\right] \le K\Delta\left(1 + \|\psi\|_\infty\left(1 + T^2\|\psi\|_\infty^2\right)\right) + K\left(\sqrt{\|h\|_{p\text{-var}} T^{\frac{2p-1}{p}} \Delta^{\frac{1}{p}} + T\Delta\|h\|_\infty} + \|h\|_{p\text{-var}} T^{\frac{2p-1}{p}} \Delta^{\frac{1}{p}} + T\Delta\|h\|_\infty\right).$$

*Proof.* Taking the expected value of the inequality (9), we have

$$\begin{aligned}
\mathbb{E}\left[d_S\left(R, R^\Delta\right)\right] &\le \Delta + \mathbb{E}[\omega_R'(\Delta)] + \mathbb{E}\left[\max_{k=1,\dots,M} |R_{k\Delta} - R_{k\Delta}^\Delta|\right] \\
&\le \Delta + \mathbb{E}[\omega_R'(\Delta)] + \mathbb{E}\left[\max_{k=1,\dots,M} |\Xi_k - \Xi_k^\Delta|\right] \\
&\quad + \mathbb{E}\left[\max_{k=1,\dots,M} \left|\int_0^{k\Delta} \lambda_s ds - \sum_{i=1}^{k} \lambda_{i\Delta}^\Delta \Delta\right|\right] \\
&= \Delta + A_1 + A_2 + A_3.
\end{aligned}$$

The difference between the values for the discrete-time and continuous-time processes has been separated into a martingale part and a compensator part.

The martingale term $A_2$ has already been dealt with in Proposition 3. For $A_3$, we simply write

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\left|\int_0^{k\Delta}\lambda_s \mathrm{d}s - \sum_{i=1}^{k}\lambda_{i\Delta}^{\Delta}\Delta\right| = \left|\sum_{i=1}^{k}\int_{(i-1)\Delta}^{i\Delta}\lambda_s - \lambda_{i\Delta}^{\Delta}\mathrm{d}s\right|$$
$$\leq \sum_{i=1}^{M}\int_{(i-1)\Delta}^{i\Delta}|\lambda_s - \lambda_{i\Delta}^{\Delta}|\mathrm{d}s.$$

Taking the expectation of each term of the previous inequality and using the inequality (7), we obtain

$$A_3 \leq KTC_R(h,\,\Delta).$$

The last term is $A_1$. Since the jump rate is bounded by $\|\psi\|_\infty$, the compound Poisson process

$$\Pi_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} |y|\mathbf{1}_{\theta\leq\|\psi\|_\infty}P(\mathrm{d}s,\,\mathrm{d}\theta,\,\mathrm{d}y)$$

dominates the process $R$. That is, for any $0 \leq a \leq b \leq T$, $\Pi_b - \Pi_a \geq |R_b - R_a|$ . The problem now is to determine the behavior of the modulus of continuity of a compound Poisson process of intensity $\|\psi\|_\infty$, which is solved in Lemma 8. □

The case of a bounded jump rate is quite restrictive; for instance, the results of Theorem 2 cannot be applied to the standard linear Hawkes process. We therefore give a generalization to unbounded jump rates in the following theorem.

**Theorem 3.** *Suppose that Assumptions 1 and 3 or Assumptions 1 and 2 are in force. Assume that h is bounded and that ν has a finite second moment. There exists a positive constant K that does not depend on T and Δ such that*

$$\mathbb{E}\left[d_S\left(R,\,R^{\Delta}\right)\right] \leq K\left(\Delta^{\frac{1}{4}}\left(1+T^2\right) + \sqrt{TC_R(h,\,\Delta)} + TC_R(h,\,\Delta)\right).$$

*Moreover, if h is of finite p-variations for p ≥ 1, then Assumption 3 automatically holds and*

$$\mathbb{E}\left[d_S(R,\,R^{\Delta})\right] \leq K\left(\Delta^{\frac{1}{4}}\left(1+T^2\right) + T\Delta\right)$$
$$+ K\left(\sqrt{\|h\|_{p\text{-var}}T^{\frac{2p-1}{p}}\Delta^{\frac{1}{p}} + T\Delta\|h\|_\infty} + \|h\|_{p\text{-var}}T^{\frac{2p-1}{p}}\Delta^{\frac{1}{p}} + T\Delta\|h\|_\infty\right).$$

*Proof.* Let $C$ a positive real number to be fixed later and let $\psi^C = \psi \wedge C$. We also denote by $\lambda^C$, $N^C$, $R^C$ the intensity, Hawkes process and risk process, and by $\lambda^{C,\Delta}$, $N^{C,\Delta}$, $R^{C,\Delta}$ the discrete intensity, Hawkes process and risk process associated with $\psi^C$. On the event $\Gamma_C = \{\omega,\, \sup_{t\leq T}\lambda_t \leq C\}$, the process $\left(\lambda_t^C,\, N_t^C,\, R_t^C,\ \ t\in[0,\,T]\right)$ and $(\lambda_t,\, N_t,\, R_t,\ \ t\in[0,\,T])$ coincide, and on the event $\Gamma_{C,\Delta} = \left\{\omega,\, \sup_{t\leq T}\lambda_t^{\Delta} \leq C\right\}$, the processes $\left(\lambda_t^{C,\Delta},\, N_t^{C,\Delta},\, R_t^{C,\Delta},\ \ t\in[0,\,T]\right)$ and $\left(\lambda_t^{\Delta},\, N_t^{\Delta},\, R_t^{\Delta},\ \ t\in[0,\,T]\right)$ coincide.

On the event $\Omega \setminus \left(\Gamma_C \cap \Gamma_{C,\Delta}\right)$, we estimate

$$d_S\left(R,\,R^{\Delta}\right) \leq \sup_{s\leq T}|R_s| + \sup_{s\leq T}|R_s^{\Delta}| \leq \sum_{k=1}^{N_T}|Y_k| + \sum_{k=1}^{N_T^{\Delta}}|Y_k|.$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

Thus,

$$\mathbb{E}\left[d_S\left(R, R^{\Delta}\right)\right] = \mathbb{E}\left[d_S\left(R, R^{\Delta}\right)\mathbf{1}_{\Gamma_C \cap \Gamma_{C,\Delta}}\right] + \mathbb{E}\left[d_S\left(R, R^{\Delta}\right)\left(\mathbf{1}_{\Gamma_C^c \cup \Gamma_{C,\Delta}^c}\right)\right]$$
$$\leq \mathbb{E}\left[d_S\left(R^C, R^{C,\Delta}\right)\right] + \mathbb{E}\left[d_S\left(R, R^{\Delta}\right)\left(\mathbf{1}_{\Gamma_C^c} + \mathbf{1}_{\Gamma_{C,\Delta}^c}\right)\right].$$

Using Cauchy-Schwarz and Markov inequalities, we obtain

$$\mathbb{E}\left[d_S\left(R, R^{\Delta}\right)\right]$$
$$\leq \mathbb{E}\left[d_S\left(R^C, R^{C,\Delta}\right)\right] + \sqrt{4\mathbb{E}\left(\left[\sum_{k=1}^{N_T}|Y_k|\right]^2 + \left[\sum_{k=1}^{N_T^{\Delta}}|Y_k|\right]^2\right)\left(\mathbb{P}\left(\Gamma_C^c\right) + \mathbb{P}\left(\Gamma_{C,\Delta}^c\right)\right)}$$
$$\leq \mathbb{E}\left[d_S\left(R^C, R^{C,\Delta}\right)\right] + \frac{2}{C}\sqrt{\mathbb{E}\left(\left[\sum_{k=1}^{N_T}|Y_k|\right]^2 + \left[\sum_{k=1}^{N_T^{\Delta}}|Y_k|\right]^2\right)\mathbb{E}\left(\sup_{s\leq T}\lambda_s^2 + \sup_{s\leq T}(\lambda_s^{\Delta})^2\right)}.$$

We now bound each term under the square root.
First, since $N_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\theta\leq\lambda_s} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)$,

$$\mathbb{E}[N_T^2] \leq 2\int_0^T \mathbb{E}[\lambda_s]\mathrm{d}s + 2\mathbb{E}\left[\int_0^T \lambda_s \mathrm{d}s\right]^2$$
$$\leq 2T\sup_{s\leq T}\mathbb{E}[\lambda_s] + 2T^2\sup_{s\leq T}\mathbb{E}[\lambda_s^2].$$

According to Lemmas 5 and 7,

$$\mathbb{E}[N_T^2] \leq K\left(1+T^2\right). \tag{10}$$

Second, since $N_t^{\Delta} = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\theta\leq\lambda_s^{\Delta}} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)$,

$$\mathbb{E}[\left(N_T^{\Delta}\right)^2] \leq 2\int_0^T \mathbb{E}[\lambda_s^{\Delta}]\mathrm{d}s + 2\mathbb{E}\left[\int_0^T \lambda_s^{\Delta}\mathrm{d}s\right]^2$$
$$\leq 2T\sup_{s\leq T}\mathbb{E}[\lambda_s^{\Delta}] + 2T^2\sup_{s\leq T}\mathbb{E}[\left(\lambda_s^{\Delta}\right)^2].$$

According to Lemmas 4 and 6,

$$\mathbb{E}[\left(N_T^{\Delta}\right)^2] \leq K\left(1+T^2\right).$$

Third, recall that

$$\lambda_t = \psi\left(\int_0^{t-} h(t-s)\mathrm{d}\xi_s\right).$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

Thus, since $\psi$ is Lipschitz continuous,

$$\lambda_s \leq \psi(0) + L \int_0^{t-} h(t-s)\mathrm{d}\xi_s.$$

Since $\xi_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} b(y)\mathbf{1}_{\theta\leq\lambda_s} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y)$ is an increasing process and $h$ is bounded,

$$\lambda_s \leq \psi(0) + L\|h\|_\infty \xi_T.$$

Thus,

$$\sup_{s\leq T} \lambda_s^2 \leq 2\psi(0) + 2L^2\|h\|_\infty^2 \xi_T^2.$$

Then, taking the expectation of each term, we obtain

$$\mathbb{E}\left[\sup_{s\leq T} \lambda_s^2\right] \leq 2\psi(0) + 2L^2\|h\|_\infty^2 \mathbb{E}\left[\xi_T^2\right].$$

Using the same line as in the proof of the estimation of $\mathbb{E}[N_T^2]$, (10), we have

$$\mathbb{E}\left[\sup_{s\leq T} \lambda_s^2\right] \leq K\left(1+T^2\right). \tag{11}$$

Fourth, recall that $\lambda_t^\Delta = \psi\left(\sum_{k=1}^{n_t-1} h(n_t-k)X_k^\Delta\right)$. Since $\psi$ is Lipschitz continuous,

$$\lambda_t^\Delta \leq \psi(0) + L\sum_{k=1}^{n_t-1} h(n_t-k)X_k^\Delta.$$

Using the fact that $h$ is bounded and

$$X_n^\Delta = \int_{((n-1)\Delta, n\Delta]\times\mathbb{R}_+\times\mathbb{R}} b(y)\mathbf{1}_{\theta\leq l_n^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y),$$

we have

$$\sup_{s\leq T} \lambda_s^\Delta \leq \psi(0) + L\|h\|_\infty \int_{(0,T]\times\mathbb{R}_+\times\mathbb{R}} b(y)\mathbf{1}_{\theta\leq\lambda_s^\Delta} P(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}y).$$

Using the same lines as in the proof of the estimation (11), we obtain

$$\mathbb{E}\left[\sup_{s\leq T}\left(\lambda_s^\Delta\right)^2\right] \leq K\left(1+T^2\right).$$

Then for a universal constant $K$,

$$\mathbb{E}\left[d_S\left(R, R^\Delta\right)\right] \leq \mathbb{E}\left[d_S\left(R^C, R^{C,\Delta}\right)\right] + K\frac{1+T^2}{C}.$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

The quantity $\mathbb{E}\left[d_S\left(R^C, R^{C,\Delta}\right)\right]$ is bounded with the use of Theorem 2 for $\psi^C$ with $\|\psi^C\|_\infty \le C$,

$$\mathbb{E}\left[d_S\left(R, R^\Delta\right)\right] \le K\left(\Delta\left[1 + C\left(1 + C^2T^2\right)\right] + \sqrt{TC_R(h,\Delta)} + TC_R(h,\Delta)\right) + K\frac{1+T^2}{C}.$$

Taking $C = \frac{1}{\Delta^{\frac{1}{4}}}$, we derive

$$\mathbb{E}\left[d_S(R, R^\Delta)\right] \le K\Delta^{\frac{1}{4}}\left(1+T^2\right) + K\left(\sqrt{TC_R(h,\Delta)} + TC_R(h,\Delta)\right),$$

where $K$ is a constant independent of $T$ and $\Delta$. □

**Remark 5.** If in the proof of Theorem 3, we bound $\mathbb{P}(\Gamma_C^c)$ by $\frac{\mathbb{E}[\sup_{s\le T}\lambda_s]}{C}$ instead of $\frac{\mathbb{E}[\sup_{s\le T}\lambda_s^2]}{C^2}$, we obtain

$$\mathbb{E}\left[d_S\left(R, R^{C,\Delta}\right)\right] \le K\left(\Delta^{\frac{1}{7}}\left(1+T^{\frac{11}{7}}\right) + \sqrt{TC_R(h,\Delta)} + TC_R(h,\Delta)\right).$$

The power of $T$ in the constant is smaller than in the bound given in Theorem 3 but the rate of convergence in $\Delta$ is slower.

## 4. Conclusion

Using coupling arguments based on thinning from a given Poisson measure, we derived explicit bounds on the distance between the continuous-time embedding of nonlinear Poisson autoregression (referred to here as the discrete-time Hawkes process) and the standard continuous-time Hawkes process, both in the Sobolev space and in the Skorokhod space. Our bounds yield a quantitative generalization of the convergence result proven in [23]. More specifically, the speed of convergence is given both in the time step of the discretization $\Delta$ and in the time horizon $T$.

We point out that the convergence results proven in this article can be generalized to multivariate Hawkes processes involving cross-excitation and inhibition. The difference lies in the fact that we need to perform the thinning procedure from $d$ independent Poisson measures instead of one.

Another interesting development from the results shown in this article is their extension to SDEs involving both Brownian noise and state-dependent Hawkes jumps in the Skorokhod metric; see [27]. To the best of our knowledge, such results do not exist in the literature where jumps are supposed to follow a Poisson process [7]. We thus consider that they would constitute a valuable contribution to the approximation of jump-diffusion SDEs with potential numerical applications in many fields, such as in option pricing or neuroscience.

## Appendix A. Proofs of the Convergence Results

### A.1 Proof of Lemma 2

First, we compute the distance between the two intensities at a point of the grid. For a given $u \in [0, T)$ we write the expressions for the intensities in order to find an upper bound of $\mathbb{E}\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|$:

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\mathbb{E}\left[\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|\right] = \mathbb{E}\left[\left|\psi\left(\int_0^{(u)_\Delta} h\left((u)_\Delta - v\right) \mathrm{d}\xi_v\right) - \psi\left(\sum_{k=1}^{n_u-1} h_{n_u-k} X_k^\Delta\right)\right|\right]$$

$$\leq L\mathbb{E}\left[\left|\int_{(t_{n_u-1},(u)_\Delta]\times\mathbb{R}_+^2} h((u)_\Delta - v) b(y) \mathbf{1}_{\theta\leq\lambda_v} P(\mathrm{d}v, \mathrm{d}\theta, \mathrm{d}y)\right|\right]$$

$$+ L\mathbb{E}\left[\left|\sum_{k=1}^{n_u-1} \int_{(t_{k-1},t_k]\times\mathbb{R}_+^2} \left(h((u)_\Delta - v)\mathbf{1}_{\theta\leq\lambda_v} - h_{n_u-k}\mathbf{1}_{\theta\leq\lambda_{t_k}^\Delta}\right) b(y) P(\mathrm{d}v, \mathrm{d}\theta, \mathrm{d}y)\right|\right],$$

$$\mathbb{E}\left[\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|\right] \leq \mathbb{E}[b(Y)]L \int_{((u)_\Delta-\Delta,(u)_\Delta]} |h((u)_\Delta - v)| \mathbb{E}[\lambda_v] \mathrm{d}v$$

$$+ L\sum_{k=1}^{n_u-1} \mathbb{E}\left[\left|\int_{(t_{k-1},t_k]\times\mathbb{R}_+^2} \left(h((u)_\Delta - v)\mathbf{1}_{\theta\leq\lambda_v} - h((u)_\Delta - v)\mathbf{1}_{\theta\leq\lambda_{t_k}^\Delta}\right) b(y) P(\mathrm{d}v, \mathrm{d}\theta, \mathrm{d}y)\right|\right]$$

$$+ L\sum_{k=1}^{n_u-1} \mathbb{E}\left[\left|\int_{(t_{k-1},t_k]\times\mathbb{R}_+^2} \left(h((u)_\Delta - v)\mathbf{1}_{\theta\leq\lambda_{t_k}^\Delta} - h_{n_u-k}\mathbf{1}_{\theta\leq\lambda_{t_k}^\Delta}\right) b(y) P(\mathrm{d}v, \mathrm{d}\theta, \mathrm{d}y)\right|\right]$$

$$\leq K \int_{((u)_\Delta-\Delta,(u)_\Delta]} |h((u)_\Delta - v)| \mathbb{E}[\lambda_v] \mathrm{d}v$$

$$+ L\sum_{k=1}^{n_u-1} \int_{(t_{k-1},t_k]} |h((u)_\Delta - v)| \, \mathbb{E}\left[\left|\lambda_v - \lambda_v^\Delta\right|\right] \mathrm{d}v$$

$$+ L\sum_{k=1}^{n_u-1} \int_{(t_{k-1},t_k]} \left|h((u)_\Delta - v) - h_{n_u-k}\right| \mathbb{E}\left[\lambda_v^\Delta\right] \mathrm{d}v,$$

where $K$ is a constant independent of $T$ and $\Delta$. To apply the same induction used in the proof of Lemma 4, the quantity in the integral $\int_{(0,(u)_\Delta-\Delta]}$ and the term to the left $\mathbb{E}\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|$ should coincide. That is why we take the projection of $\lambda_v$ on the discretization grid. Therefore, using the upper bounds on $\mathbb{E}\lambda_v$ and $\mathbb{E}\lambda_v^\Delta$ (see Lemmas 5 and 4) we have

$$\mathbb{E}\left[\left|\lambda_{(u)_\Delta} - \lambda_u^\Delta\right|\right] \leq K\left(\frac{1}{1-\rho_h} \int_{((u)_\Delta-\Delta,(u)_\Delta]} |h((u)_\Delta - v)| \mathrm{d}v\right)$$

$$+ K\left(\frac{1}{1-\rho_{h,\Delta}} \sum_{k=1}^{n_u-1} \int_{(t_{k-1},t_k]} \left|h((u)_\Delta - v) - h_{n_u-k}\right| \mathrm{d}v\right)$$

$$+ \mathbb{E}[b(Y)]L \sum_{k=1}^{n_u-1} \int_{(t_{k-1},t_k]} |h((u)_\Delta - v)| \, \mathbb{E}\left[\left|\lambda_{t_{k-1}} - \lambda_{t_{k-1}}^\Delta\right|\right] \mathrm{d}v$$

$$+ \int_{(t_{k-1},t_k]} |h((u)_\Delta - v)| \, \mathbb{E}\left[\left|\lambda_v - \lambda_{(v)_\Delta}\right|\right] \mathrm{d}v$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$
\begin{aligned}
&\le K\left(\frac{1}{1-\rho_h}\int_0^{\Delta}|h(y)|\mathrm{d}y\right)\\
&\quad + K\left(\frac{1}{1-\rho_{h,\Delta}}\sum_{k=1}^{n_u-1}\int_{(t_{k-1},t_k]}\left|h((u)_\Delta - v) - h_{n_u-k}\right|\mathrm{d}v\right)\\
&\quad + \mathbb{E}[b(Y)]L\sum_{k=1}^{n_u-1}\mathbb{E}\left[|\lambda_{t_{k-1}} - \lambda^{\Delta}_{t_{k-1}}|\right]\int_{(t_{k-1},t_k]}|h((u)_\Delta - v)|\,\mathrm{d}v\\
&\quad + \frac{K}{1-\rho_h}\left(\int_0^{\Delta}|h(y)|\mathrm{d}y + \sup_{\epsilon\in[0,\Delta]}\int_0^{T-\Delta}|h(y+\epsilon)-h(y)|\,\mathrm{d}y\right),
\end{aligned}
$$

where we got the fourth line from Lemma 1 and the first line using a linear time change.

For the second line, we notice that $t_{k-1} = t_k - \Delta = (v)_\Delta$; hence,

$$
\begin{aligned}
\sum_{k=1}^{n_u-1}\int_{(t_{k-1},t_k]}\left|h((u)_\Delta - v) - h_{n_u-k}\right|\mathrm{d}v &= \sum_{k=1}^{n_u-1}\int_{(t_{k-1},t_k]}|h((u)_\Delta - v) - h((u)_\Delta - k\Delta)|\,\mathrm{d}v\\
&= \sum_{k=1}^{n_u-1}\int_{(t_{k-1},t_k]}|h((u)_\Delta - v) - h((u)_\Delta - (v)_\Delta + \Delta)|\,\mathrm{d}v\\
&= \int_0^{(u)_\Delta-\Delta}|h((u)_\Delta - v) - h((u)_\Delta - (v)_\Delta + \Delta)|\,\mathrm{d}v.
\end{aligned}
$$

Using the change of variables $y = (u)_\Delta - v$ and noticing that $(u)_\Delta$ is already on the discretization grid, we have

$$
\begin{aligned}
\sum_{k=1}^{n_u-1}\int_{(t_{k-1},t_k]}\left|h((u)_\Delta - v) - h_{n_u-k}\right|\mathrm{d}v &= \int_\Delta^{(u)_\Delta}|h(y) - h\left((u)_\Delta - ((u)_\Delta - y)_\Delta + \Delta\right)|\,\mathrm{d}y\\
&= \int_\Delta^{(u)_\Delta}|h(y) - h\left((y)_\Delta + \Delta\right)|\,\mathrm{d}y\\
&\le \int_0^{T-\Delta}|h(y) - h\left((y)_\Delta + \Delta\right)|\,\mathrm{d}y.
\end{aligned}
$$

For the third line we notice that $\int_{(t_{k-1},t_k]}|h((u)_\Delta - v)|\,\mathrm{d}v$ depends only on the difference $n_u - k$; indeed,

$$
\begin{aligned}
\mathbb{E}[b(Y)]L\int_{t_{k-1}}^{t_k}|h((u)_\Delta - v)|\,\mathrm{d}v &= \mathbb{E}[b(Y)]L\int_{(n_u-k)\Delta}^{(n_u-k+1)\Delta}|h(y)|\mathrm{d}y\\
&:= \eta^{\Delta}_{n_u-(k-1)}.
\end{aligned}
$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

By denoting $n_u = n$, $g_n = \mathbb{E}\left|\lambda_{n\Delta} - \lambda^{\Delta}_{n\Delta}\right|$, and $\tilde{\eta}^{\Delta}_j = \eta^{\Delta}_j \mathbf{1}_{j\geq 2}$, we have

$$g_n \leq C(h,\,\Delta) + \sum_{k=1}^{n-1} g_k \tilde{\eta}^{\Delta}_{n-k},$$

where

$$C(h,\,\Delta) = K\frac{1}{1-\rho_h}\left(\int_0^{\Delta} |h(y)|\mathrm{d}y + \sup_{\epsilon\in[0,\,\Delta]}\int_0^{T-\Delta} |h(y+\epsilon) - h(y)|\,\mathrm{d}y\right)$$
$$+ K\frac{1}{1-\rho_{h,\,\Delta}}\int_0^{T-\Delta} |h(y) - h\left((y)_{\Delta} + \Delta\right)|\,\mathrm{d}y.$$

Since $\tilde{\eta}^{\Delta} \in l_1$ and its sum is bounded by $\mathbb{E}[b(Y)]L\|h\|_1 < 1$, we apply the same induction as in the proof of Lemma 4 to obtain

$$\mathbb{E}\left[\left|\lambda_{(u)_{\Delta}} - \lambda^{\Delta}_u\right|\right] \leq \frac{C(h,\,\Delta)}{1-\rho_h}.$$

Using Definition 3, we finally conclude that

$$\mathbb{E}\left[\left|\lambda_{(u)_{\Delta}} - \lambda^{\Delta}_u\right|\right] \leq KC_R(h,\,\Delta)C_S(h,\,\Delta).$$

### A.2 Proof of Proposition 2

*Proof.* We first start by projecting both $s$ and $t$ on the time grid:

$$\mathbb{E}\left[|(R_t - R_s) - (R^{\Delta}_t - R^{\Delta}_s)|\right] = \mathbb{E}\left[|(R_t - R_{(t)_{\Delta}}) + (R_{(s)_{\Delta}} - R_s) + (R_{(t)_{\Delta}} - R_{(s)_{\Delta}}) - (R^{\Delta}_t - R^{\Delta}_s)|\right]$$
$$\leq \mathbb{E}\left[|R_t - R_{(t)_{\Delta}}|\right] + \mathbb{E}\left[|R_s - R_{(s)_{\Delta}}|\right]$$
$$+ \mathbb{E}\left[|(R_{(t)_{\Delta}} - R_{(s)_{\Delta}}) - (R^{\Delta}_{(t)_{\Delta}} - R^{\Delta}_{(s)_{\Delta}})|\right].$$

We handle the first two terms as follows:

$$\mathbb{E}\left[|R_t - R_{(t)_{\Delta}}|\right] \leq \mathbb{E}\left[\left|\int_{((t)_{\Delta},\,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\theta\leq\lambda_u} y P(\mathrm{d}u,\,\mathrm{d}\theta,\,\mathrm{d}y)\right|\right]$$
$$\leq \int_{\mathbb{R}_+} |y|\nu(\mathrm{d}y)\mathbb{E}\left[\int_{((t)_{\Delta},\,t]} \lambda_u \mathrm{d}u\right]$$
$$\leq \mathbb{E}[|Y|]\int_{((t)_{\Delta},\,t]} \mathbb{E}[\lambda_u]\mathrm{d}u$$
$$\leq \frac{K\Delta}{1-\rho_h}.$$

Hence,

$$\mathbb{E}\left[|R_t - R_{(t)_{\Delta}}|\right] + \mathbb{E}\left[|R_s - R_{(s)_{\Delta}}|\right] \leq K\frac{\Delta}{1-\rho_h}. \tag{A1}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

When it comes to the difference between the increments, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\left(R_{(t)_\Delta}-R_{(s)_\Delta}\right)-\left(R^\Delta_{(t)_\Delta}-R^\Delta_{(s)_\Delta}\right)\right|\right]\\
&\quad\le \mathbb{E}\left[\left|\int_{((s)_\Delta,(t)_\Delta]\times\mathbb{R}_+\times\mathbb{R}} y\left(\mathbf{1}_{\theta\le\lambda_u}-\mathbf{1}_{\theta\le\lambda^\Delta_u}\right)P(\mathrm{d}u,\mathrm{d}\theta,\mathrm{d}y)\right|\right]\\
&\quad\le \int_{\mathbb{R}_+}|y|\nu(\mathrm{d}y)\mathbb{E}\left[\int_{((s)_\Delta,(t)_\Delta]}\left|\lambda_u-\lambda^\Delta_u\right|\mathrm{d}u\right]\\
&\quad= \mathbb{E}[|Y|]\int_{((s)_\Delta,(t)_\Delta]}\mathbb{E}\left[\left|\lambda_u-\lambda^\Delta_u\right|\right]\mathrm{d}u\\
&\quad\le K\left(\int_{((s)_\Delta,(t)_\Delta]}\mathbb{E}\left[\left|\lambda_{(u)_\Delta}-\lambda^\Delta_u\right|\right]+\mathbb{E}\left[\left|\lambda_{(u)_\Delta}-\lambda_u\right|\right]\mathrm{d}u\right),
\end{aligned}$$

where the first term in the integral measures the discretization error on the grid and the second term measures the variation between a given point and its projection on the grid.

The second term in the integral can be bounded with the use of Lemma 1:

$$\begin{aligned}
&\int_{(s)_\Delta}^{(t)_\Delta}\mathbb{E}\left[\left|\lambda_{(u)_\Delta}-\lambda_u\right|\right]\mathrm{d}u\\
&\qquad\le\frac{K}{1-\rho_h}\left(\int_0^\Delta|h(y)|\mathrm{d}y+\sup_{\epsilon\in[0,\Delta]}\int_0^{T-\Delta}|h(y+\epsilon)-h(y)|\,\mathrm{d}y\right)((t)_\Delta-(s)_\Delta).
\end{aligned}\tag{A2}$$

The first term is upper-bounded with the use of Lemma 2:

$$\int_{(s)_\Delta}^{(t)_\Delta}\mathbb{E}\left[\left|\lambda_{(u)_\Delta}-\lambda^\Delta_u\right|\right]\mathrm{d}u\le KC_R(h,\Delta)\,((t)_\Delta-(s)_\Delta).\tag{A3}$$

Combining the inequalities (A1), (A2), and (A3), we obtain the result. □

### A.3. Proof of Proposition 3

Note that $\left(R_t-\mathbb{E}Y\int_0^t\lambda_s\mathrm{d}s\right)_t$ is a continuous-time $\mathcal{F}$-martingale and $(\Xi_k)_{k=0,\dots,M}$ is a discrete-time martingale. As for the discrete-time process, we have

$$\begin{aligned}
\Xi^\Delta_k&=R^\Delta_{k\Delta}-\mathbb{E}[Y]\sum_{i=1}^k\lambda^\Delta_{i\Delta}\Delta\\
&=\sum_{i=1}^k\int_{((i-1)\Delta,i\Delta]}\int_{\mathbb{R}_+\times\mathbb{R}}y\mathbf{1}_{\theta\le\lambda^\Delta_{i\Delta}}\left(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)-\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y)\right).
\end{aligned}$$

By construction we have $\int_{((i-1)\Delta,i\Delta]}\int_{\mathbb{R}_+\times\mathbb{R}}y\mathbf{1}_{\theta\le\lambda^\Delta_{i\Delta}}(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)-\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y))\in\mathcal{F}_{i\Delta}$ and $\lambda^\Delta_{i\Delta}\in\mathcal{F}_{(i-1)\Delta}$, which yields the result.

For the upper bound, we start by combining the Cauchy–Schwarz inequality with Doob's maximal inequality:

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\mathbb{E}\left[\max_{k=1,\dots,M}|\Xi_k-\Xi_k^\Delta|\right]\le\mathbb{E}\left[\max_{k=1,\dots,M}|\Xi_k-\Xi_k^\Delta|^2\right]^{1/2}$$
$$\le 2\sqrt{\mathbb{E}[|\Xi_M-\Xi_M^\Delta|^2]}.$$

Keeping in mind that $\Delta M = T$, we have

$$\Xi_M-\Xi_M^\Delta=\int_0^T\int_{\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\theta\le\lambda_s}\,(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)-\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y))$$
$$-\sum_{i=1}^M\int_{((i-1)\Delta,i\Delta]}\int_{\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\theta\le\lambda_{i\Delta}^\Delta}\,(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)-\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y))$$
$$=\int_0^T\int_{\mathbb{R}_+\times\mathbb{R}}\sum_{i=1}^M y\mathbf{1}_{(i-1)\Delta<s\le i\Delta}\left(\mathbf{1}_{\theta\le\lambda_s}-\mathbf{1}_{\theta\le\lambda_{i\Delta}^\Delta}\right)(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)-\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y)).$$

Therefore,

$$\mathbb{E}\left[|\Xi_M-\Xi_M^\Delta|^2\right]=\mathbb{E}\left[[\Xi-\Xi^\Delta]_T\right]$$
$$=\mathbb{E}\left[\int_0^T\int_{\mathbb{R}_+\times\mathbb{R}}\left(\sum_{i=1}^M y\mathbf{1}_{(i-1)\Delta<s\le i\Delta}(\mathbf{1}_{\theta\le\lambda_s}-\mathbf{1}_{\theta\le\lambda_{i\Delta}^\Delta})\right)^2\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y)\right]$$
$$=\mathbb{E}\left[\int_0^T\int_{\mathbb{R}_+\times\mathbb{R}}\sum_{i=1}^M y^2\mathbf{1}^2_{(i-1)\Delta<s\le i\Delta}\left(\mathbf{1}_{\theta\le\lambda_s}-\mathbf{1}_{\theta\le\lambda_{i\Delta}^\Delta}\right)^2\mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y)\right],$$

because $\mathbf{1}_{(i-1)\Delta<s\le i\Delta}\mathbf{1}_{(j-1)\Delta<s\le j\Delta}=0$ if $i\neq j$. Thus,

$$\mathbb{E}\left[|\Xi_M-\Xi_M^\Delta|^2\right]=\mathbb{E}[Y^2]\sum_{i=1}^M\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}[|\lambda_s-\lambda_{t_i}^\Delta|]\mathrm{d}t.$$

By adding and subtracting $\lambda_{t_i}$, we have for all $s\in((i-1)\Delta,\,i\Delta]$,

$$\mathbb{E}\left[|\lambda_s-\lambda_{\Delta i}^\Delta|\right]\le\mathbb{E}\left[|\lambda_s-\lambda_{\Delta i}|\right]+\mathbb{E}\left[|\lambda_{\Delta i}-\lambda_{\Delta i}^\Delta|\right].$$

Following the proof of Lemma 1 and projecting $s$ on the nearest upper point of the grid (instead of the lower), we have the bound

$$\mathbb{E}\left[|\lambda_s-\lambda_{\Delta i}|\right]\le K\left(\int_0^\Delta|h(y)|\mathrm{d}y+\sup_{\epsilon\in[0,\Delta]}\int_0^{T-\Delta}|h(y+\epsilon)-h(y)|\,\mathrm{d}y\right).$$

Thus, using the inequality (7) to bound $\mathbb{E}\left[|\lambda_{\Delta i}-\lambda_{\Delta i}^\Delta|\right]$, we obtain with the expression for $C_R(h,\,\Delta)$ given in Definition 3

$$\mathbb{E}\left[|\Xi_M-\Xi_M^\Delta|^2\right]\le KC_R(h,\,\Delta)T$$

for a constant $K$ that does not depend on $T$ or $\Delta$.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

## Appendix B. Technical Lemmata

### B.1. Bounds on the moments of the intensities

In this section, for completeness we prove some uniform estimations on the first and second moments of the continuous-time and discrete-time intensities.

**Lemma 4.** *Assume that Assumption 2 holds and* $\Delta \in (0, T]$. *Then*

$$\mathbb{E}\left[\lambda_t^{\Delta}\right] \leq \frac{\psi(0)}{1-\rho_{h,\Delta}}$$

*for any* $t \in (0, T]$.

*Proof.* Since $\lambda_t^{\Delta} = \lambda_{n_t\Delta}^{\Delta}$, the inequality on the expected value of the discrete-time intensity will be proven by induction on $n$. We have for any $0 \leq n \leq M$

$$\lambda_{n\Delta}^{\Delta} = \psi\left(\sum_{k=1}^{n-1} h_{n_t-k} X_k^{\Delta}\right).$$

In particular, for $n = 0, 1$,

$$\lambda_{n\Delta}^{\Delta} = \psi(0) \leq \frac{\psi(0)}{1-\rho_{h,\Delta}}.$$

Let $n \geq 2$ and assume that for all $j \leq n < M$

$$\mathbb{E}\left[\lambda_{j\Delta}^{\Delta}\right] \leq \frac{\psi(0)}{1-\rho_{h,\Delta}}.$$

Using the fact that $\psi$ is $L$-Lipschitz, we have

$$\begin{aligned}\lambda_{(n+1)\Delta}^{\Delta} &\leq \psi(0) + L\left|\sum_{k=1}^{n} h_{n+1-k} X_k^{\Delta}\right| \\ &\leq \psi(0) + L\sum_{k=1}^{n} |h|_{n+1-k} X_k^{\Delta}.\end{aligned}$$

Since $\mathbb{E}\left[X_k^{\Delta}\right] = \mathbb{E}\left[\mathbb{E}\left[X_k^{\Delta}|\lambda_{k\Delta}^{\Delta}\right]\right] = \mathbb{E}[b(Y)]\Delta \cdot \mathbb{E}\left[\lambda_{k\Delta}^{\Delta}\right]$ for any $k \geq 1$, we have

$$\begin{aligned}\mathbb{E}\left[\lambda_{(n+1)\Delta}^{\Delta}\right] &\leq \psi(0) + L\sum_{k=1}^{n} |h|_{n+1-k}\mathbb{E}\left[X_k^{\Delta}\right] \\ &= \psi(0) + L\sum_{k=1}^{n} |h|_{n+1-k}\mathbb{E}[b(Y)]\Delta \cdot \mathbb{E}\left[\lambda_{k\Delta}^{\Delta}\right],\end{aligned}$$

which, according to the induction's hypothesis, is bounded by

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\mathbb{E}\left[\lambda^{\Delta}_{(n+1)\Delta}\right] \le \psi(0) + \frac{L\psi(0)}{1-\mathbb{E}[b(Y)]L\sum_{k=1}^{M-1}|h|_k\Delta}\mathbb{E}[b(Y)]\sum_{k=1}^{n}|h|_{n+1-k}\Delta$$

$$\le \psi(0) + \frac{L\psi(0)}{1-\mathbb{E}[b(Y)]L\sum_{k=1}^{M-1}|h|_k\Delta}\mathbb{E}[b(Y)]\sum_{k=1}^{M-1}|h|_k\Delta,$$

which yields the result for $n+1 \le M$. □

**Lemma 5.** *Suppose that Assumption 1 is in force. Let $t \in [0,T]$. We have*

$$\mathbb{E}[\lambda_t] \le \frac{\psi(0)}{1-\rho_h}.$$

*Proof.* The facts that $\lambda_t$ is integrable and $t \mapsto \mathbb{E}(\lambda_t)$ is locally bounded follow from the same lines as in the proof of Theorem 1 in [6]. Using the fact that $\psi$ is $L$-Lipschitz, we have

$$\lambda_t = \psi\left(\int_0^{t-} h(t-s)\mathrm{d}\xi_s\right)$$

$$\le \psi(0) + L\left|\int_0^{t-} h(t-s)\mathrm{d}\xi_s\right|$$

$$\le \psi(0) + L\int_0^{t} |h|(t-s)\mathrm{d}\xi_s,$$

which yields by our taking the expected value

$$\mathbb{E}[\lambda_t] \le \psi(0) + \int_0^t L|h|(t-s)\mathbb{E}[b(Y)]\mathbb{E}[\lambda_s]\mathrm{d}s. \tag{B1}$$

Let $*$ denote the convolution operator $(f*g)(t) = \int_0^t f(s)g(t-s)\mathrm{d}s$ for any integrable $f$ and $g$. Let

$$S = \sum_{n\ge 1}\left(\mathbb{E}[b(Y)]L|h|\right)^{(n)}, \tag{B2}$$

where $\left(\mathbb{E}[b(Y)]L|h|\right)^{(n)} = \mathbb{E}[b(Y)]L|h| \underbrace{*\cdots*}_{n \text{ times}} \mathbb{E}[b(Y)]L|h|$. The stability condition $\mathbb{E}[b(Y)]L\|h\|_1 < 1$ ensures that $S$ is well defined and that $\|S\|_1 = \mathbb{E}[b(Y)]L\|h\|_1(1-\mathbb{E}[b(Y)]L\|h\|_1)^{-1}$.

For a given $k \ge 0$, convoling the inequality (B1) with $L|h|^{(k)}$, we obtain

$$\left(\left(\mathbb{E}[b(Y)]L|h|\right)^{(k)} * \mathbb{E}[\lambda]\right)(t) - \left(\left(\mathbb{E}[b(Y)]L|h|\right)^{(k+1)} * \mathbb{E}[\lambda]\right)(t) \le \left(\left(\mathbb{E}[b(Y)]L|h|\right)^{(k)} * \psi(0)\right)(t),$$

which yields after telescoping

$$\mathbb{E}[\lambda_t] - \left(\left(\mathbb{E}[b(Y)]L|h|\right)^{(n+1)} * \mathbb{E}[\lambda]\right)(t) \le \psi(0) + \sum_{k=1}^{n}\left(\left(\mathbb{E}[b(Y)]L|h|\right)^{(k)} * \psi(0)\right)(t).$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

Since $\mathbb{E}[\lambda_t]$ is finite, $(\mathbb{E}[b(Y)]L|h|)^{(n+1)} * \mathbb{E}[\lambda]$ converges to zero as $n$ tends to infinity, and finally

$$\begin{aligned}\mathbb{E}[\lambda_t] &\leq \psi(0) + \sum_{k=1}^{M} \left( (\mathbb{E}[b(Y)]L|h|)^{(k)} * \psi(0) \right)(t) \\ &= \psi(0) \left( 1 + \sum_{k=1}^{M} (\mathbb{E}[b(Y)]L|h|)^{(k)} * 1 \right) \\ &= \psi(0) \left( 1 + \frac{\mathbb{E}[b(Y)]L\|h\|_1}{1 - \mathbb{E}[b(Y)]L\|h\|_1} \right),\end{aligned}$$

hence the result. □

**Lemma 6.** *Assume that* $\rho_{h,\Delta} = L\mathbb{E}[b(Y)] \sum_{k=1}^{M} |h_k|\Delta < 1$ *and that* $\mathbb{E}[b(Y)^2] < \infty$. *Then*

$$\sup_{t\in[0,T]} \mathbb{E}\left[(\lambda_t^{\Delta})^2\right] \leq \left( \psi(0)^2 + L^2\mathbb{E}[b^2(Y)] \frac{\psi(0)}{1-\rho_{h,\Delta}} \|h\|_{\Delta,2}^2 \right) \frac{1}{(1-\rho_{h,\Delta})^2},$$

*where* $\|h\|_{\Delta,2} := \left( \sum_{j=1}^{M} |h_j|^2 \Delta \right)^{1/2}$.

*Proof.* The process $\lambda^{\Delta}$ is piecewise constant. Let us recall the definition of $\lambda_{\Delta n}^{\Delta} = l_n^{\Delta}$ given in Definition 2:

$$l_n^{\Delta} = \psi \left( \sum_{k=1}^{n-1} h_{n-k} X_k \right).$$

Since $\psi$ is $L$-Lipschitz continuous, we have

$$l_n^{\Delta} \leq \psi(0) + L \left( \sum_{k=1}^{n-1} |h_{n-k}| X_k \right).$$

Let us introduce $\tilde{X}_k = X_k - \Delta\mathbb{E}(b(Y))l_k^{\Delta}$; then

$$\begin{aligned}l_n^{\Delta} &\leq \psi(0) + L \left( \sum_{k=1}^{n-1} |h_{n-k}| \tilde{X}_k \right) + \sum_{k=1}^{n-1} L\Delta\mathbb{E}[b(Y)]|h_{n-k}| l_k^{\Delta} \\ &= \tilde{\nu}_n + \sum_{k=1}^{n-1} L\Delta\mathbb{E}[b(Y)]|h_{n-k}| l_k^{\Delta},\end{aligned}$$

where $\tilde{\nu}_n := \psi(0) + L \left( \sum_{k=1}^{n-1} |h_{n-k}| \tilde{X}_k \right)$. Since $l_n^{\Delta}$ is nonnegative, we have

$$\begin{aligned}l_n^{\Delta} &\leq \tilde{\nu}_n + \sum_{k=0}^{n} L\Delta\mathbb{E}[b(Y)]|h_{n-k}| l_k^{\Delta} \\ &= \tilde{\nu}_n + \left( L\Delta\mathbb{E}[b(Y)]|h| * l^{\Delta} \right)_n, \end{aligned} \tag{B3}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$*$ here being the discrete convolution operator for sequences defined on $\mathbb{N}$ and that take the value 0 for $n > M$. For a given $j \in \mathbb{N}$ we recursively define for any $n = 0, \ldots, M$

$$\begin{cases} a_n^{(0)} & = \mathbf{1}_{n=0}, \\ a_n^{(1)} & = a_n = L\Delta\mathbb{E}\left[b(Y)\right]|h_n|, \\ a_n^{(j)} & = (a^{(j-1)} * a)_n. \end{cases}$$

For a given integer $j$ we take the convolution of the inequality (B3) (whose right-hand side is nonnegative) with the nonnegative sequence $a^{(j)}$, yielding for any $n = 0, \ldots, M$

$$\left(a^{(j)} * l^{\Delta}\right)_n \leq \left(a^{(j)} * \tilde{\nu}\right)_n + \left(a^{(j+1)} * l^{\Delta}\right)_n,$$

and therefore by telescoping

$$l_n^{\Delta} - \left(a^{(j+1)} * l^{\Delta}\right)_n \leq \sum_{i=0}^{j+1} \left(a^{(i)} * \tilde{\nu}\right)_n.$$

Since $l_n^{\Delta}$ is almost surely finite for any $n = 0, \ldots, M$ (in fact it has a finite first moment) and since $\|a\|_1 = \|L\Delta\mathbb{E}[b(Y)]|h|\|_1 < 1$, we have

$$\begin{aligned} \left(a^{(j+1)} * l^{\Delta}\right)_n &\leq \|a^{(j+1)} * l^{\Delta}\|_1 \\ &\leq \|a\|_1^{j+1} \|l^{\Delta}\|_1 \xrightarrow[j\to+\infty]{} 0. \end{aligned}$$

Hence, for any $n = 0, \ldots, M$

$$l_n^{\Delta} \leq \tilde{\nu}_n + \sum_{k=0}^{n} A_{n-k}\tilde{\nu}_k,$$

where $A := \sum_{j=1}^{+\infty} a^{(j)}$ is well defined because $\|a\|_1 < 1$ and satisfies $\|A\|_1 \leq \frac{\|a\|_1}{1-\|a\|_1}$. Taking the square, we obtain

$$\begin{aligned} (l_n^{\Delta})^2 &\leq (\tilde{\nu}_n)^2 + 2\sum_{k=0}^{n} A_{n-k}\tilde{\nu}_k\tilde{\nu}_n + \left(\sum_{k=0}^{n} A_{n-k}\tilde{\nu}_k\right)^2 \\ &= (\tilde{\nu}_n)^2 + 2\sum_{k=0}^{n} A_{n-k}\tilde{\nu}_k\tilde{\nu}_n + 2\sum_{0\leq j<k\leq n} A_{n-k}A_{n-j}\tilde{\nu}_k\tilde{\nu}_j + \sum_{k=0}^{n} (A_{n-k})^2(\tilde{\nu}_k)^2. \end{aligned}$$

Using the definition of $\tilde{\nu}$ and the fact that $\tilde{X}$ is a martingale increment sequence, we have for $k \leq n$

$$\begin{aligned} \mathbb{E}\left[\tilde{\nu}_k\tilde{\nu}_n\right] &= \psi(0)^2 + L^2 \sum_{j=1}^{k} |h_{n-j}||h_{k-j}|\mathbb{E}[l_j^{\Delta}]\Delta\mathbb{E}[b^2(Y)] \\ &\leq \psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}} \sum_{j=1}^{k} |h_{n-j}||h_{k-j}|\Delta \end{aligned}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\leq \psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}}\left(\sum_{j=1}^{k}|h_{k-j}|^2\Delta\right)^{1/2}\left(\sum_{j=1}^{n}|h_{n-j}|^2\Delta\right)^{1/2}$$

$$\leq \psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}}\|h\|_{\Delta,2}^2,$$

where $\|h\|_{\Delta,2} := \left(\sum_{j=1}^{M}|h_j|^2\Delta\right)^{1/2}$. Since $\|A\|_1 \leq \frac{\rho_{h,\Delta}}{1-\rho_{h,\Delta}}$ we have

$$\mathbb{E}\left[\left(l_n^\Delta\right)^2\right] \leq \left(\psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}}\|h\|_{\Delta,2}^2\right)\left(1 + 2\frac{\rho_{h,\Delta}}{1-\rho_{h,\Delta}} + \left(\frac{\rho_{h,\Delta}}{1-\rho_{h,\Delta}}\right)^2\right)$$

$$= \left(\psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}}\|h\|_{\Delta,2}^2\right)\left(1 + \frac{\rho_{h,\Delta}}{1-\rho_{h,\Delta}}\right)^2$$

$$\leq \left(\psi(0)^2 + L^2\mathbb{E}[b^2(Y)]\frac{\psi(0)}{1-\rho_{h,\Delta}}\|h\|_{\Delta,2}^2\right)\frac{1}{(1-\rho_{h,\Delta})^2}.$$

□

**Lemma 7.** *Under Assumption 1 and* $\mathbb{E}b(Y)^2 < \infty$,

$$\sup_{t\in[0,T]}\mathbb{E}[\lambda_t^2] \leq \left(\psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\|h\|_2^2\right)\frac{1}{(1-\rho_h)^2}.$$

**Remark 6.** Unlike Theorem 2.4 in [20] in which an exact expression for $\mathbb{E}[\lambda_t^2]$ is given, we provide here an upper bound on that quantity. Our result has the advantage of being explicit in the parameters of the Hawkes process and of illustrating that the second moment is also bounded in $t$ when the stability condition is met.

*Proof.* Recall from Definition 1 that

$$\lambda_t = \psi\left(\int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}} h(t-s)\mathbf{1}_{\theta\leq\lambda_s}b(y)P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)\right).$$

Since $\Psi$ is Lipschitz continuous, we have

$$\lambda_t \leq \psi(0) + L\int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}}|h(t-s)|\mathbf{1}_{\theta\leq\lambda_s}b(y)P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y)$$

$$\leq \tilde{M}_t + L\mathbb{E}[b(Y)]\int_0^t|h(t-s)|\lambda_s ds,$$

where

$$\tilde{M}_t = \psi(0) + L\int_{[0,t)\times\mathbb{R}_+\times\mathbb{R}}|h(t-s)|\mathbf{1}_{\theta\leq\lambda_s}b(y)\,(P(\mathrm{d}s,\mathrm{d}\theta,\mathrm{d}y) - \mathrm{d}s\mathrm{d}\theta\nu(\mathrm{d}y)).$$

Using the same lines as in the proof of Lemma 5 or Lemma 3 in [3], we have

$$\lambda_t \leq \tilde{M}_t + \int_0^t S(t-s)\tilde{M}_s\mathrm{d}s,$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

where $S$ is defined in (B2). Then

$$\lambda_t^2 \le \tilde{M}_t^2 + 2\int_0^t S(t-s)\tilde{M}_s\tilde{M}_t \mathrm{d}s + \left(\int_0^t S(t-s)\tilde{M}_s \mathrm{d}s\right)^2$$
$$\le \tilde{M}_t^2 + 2\int_0^t S(t-s)\tilde{M}_s\tilde{M}_t \mathrm{d}s + \int_{[0,t]^2} S(t-s)S(t-u)\tilde{M}_s\tilde{M}_u \mathrm{d}s\mathrm{d}u.$$

According to the definition of $\tilde{M}_t$, the fact that $h$ is bounded, $\rho_h = L\mathbb{E}[b(Y)]\|h\|_1 < 1$, and Lemma 1,

$$\mathbb{E}\left[\tilde{M}_s\tilde{M}_u\right] = \psi(0)^2 + L^2\int_0^{\min(u,s)} \mathbb{E}[b(Y)^2]|h(s-r)||h(u-r)|\mathbb{E}[\lambda_r]\mathrm{d}r$$
$$\le \psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\int_0^{\min(u,s)} |h(s-r)||h(u-r)|\mathrm{d}r$$
$$\le \psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\left(\int_0^s h^2(s-r)\mathrm{d}r\right)^{1/2}\left(\int_0^u h^2(u-r)\mathrm{d}r\right)^{1/2}$$
$$= \psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\|h\|_2^2.$$

Since $\|S\|_1 = \rho_h[1-\rho_h]^{-1}$ we have

$$\mathbb{E}[\lambda_t^2] \le \left(\psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\|h\|_2^2\right)\left(1 + \frac{2\rho_h}{1-\rho_h} + \frac{\rho_h}{(1-\rho_h)^2}\right)$$
$$= \left(\psi(0)^2 + L^2\mathbb{E}[b(Y)^2]\frac{\psi(0)}{1-\rho_h}\|h\|_2^2\right)\frac{1}{(1-\rho_h)^2}.$$

□

### B.2. Estimation on the modulus of continuity of compound Poisson processes

**Lemma 8.** *Let $N$ be a Poisson process of intensity $I$ on [0,1] and $R_t = \sum_{k=0}^{N_t} |Y_k|$, where $(Y_k)_k$ is a sequence of i.i.d. random variables with common distribution $\nu$ such that $\int_{\mathbb{R}} |y|\nu(\mathrm{d}y) < +\infty$. Then its average modulus of continuity in $\mathbb{D}([0,T],\mathbb{R})$ is bounded by*

$$\mathbb{E}\left[\omega_R'(\Delta,[0,T])\right] \le \mathbb{E}[|Y_1|]IT\frac{\Delta}{T}(5+4I^2T^2) = \mathbb{E}[|Y_1|]I\Delta(5+4I^2T^2).$$

*Proof.* First, we assume that $T=1$. Let $\tau_1, \tau_2, \ldots$ denote the arrival times of the Poisson process and $S_1, S_2, \ldots$ be the interarrival times. Using the law of total probability, we have

$$\mathbb{E}\left[\omega_R'(\Delta,[0,1])\right] = \sum_{n=1}^{+\infty}\mathbb{E}[\omega_R'(\Delta)|N_T=n]\mathbb{P}[N_T=n]$$
$$= \sum_{n=1}^{+\infty}\mathbb{E}_n[\omega_R'(\Delta)]\frac{(I)^n}{n!}e^{-I}.$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

Knowing that $N_T = n$, the distribution of the arrival times is that of the order statistics of $n$ uniform random variables on [0,1], that is, of density $p(\tau_1 = t_1, \dots, \tau_n = t_n) = \mathbf{1}_{0 \le t_1 \le \dots \le t_n \le 1} n!$. Using an affine change of variables, we obtain a similar formula for the density of the interarrival times:

$$p(S_1 = s_1, \dots, S_n = s_n) = \mathbf{1}_{0 \le s_1 + \dots + s_n \le 1} n! \prod_{i=1}^{n} \mathbf{1}_{s_i \ge 0}.$$

Two scenarios are possible:

(i) All of the interarrival times are greater than $\Delta$ (which is possible only if $n\Delta \le T$). In this case we have only two possibilities:

- $\omega'_R(\Delta, [0, 1]) = |Y_n|$ if the last arrival time $\tau_n$ is at a distance less than $\Delta$ from 1.
- $\omega'_R(\Delta, [0, 1]) = 0$ otherwise.

(ii) At least one interarrival time $S_i$ for $i = 1, \dots, n$ is less than $\Delta$. In this case, the worst-case scenario is to have all of the jumps in one interval of size at most $\Delta$, yielding

$$\omega'_R(\Delta, [0, 1]) \le \sum_{k=1}^{n} |Y_k|.$$

Hence, we have

$$\begin{aligned}\mathbb{E}_n \left[\omega'_R(\Delta, [0, 1])\right] &\le \mathbb{E}_n[|Y_n|]\mathbb{P}_n[A_1] + \sum_{k=1}^{n} \mathbb{E}_n[|Y_k|]\mathbb{P}_n[A_2] \\ &= \mathbb{E}[|Y_1|]\mathbb{P}_n[A_1] + n\mathbb{E}[|Y_1|]\mathbb{P}_n[A_2],\end{aligned}$$

where $A_1 = \{S_i \ge \Delta \text{ for all } i = 1, \dots, n \text{ and } 1 - \tau_n < \Delta\}$ and $A_2 = \{\text{there exists } i \in [1, n], S_i \le \Delta\}$.

Keeping in mind that $\tau_n$ is the maximum of $n$ uniform i.i.d. variables on [0,1], we have

$$\begin{aligned}\mathbb{P}_n[A_1] &\le \mathbb{P}_n[1 - \Delta < \tau_n] \\ &= 1 - \mathbb{P}_n[1 - \Delta > \tau_n] \\ &= 1 - \mathbb{P}_n[1 - \Delta > \mathcal{U}[0, 1]]^n \quad \text{because } \tau_n = \max_{i=1,\dots,n} U_i \\ &= 1 - (1 - \Delta)^n \\ &\le (n\Delta) \wedge 1.\end{aligned}$$

For the event $A_2$ we have

$$\begin{aligned}\mathbb{P}_n[A_2] &= \mathbb{P}_n[\min_{1 \le i \le n} (S_i) \le \Delta] \\ &= \int_{\mathbb{R}_+^n} \mathbf{1}_{\min_{1 \le i \le n}(s_i) \le \Delta} \mathbf{1}_{0 \le s_1 + \dots + s_n \le 1} n! \mathrm{d}s_1 \dots \mathrm{d}s_n\end{aligned}$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$
\begin{aligned}
&\leq \sum_{i=1}^{n} \int_{0\leq s_1+\cdots+s_n\leq 1} \mathbf{1}_{s_i\leq\Delta} n! \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&= n \int_{0\leq s_1+\cdots+s_n\leq 1} \mathbf{1}_{s_1\leq\Delta} n! \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&\leq nn! \int_0^{\Delta} \mathrm{d}s_1 \int_{0\leq s_2+\cdots+s_n\leq 1} \mathrm{d}s_2 \ldots \mathrm{d}s_n \\
&= nn!\Delta \frac{1}{(n-1)!} \\
&= n^2\Delta.
\end{aligned}
$$

Therefore,

$$
\mathbb{E}_n\left[\omega'_R(\Delta, [0,1])\right] \leq \left[n\mathbb{E}[|Y_1|] + n\mathbb{E}[|Y_1|]n^2\right]\Delta.
$$

By averaging over the Poisson variable, we obtain

$$
\begin{aligned}
\mathbb{E}[\omega'_R(\Delta, [0,1])] &\leq \mathbb{E}[|Y_1|]\Delta(2+3I+I^2)I \\
&\leq \mathbb{E}[|Y_1|]\Delta I(5+4I^2). \qquad \text{(B4)}
\end{aligned}
$$

We now define the time-scaled process $R^T_v = R_{vT}$, where $v \in [0, T]$. The process $R^T$ is also a compound Poisson process of intensity $IT$. Using the fact that $\Delta$-sparse subdivisions of $[0, T]$ are exactly the $\frac{\Delta}{T}$-sparse subdivisions of [0,1] multiplied by $T$, we have

$$
\begin{aligned}
\omega'_R(\Delta, [0,T]) &= \inf_{\Delta\text{-sparse}} \max_{1\leq i\leq K} \sup_{u,v\in[t_{i-1},t_i)} |R_u - R_v| \\
&= \inf_{\Delta\text{-sparse}} \max_{1\leq i\leq K} R_{t_i-} - R_{t_{i-1}} \\
&= \inf_{\Delta\text{-sparse}} \max_{1\leq i\leq K} R_{T\frac{t_i-}{T}} - R_{T\frac{t_{i-1}}{T}} \\
&= \inf_{\frac{\Delta}{T}\text{-sparse}} \max_{1\leq i\leq K} R^T_{s_i-} - R^T_{s_{i-1}} = \omega'_{R^T}\left(\frac{\Delta}{T}, [0,1]\right).
\end{aligned}
$$

We now take the expected value and use the upper bound (B4) with intensity $IT$, time step $\frac{\Delta}{T}$, and time horizon 1 to obtain

$$
\mathbb{E}\left[\omega'_R(\Delta, [0,T])\right] \leq \mathbb{E}[|Y_1|]IT\frac{\Delta}{T}(5+4I^2T^2) = \mathbb{E}[|Y_1|]I\Delta(5+4I^2T^2).
$$

□

### B.3. Estimation for finite *p*-variation kernels

We now give a more exploitable bound instead of $C_R(\Delta)$ for a class of kernels.

**Lemma 9.** *Let $T > 0$ and $p \geq 1$. Assume that $h$ is of bounded $p$-variation on $[0, T]$. Then there exists a constant $K$, independent of $T$, $h$, and $\Delta$, such that*

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

$$\int_0^{\Delta} |h(y)|\mathrm{d}y \le \Delta \|h\|_{\infty},$$

$$\sup_{\varepsilon \le \Delta} \int_0^T |h(t+\epsilon) - h(t)|\mathrm{d}t \le K\|h\|_{p\text{-var},T} \left( T^{\frac{p-1}{p}} \Delta^{\frac{1}{p}} + \Delta \right),$$

$$\int_0^T |h(y) - h\left((y)_{\Delta}\right)|\mathrm{d}y \le K\|h\|_{p\text{-var},T} \left( T^{\frac{p-1}{p}} \Delta^{\frac{1}{p}} + \Delta \right),$$

*where* $\|h\|_{p\text{-var},T}$ *is the p-variation seminorm of h on [0,T] and* $\|h\|_{\infty,T} = \sup_{0\le t\le T} |h(t)|$. *Moreover, for* $\Delta$ *small enough, Assumption 2 is fulfilled.*

*Proof.* We start the proof by providing an upper bound on the modulus of continuity of the shift operator in $L^1$, along the lines of Lemma A.1 in [22]. Let $0 < \epsilon \le \Delta$. Then

$$\begin{aligned}\int_0^T |h(t+\epsilon) - h(t)|\mathrm{d}t &= \sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} \int_{(j-1)\epsilon}^{j\epsilon \wedge T} |h(t+\epsilon) - h(t)|\mathrm{d}t \\ &= \sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} \int_0^{\epsilon} |h((t+j\epsilon)\wedge T) - h(t+(j-1)\epsilon)|\mathrm{d}t \\ &= \int_0^{\epsilon} \sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} |h((t+j\epsilon)\wedge T) - h(t+(j-1)\epsilon)|\mathrm{d}t.\end{aligned}$$

Using Hölder's inequality, we have for $t \in [0, \epsilon]$ and $\frac{1}{q} = 1 - \frac{1}{p}$,

$$\begin{aligned}&\sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} |h((t+j\epsilon)\wedge T) - h(t+(j-1)\epsilon)| \\ &\qquad \le \left( \sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} |h((t+j\epsilon)\wedge T) - h(t+(j-1)\epsilon)|^p \right)^{1/p} \left( \sum_{j=1}^{\lfloor T/\epsilon \rfloor + 1} 1 \right)^{1/q} \\ &\qquad \le \|h\|_{p\text{-var},T} \left( \lfloor T/\epsilon \rfloor + 1 \right)^{\frac{p-1}{p}} \\ &\qquad \le K\|h\|_{p\text{-var},T} \left( \left( \frac{T}{\epsilon} \right)^{\frac{p-1}{p}} + 1 \right).\end{aligned}$$

By integration from 0 to $\epsilon$ we get

$$\int_0^T |h(t+\epsilon) - h(t)|\mathrm{d}t \le K\|h\|_{p\text{-var},T} \left( T^{\frac{p-1}{p}} \epsilon^{\frac{1}{p}} + \epsilon \right),$$

which by our taking the supremum of $\epsilon$ between 0 and $\Delta$ yields

$$\sup_{\epsilon \in [0,T]} \int_0^T |h(t+\epsilon) - h(t)|\mathrm{d}t \le K\|h\|_{p\text{-var},T} \left( T^{\frac{p-1}{p}} \Delta^{\frac{1}{p}} + \Delta \right).$$

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

In a similar fashion, we also show that

$$\int_0^T |h(y) - h\left((y)_\Delta\right)| \leq K\|h\|_{p\text{-var},T}\left(T^{\frac{p-1}{p}}\Delta^{\frac{1}{p}} + \Delta\right).$$

Indeed,

$$\begin{aligned}\int_0^T |h(y) - h((y)_\Delta)|\mathrm{d}y &= \sum_{k=1}^{M}\int_{t_{k-1}}^{t_k} |h(y) - h((y)_\Delta)|\mathrm{d}y \\ &= \int_0^\Delta \sum_{k=1}^{M} |h((t_{k-1}+r)\wedge T) - h(t_{k-1})|\mathrm{d}r \\ &\leq \int_0^\Delta \sum_{k=1}^{M} |h((t_{k-1}+r)\wedge T) - h(t_{k-1})| + |h(t_k\wedge T) \\ &\quad - h((t_{k-1}+r)\wedge T)|\mathrm{d}r.\end{aligned}$$

Thus, we also show that

$$\int_0^T |h(y) - h\left((y)_\Delta\right)| \leq K\|h\|_{p\text{-var},T}\left(T^{\frac{p-1}{p}}\Delta^{\frac{1}{p}} + \Delta\right).$$

In a similar fashion, we can show that

$$\int_0^{T-\Delta} |h((t)_\Delta + \Delta) - h(t)|\mathrm{d}t \leq \|h\|_{p\text{-var},T}T^{\frac{p-1}{p}}\Delta^{\frac{1}{p}}.$$

This means that $\lim_{\Delta\to 0}\int_0^{T-\Delta} |h((t)_\Delta + \Delta) - h(t)|\mathrm{d}t = 0$, and hence, thanks to the inequality (8) we have that Assumption 2 is in force. □

## Acknowledgements

We thank the anonymous reviewers and the associate editor for their insightful comments and constructive suggestions, which significantly improved the manuscript.

## Funding information

This work was supported by the ANR EDDA Project-ANR-20-IADJ-0003. Mahmoud Khabou acknowledges support from EPSRC NeST Programme grant EP/X002195/1.

## Competing interests

There are no competing interests to declare that arose during the preparation of or the publication process for this article.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

## References


[1] ALZAID, A. A. AND AL-OSH, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* **27**(2), 314–324. issn: 00219002. Available at http://www.jstor.org/stable/3214650 (visited on 12/27/2023).

[2] ARMILLOTTA, M. AND FOKIANOS, K. (2024). Count network autoregression. *J. Time Series Anal.* **45**(4), 584–612. issn: 0143-9782,1467-9892. Available at https://doi.org/10.1111/jtsa.12728.

[3] BACRY, E., DELATTRE, S., HOFFMANN, M. AND MUZY, J. F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123**(7). A Special Issue on the Occasion of the 2013 International Year of Statistics, pp. 2475–2499. issn: 0304-4149. doi: 10.1016/j.spa.2013.04.007. Available at https://www.sciencedirect.com/science/article/pii/S0304414913001026.

[4] BERGOUNIOUX, M., LEACI, A., NARDI, G. AND TOMARELLI, F. (2017). Fractional Sobolev spaces and functions of bounded variation of one variable. *Fract. Calc. Appl. Anal.* **20**(4), 936–962. issn: 1311-0454,1314-2224. Available at https://doi.org/10.1515/fca-2017-0049.

[5] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. Wiley Series in Probability and Statistics. Wiley-Interscience Publication. John Wiley & Sons, Inc., New York. Available at https://doi.org/10.1007/978-3-662-47507-2.

[6] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24**(3), 1563–1588. issn: 0091-1798,2168-894X. Available at https://doi.org/10.1214/aop/1065725193.

[7] BRUTI-LIBERATI, N. AND PLATEN, E. (2007). Strong approximations of stochastic differential equations with jumps. *J. Comput. Appl. Math.* **205**(2). Special issue on evolutionary problems, pp. 982–1001. issn: 0377-0427. doi: 10.1016/j.cam.2006.03.040. Available at https://www.sciencedirect.com/science/article/pii/S0377042706004183 .

[8] COUTIN, L. AND DECREUSEFOND, L. (2013). Stein's method for Brownian approximations. *Commun. Stoch. Anal.* **7**(3), 349–372. issn: 2688-6669. Available at https://doi.org/10.31390/cosa.7.3.01.

[9] COUTIN, L., DECREUSEFOND, L. AND HUANG, L. (2025). Rate of convergence in the functional central limit theorem for stable processes. *Potential Anal.* **63**, 1–25.

[10] DASSIOS, A. AND ZHAO, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Probab.* **18**(62), 13. issn: 1083-589X. Available at https://doi.org/10.1214/ECP.v18-2717.

[11] DAW, A. AND PENDER, J. (2018). Queues driven by Hawkes processes. *Stoch. Syst.* **8**(3), 192–229. issn: 1946-5238. Available at https://doi.org/10.1287/stsy.2018.0014.

[12] DUARTE A., LAXA, K., LÖCHERBACH, E. AND LOUKIANOVA, D. (2025). *Nonparametric estimation of the jump rate in mean field interacting systems of neurons.* Available at https://arxiv.org/abs/2506.24065.

[13] ERRAIS, E., GIESECKE, K. AND GOLDBERG, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financ. Math.* **1**(1), 642–665.

[14] FERLAND, R., LATOUR, A. AND ORAICHI, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal.* **27**(6), 923–942. issn: 0143-9782,1467-9892. Available at https://doi.org/10.1111/j.1467-9892.2006.00496.x.

[15] FOKIANOS, K., RAHBEK, A. AND TJØSTHEIM, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.* **104**(488). With electronic supplementary materials available online, pp. 1430–1439. issn: 0162-1459,1537-274X. Available at https://doi.org/10.1198/jasa.2009.tm08270.

[16] FOKIANOS, K. AND TJØSTHEIM, D. (2012). Nonlinear Poisson autoregression. *Ann. Inst. Statist. Math.* **64**(6), 1205–1225. issn: 0020-3157,1572-9052. Available at https://doi.org/10.1007/s10463-012-0351-3.

[17] FOKIANOS, K., STØVE, B., TJØSTHEIM, D. AND DOUKHAN, P. (2020). Multivariate count autoregression. *Bernoulli* **26**(1), 471–499. issn: 1350-7265,1573-9759. Available at https://doi.org/10.3150/19-BEJ1132.

[18] GRAHAM, C. (1992). McKean-Vlasov Itô-Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stochastic Process. Appl.* **40**(1), 69–82. issn: 0304-4149,1879-209X. Available at https://doi.org/10.1016/0304-4149(92)90138-G.

[19] HAWKES, A. G. (1971). Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. Ser. B (Methodol.)* **33**(3), 438–443. issn:00359246. Available at http://www.jstor.org/stable/2984686 (visited on 12/27/2023).

[20] HILLAIRET, C. AND RÉVEILLAC, A. (2023). *Explicit correlations for the Hawkes processes*. arXiv: 2304.02376 [math.PR].

[21] HILLAIRET, C., RÉVEILLAC, A. AND ROSENBAUM, M. (2023). An expansion formula for Hawkes processes and application to cyber-insurance derivatives. *Stochastic Process. Appl.* **160**, 89–119. issn: 0304-4149. doi: 10.1016/j.spa.2023.02.012. Available at https://www.sciencedirect.com/science/article/pii/S0304414923000455.

[22] HOLDEN, H. AND RISEBRO, N. H. (2015). *Front Tracking for Hyperbolic Conservation Laws*, 2nd edn., Vol. 152. Applied Mathematical Sciences. Springer, Heidelberg, pp. xiv+515. isbn: 978-3-662-47506-5; 978-3-662-47507-2. Available at https://doi.org/10.1007/978-3-662-47507-2.


https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press

[23] HUANG, L. AND KHABOU, M. (2023). Nonlinear Poisson autoregression and nonlinear Hawkes processes. *Stochastic Process. Appl.* **161**, 201–241. issn: 0304-4149,1879-209X. Available at https://doi.org/10.1016/j.spa.2023.03.015.

[24] KARABASH, D. AND ZHU, L. (2015). Limit theorems for marked Hawkes processes with application to a risk model. *Stoch. Models* **31**(3), 433–451. issn: 1532-6349,1532-4214. Available at https://doi.org/10.1080/15326349.2015.1024868.

[25] KHABOU, M., COHEN, E. A. K. AND VERAART, A. E. D. (2025). *The Markov approximation of the periodic multivariate Poisson autoregression*. Available at https://arxiv.org/abs/2504.02649.

[26] KHABOU, M., PRIVAULT, N. AND RÉVEILLAC, A. (2023). Normal approximation of compound Hawkes functionals. *J. Theoret. Probab.* **37**, 1–33.

[27] KHABOU, M. AND TALBI, M. (2025). *Markov approximation for controlled Hawkes jump-diffusions with general kernels.* Available at https://arxiv.org/abs/2507.11294.

[28] KHABOU, M. AND TORRISI, G. L. (2025). Gaussian approximation and moderate deviations of Poisson shot noises with application to compound generalized Hawkes processes. *Adv. Appl. Probab.* **57**(1), 305–345. issn: 0001-8678,1475-6064. Available at https://doi.org/10.1017/apr.2024.51.

[29] KIRCHNER, M. (2016). Hawkes and INAR($\infty$) processes. *Stochastic Process. Appl.* **126**(8), 2494–2525. issn: 0304-4149,1879-209X. Available at https://doi.org/10.1016/j.spa.2016.02.008.

[30] LAMBERT, R. C., TULEAU-MALOT, C., BESSAIH, T., RIVOIRARD, V., BOURET, Y., LERESCHE, N. AND REYNAUD-BOURET, P. (2018). Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *J. Neurosci. Methods* **297**, 9–21. issn: 0165-0270. doi: 10.1016/j.jneumeth.2017.12.026. Available at https://www.sciencedirect.com/science/article/pii/S0165027017304442.

[31] LEE, K. AND SEO, B. K. (2017). Modeling microstructure price dynamics with symmetric Hawkes and diffusion model using ultra-high-frequency stock data. *J. Econ. Dyn. Control* **79**, 154–183. issn: 0165-1889. doi: 10.1016/j.jedc.2017.04.004. Available at https://www.sciencedirect.com/science/article/pii/S0165188917300945.

[32] LOUZADA PINTO, J., CHAHED, T. AND ALTMAN, E. (2016). A framework for information dissemination in social networks using Hawkes processes. *Perform. Eval*. **103**. Performance Evaluation Methodologies and Tools: Selected Papers from ValueTools 2014, pp. 86–107. issn: 0166-5316. doi: 10.1016/j.peva.2016.06.004. Available at https://www.sciencedirect.com/science/article/pii/S016653161630058X.

[33] OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* **27**(1), 23–31. doi:https://doi.org/10.1109/TIT.1981.105630510.1109/TIT.1981.1056305.

[34] OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Stat. Assoc.* **83**(401), 9–27. Available at https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560.

[35] RYDBERG, T. H. AND SHEPHARD, N. (2000). A modelling framework for the prices and times of trades made on the New York stock exchange.

[36] SEOL, Y. (2015). Limit theorems for discrete Hawkes processes. *Stat. Probab. Lett.* **99**, 223–229. issn: 0167-7152. doi: 10.1016/j.spl.2015.01.023. Available at https://www.sciencedirect.com/science/article/pii/S0167715215000292.

[37] SULEM, D., RIVOIRARD, V. AND ROUSSEAU. J. (2024). Bayesian estimation of nonlinear Hawkes processes. *Bernoulli* **30**(2), 1257–1286. issn: 1350-7265,1573-9759. Available at https://doi.org/10.3150/23-bej1631.

https://doi.org/10.1017/apr.2025.10049 Published online by Cambridge University Press