



**HAL**  
open science

# Convergence of the Discrete-Time Compound Hawkes Process with Exponential or Erlang Kernel

Lorick Huang, Mahmoud Khabou

► **To cite this version:**

Lorick Huang, Mahmoud Khabou. Convergence of the Discrete-Time Compound Hawkes Process with Exponential or Erlang Kernel. 2021. hal-03220800v1

**HAL Id: hal-03220800**

**<https://hal.science/hal-03220800v1>**

Preprint submitted on 24 Jun 2021 (v1), last revised 4 Aug 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Convergence of the Discrete-Time Compound Hawkes Process with Exponential or Erlang Kernel

Lorick Huang<sup>\*</sup>      Mahmoud Khabou<sup>†</sup>

June 24, 2021

## Abstract

Due to its clustering and self-exciting properties, the Hawkes process has been used extensively in numerous fields ranging from seismology to finance. Since data is often acquired on regular time intervals, we propose a piece-wise constant model based on a Discrete-Time Hawkes Process (*DTHP*). We prove that this discrete-time model converges to the usual continuous-time Hawkes process as the time-step tends to zero.

## Résumé

Les propriétés d'auto-excitation des processus de Hawkes permettent une alternative de modélisation efficace au processus de Poisson à intensité déterministe dans plusieurs domaines d'application comme la finance ou la sismologie. Dans certaines applications, l'accès aux données se fait à des dates déterministes et non de façon continue dans le temps. Ainsi, seulement une approximation à temps discret du processus de Hawkes sur une grille déterministe est observable. Dans cet article nous étudions la convergence de cette approximation à temps discret lorsque le pas de la subdivision tend vers zéro.

**Subject Class:** 60J05, 60J25, 60G55.

**Keywords:** Hawkes Process, Discrete-Time, Markov Process.

## 1 Introduction and Main Result

The linear Hawkes process was first introduced in 1971 by Hawkes [6] as a point process whose intensity exhibits an interesting self-excitation property. Even though Hawkes process has initially contributed to seismology by describing the aftershocks in case of an earthquake, its self-exciting and clustering properties made it a popular model in financial and actuarial applications.

For instance Errais *et al.* used it to model the cumulative loss due to default in a portfolio of firms [4], while Bacry *et al.* used it for measuring the endogeneity of stock markets [1].

In the continuous time setting, the Hawkes process is defined as follows. Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with a filtration  $(\mathcal{F}_t)_{t \in [0, +\infty)}$  and a sequence of increasing stopping times  $0 < \theta_1 < \theta_2 < \dots$ . A point process is defined as the counting measure

$$H_t = H([0, t]) := \sum_{i=1}^{+\infty} \mathbb{1}_{\theta_i \leq t}.$$

We assume that an event at time  $\theta_n$  corresponds to a financial loss  $\zeta_n$ . The total loss a time  $t$  is the compound process

$$L_t := \sum_{i=1}^{+\infty} \zeta_i \mathbb{1}_{\theta_i \leq t} = \sum_{i=1}^{H_t} \zeta_i,$$

where  $\zeta_n$  are independent identically distributed (*i.i.d*) non-negative random variables with an integrable distribution  $\nu$  and independent from  $(H_t)_{t \in [0, +\infty)}$ .

---

<sup>\*</sup>INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France. Email: lorick.huang@insa-toulouse.fr.

<sup>†</sup>INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France. Email: mahmoud.khabou@insa-toulouse.fr

**Remark 1.1.** If  $\zeta_n$  are chosen to be deterministic and equal to 1 then  $L_t = H_t$  for all  $t \geq 0$ .

The intensity of a point process is a measure of how much it tends to jump at a certain time  $t$  and is defined as

$$\lambda_t = \lim_{\delta t \rightarrow 0} \frac{\mathbb{E}[H_{t+\delta t} - H_t | \mathcal{F}_t]}{\delta t}.$$

In the case of a Hawkes process, the realization of an event causes an increase in the probability of other events. This translates in the intensity as:

$$\begin{aligned} \lambda_t &= \mu(t) + \int_{[0,t)} \phi(t-s) dL_s, \\ &= \mu(t) + \sum_{\theta_i < t} \phi(t-\theta_i) \zeta_i, \end{aligned}$$

where  $\mu$  is a deterministic non-negative function playing the role of the baseline intensity and  $\phi$  is a non-negative decaying kernel. Indeed, more events ( $\theta_i$ ) mean more terms in the sum, thus a higher intensity which in return triggers more events. Larger losses have a bigger impact on the intensity as well.

The condition to avoid instability (i.e. infinite amount of jumps in a finite interval) is  $\|\phi\|_1 \mathbb{E}[\zeta] < 1$ . Curious readers can consult [7] for nearly unstable Hawkes processes (the kernel's norm approaches the limit of instability).

In this paper we study the case where the intensity kernel  $\phi$  is either an exponential ( $\phi(u) = \alpha e^{-\beta u}$ ) or an Erlang function ( $\phi(u) = \alpha u e^{-\beta u}$ ). The exponential kernel case has been studied extensively in the literature. This is mainly because in this case, the intensity  $(\lambda_t)_{t \in [0, +\infty)}$  is a Markov process. For example, Errais *et al.* [4] derived formulae for the Laplace transform for the Markov Hawkes process. Indeed, if the baseline intensity is chosen to be  $\mu(t) = \lambda_\infty + (x - \lambda_\infty) e^{-\beta t}$ , with the initial intensity  $x \geq 0$  and the parameter  $\lambda_\infty > 0$ , the intensity takes the form:

$$\lambda_t = \lambda_\infty + (x - \lambda_\infty) e^{-\beta t} + \int_{[0,t)} \alpha e^{-\beta(t-s)} dL_s, \quad (1)$$

where  $\alpha$  and  $\beta$  are two positive real numbers such that  $\beta > \alpha \mathbb{E}[\zeta]$ . In this case, the intensity satisfies the following stochastic differential equation (SDE):

$$(SDE_{exp}) \begin{cases} d\lambda_t = \beta(\lambda_\infty - \lambda_t) dt + \alpha dL_t, \\ \lambda_0 = x. \end{cases}$$

**Remark 1.2.** In many cases, the initial intensity  $\lambda_0$  is chosen to be equal to the parameter  $\lambda_\infty$  which yields a constant baseline intensity  $\lambda_t = \lambda_\infty + \int_{[0,t)} \alpha e^{-\beta(t-s)} dL_s$ .

If the kernel is an Erlang function, then the intensity takes the form

$$\lambda_t = \lambda_\infty + (x - \lambda_\infty) e^{-\beta t} + \int_{[0,t)} \alpha(t-s) e^{-\beta(t-s)} dL_s, \quad (2)$$

It is possible to "Markovize" the intensity by taking an auxiliary process  $\xi_t = \int_{[0,t)} \alpha e^{-\beta(t-s)} dL_s$  into account. Thus, in this case as well, the vector  $(\lambda_t, \xi_t)_{[0, +\infty)}$  follows the dynamics given by the SDE:

$$(SDE_{Erl}) \begin{cases} d\lambda_t = \beta(\lambda_\infty - \lambda_t) dt + \xi_t dt, \\ d\xi_t = -\beta \xi_t dt + \alpha dL_t, \\ \lambda_0 = x, \\ \xi_0 = 0. \end{cases}$$

So far the simulation of the Hawkes process has been based on Ogata's thinning [9], on an immigration clustering approach like in the work of Møller *et al.* [8] or in the particular Markov case on the sampling of jumping times such as the algorithm proposed by Dassios *et al.* [2].

These approaches simulate exactly the jump times of the process on a time continuum. However, in reality data is often recorded on discrete time intervals, *e.g.* every minute, every hour or every day.

This motivates the study of Discrete-Time Hawkes Processes (*DTHP*) first introduced by Seol [10], where limit

theorems have been established as time goes to infinity.

In this paper we study the behaviour as the size of the time step goes to zero instead.

The intensity (in the exponential kernel case) or the intensity-auxiliary process vector (in the Erlang kernel case) of this DTHP is considered as piece-wise constant process constructed from a Markov chain on the time grid (cf. figure 2).

**Remark 1.3.** *Knowing the intensity is sufficient for the reconstruction of  $(L_t)_{t \in [0, +\infty)}$ . This can be seen on figure 1 taken from [2]. This is why we focus on the intensity from now on. The loss process  $(L_t)_{t \in [0, +\infty)}$  is obtained by adding an independent copy of  $\zeta$  at every jumping time.*

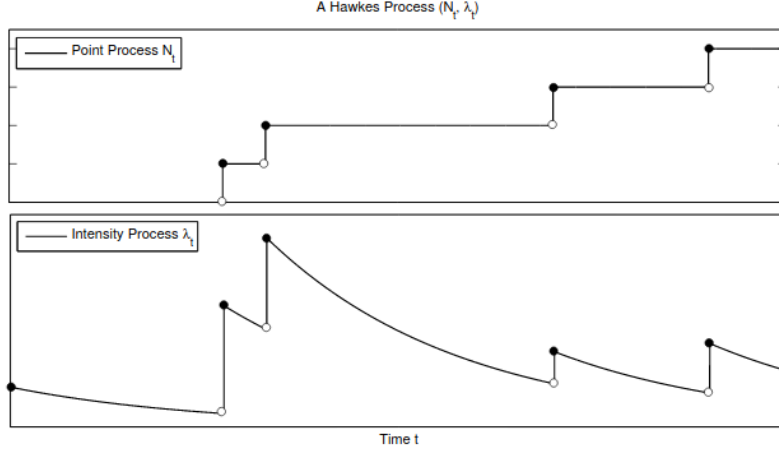


Figure 1: Hawkes process with exponential decaying intensity  $(N_t, \lambda_t)$ .

The main result is to show that the intensity (*resp.* intensity-auxiliary process vector) converges weakly to the continuous time Hawkes intensity (*resp.* to the intensity-auxiliary process vector) in the Skorokhod topology on  $[0, +\infty)$  as the grid becomes finer and finer.

Let  $\zeta$  be a positive random variable with finite expectation and let  $\nu$  be its distribution. Let  $\alpha, \beta, \lambda_\infty \in \mathbb{R}_+^*$  such that  $\alpha \mathbb{E}[\zeta] < \beta$  (exponential kernel) or  $\alpha \mathbb{E}[\zeta] < \beta^2$  (Erlang kernel) and  $x > 0$ . Let  $[0, T], 0 < T < +\infty$  be a time interval,  $N \in \mathbb{N}^*$  and  $(t_i^N := \frac{iT}{N})_{i \in [0 \dots N]}$  be a grid with a step  $h_N = \frac{T}{N}$ . In some cases we refer to  $h_N$  by  $h$  to avoid clogging up the notation.

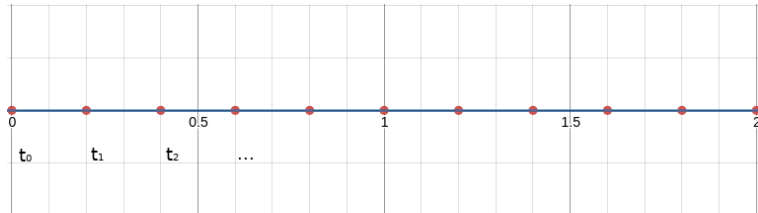


Figure 2: An example of a subdivision with  $T = 2$  and  $N = 10$

**Definition 1.4.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. Let  $N \in \mathbb{N}^*$ ,  $T > 0$  and a sequence of independent  $[0, 1]$  uniform random variables  $(U_k^N)_{k \in \mathbb{N}}$  as well as a sequence  $(\xi_k^N)_{k \in \mathbb{N}}$  of iid positive random variables with finite expectation defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ .*

1. *If  $\phi$  is an exponential kernel: The Hawkes Markov Chain  $(l^N)$  is a Markov chain defined according to the induction rule:*

$$(l^N) \begin{cases} l_{k+1}^N = \lambda_\infty(1 - e^{-\beta h}) + (l_k^N + \alpha \xi_{k+1}^N \mathbf{1}_{U_{k+1}^N < l_k^N \cdot h})e^{-\beta h}, \\ l_0^N = x. \end{cases}$$

2. If  $\phi$  is an Erlang kernel: The Hawkes Markov Chain  $(l^N, a^N)$  is a Markov chain defined according to the induction rule:

$$(l^N, a^N) \begin{cases} l_{k+1}^N = \lambda_\infty(1 - e^{-\beta h}) + l_k^N e^{-\beta h} + a_{k+1}^N h, \\ a_{k+1}^N = (a_k^N + \alpha \zeta_{k+1}^N \mathbf{1}_{U_{k+1}^N < l_k^N \cdot h}) e^{-\beta h}, \\ l_0^N = x, \\ a_0^N = 0. \end{cases}$$

**Definition 1.5.** Given  $N \in \mathbb{N}^*$  and  $T > 0$ , the  $N$ -th DTHP intensity  $(\tilde{\lambda}_t^N)_{t \in [0, +\infty)}$  and the Hawkes auxiliary process  $(\tilde{\xi}_t^N)_{t \in [0, +\infty)}$  (if the kernel is an Erlang function) are defined as the càdlàg process

$$\tilde{\lambda}_t^N = l_{\lfloor \frac{Nt}{T} \rfloor}^N,$$

$$\tilde{\xi}_t^N = a_{\lfloor \frac{Nt}{T} \rfloor}^N,$$

where  $l_{\lfloor \frac{Nt}{T} \rfloor}^N$  and  $a_{\lfloor \frac{Nt}{T} \rfloor}^N$  are defined in 1.4.

This process takes the values of the Markov chain on the grid points. Indeed

$$\tilde{\lambda}_{t_i}^N = l_{\lfloor \frac{iTN}{TN} \rfloor}^N = l_i^N \quad \text{and} \quad \tilde{\xi}_{t_i}^N = a_{\lfloor \frac{iTN}{TN} \rfloor}^N = a_i^N.$$

The following theorem, which will be proven in the following sections, states the main result:

**Theorem 1.6.** Let  $(H_t)_{t \in [0, +\infty)}$  be a Hawkes process,  $(L_t)_{t \in [0, +\infty)}$  its loss and  $(\lambda_t)_{t \in [0, +\infty)}$  its intensity.

1. If  $\phi$  is an exponential kernel: Let  $(\tilde{\lambda}_t^N)_{t \in [0, +\infty)}$  be an  $N$ -th DTHP intensity (defined in 1.5). Then we have the convergence

$$(\tilde{\lambda}_t^N)_{t \in [0, +\infty)} \Longrightarrow_{N \rightarrow +\infty} (\lambda_t)_{t \in [0, +\infty)}$$

weakly in the Skorokhod space  $D_{\mathbb{R}_+}[0, +\infty)$ , the set of all right continuous with left limits (càdlàg) non-negative functions on  $\mathbb{R}_+ = [0, +\infty)$ .

2. If  $\phi$  is an Erlang kernel: Let  $(\tilde{\lambda}_t^N, \tilde{\xi}_t^N)_{t \in [0, +\infty)}$  be an  $N$ -th DTHP intensity and auxiliary process (defined in 1.5). Then we have the convergence

$$(\tilde{\lambda}_t^N, \tilde{\xi}_t^N)_{t \in [0, +\infty)} \Longrightarrow_{N \rightarrow +\infty} (\lambda_t, \xi_t)_{t \in [0, +\infty)}$$

weakly in the Skorokhod space  $D_{\mathbb{R}_+^2}[0, +\infty)$ .

**Remark 1.7.** Normally the intensity is a càglàd process because it should be predictable (beyond the scope of this paper) but we work with the càdlàg version because the convergence results that we have in [5] as well as the Markov generator expression in [4] are for the càdlàg version.

Therefore we make the change  $\lambda_t \leftarrow \lambda_{t+} = \lim_{\delta \downarrow 0} \lambda_{t+\delta}$ .

## 2 Preliminary Results

### 2.1 General Notations and Lemmas

We denote by  $\mathbb{R}_+ = [0, +\infty)$  and we set  $E = \mathbb{R}_+$  or  $\mathbb{R}_+^2$ .  $\hat{C}(E)$  the space of real continuous functions on  $E$  vanishing at infinity.

$D_E[0, +\infty)$  refers to the set of all right continuous with left limits (càdlàg) functions  $x : [0, +\infty) \rightarrow E$ . On the other hand, càglàd is used to refer to left continuous functions with right limits.

**Lemma 2.1.**  $(\mathbb{R}_+, |\cdot|)$  is locally compact for the topology induced by the absolute value.

*Proof.*  $(\mathbb{R}, |\cdot|)$  is locally compact: every point has a compact neighbourhood.

The topology induced on  $\mathbb{R}_+$  is simply the set  $Top^+ = \{\mathbb{R}_+ \cap O, O \in Top\}$  with  $Top$  being the usual topology on  $\mathbb{R}$ . Thus  $[0, 1)$  is an open set containing 0 for  $(\mathbb{R}_+, |\cdot|)$ , which means that  $[0, 1]$  is a compact neighbourhood of 0. Any  $x > 0$  has a compact neighbourhood  $[x - \epsilon, x + \epsilon]$  for  $\epsilon$  small enough.  $\square$

**Lemma 2.2.**  $(\hat{C}(\mathbb{R}_+), \|\cdot\|)$  is a Banach space for  $\|f\| = \sup_{x \in \mathbb{R}_+} |f(x)|$ .

*Proof.* Let  $(f_n)_{n \in \mathbb{N}}$  be a Cauchy sequence in  $\hat{C}(\mathbb{R}_+)$ . Let  $\epsilon > 0$ , there exists  $M$  such that  $\forall n, p \geq M, \|f_n - f_p\| \leq \epsilon$ . Set  $x \in \mathbb{R}_+, |f_n(x) - f_p(x)| \leq \|f_n - f_p\| \leq \epsilon$  for  $n, p \geq M$ . Since  $\mathbb{R}$  is complete,  $(f_n(x))_{n \in \mathbb{N}}$  converges for every  $x \in \mathbb{R}_+$ . We call the point-wise limit  $f(x)$ . Set  $p \geq M$  and  $x \in \mathbb{R}_+$ ,

$$\begin{aligned} |f_p(x) - f(x)| &= |f_p(x) - \lim_{n \rightarrow +\infty} f_n(x)|, \\ &= \lim_{n \rightarrow +\infty} |f_p(x) - f_n(x)|, \\ &\leq \lim_{n \rightarrow +\infty} \|f_n - f_p\|, \end{aligned}$$

and since  $n \geq M$  (it goes to infinity) we have  $|f_p(x) - f(x)| \leq \epsilon$ . Because  $M$  is independent from  $x$  we have the uniform convergence  $\|f_p - f\| \leq \epsilon$ .

Let  $n$  be such that  $\|f_n - f\| \leq \epsilon$  and  $K$  such that  $|f_n(x)| \leq \epsilon$  if  $x > K$  (remember that the functions vanish at infinity). For all  $x \in \mathbb{R}_+$ , by the triangle inequality

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \|f - f_n\| + |f_n(x)|.$$

If  $x > K$  then  $|f(x)| \leq 2\epsilon$ , which means that  $f$  vanishes at infinity.

To prove the continuity of the limit function, let  $a \in \mathbb{R}_+$  and  $n$  such that  $\|f_n - f\| < \epsilon$ .  $f_n$  is continuous at  $a$  therefore there exists  $\eta > 0$  such that  $|x - a| < \eta \implies |f_n(x) - f_n(a)| < \epsilon$ . By the triangle inequality:

$$\begin{aligned} |f(x) - f(a)| &\leq |f_n(x) - f(x)| + |f_n(x) - f_n(a)| + |f(a) - f_n(a)|, \\ &\leq 2\|f_n - f\| + |f_n(x) - f_n(a)|, \end{aligned}$$

thus  $|f(x) - f(a)| \leq 3\epsilon$  if  $|x - a| < \eta$ .

In conclusion,  $f$  is continuous and vanishes at infinity thus  $\hat{C}(\mathbb{R}_+)$  is a Banach space.  $\square$

From now on the convergence in  $\hat{C}(E)$  refers to the convergence in the uniform norm  $\|f\| = \sup_{x \in E} |f(x)|$ .

**Lemma 2.3.** The set of twice continuously differentiable functions with compact support  $\hat{C}_c^2(\mathbb{R}_+)$  is dense in  $\hat{C}(\mathbb{R}_+)$  for the norm  $\|\cdot\|$ .

*Proof.* Let  $K \in \mathbb{R}_+$ . Take a non-negative infinitely differentiable function  $\phi_k$  with a compact support  $[K, K+1]$ .  $\phi_K$  is integrable and one can define  $\psi_K(x) = \frac{1}{\int_0^{+\infty} \phi_K(t) dt} \int_x^{+\infty} \phi_K(t) dt$ , an infinitely differentiable function.

$$\begin{cases} \psi_K(x) = \frac{1}{\int_0^{+\infty} \phi_K(t) dt} \int_x^{+\infty} \phi_K(t) dt = 1 & \text{if } x < K, \\ \psi_K(x) \in [0, 1] & \text{if } x \in [K, K+1], \\ \psi_K(x) = \frac{1}{\int_0^{+\infty} \phi_K(t) dt} \int_x^{+\infty} \phi_K(t) dt = 0 & \text{if } x > K+1. \end{cases}$$

Here is an illustration of  $\psi_K$ :  $\hat{C}_c^2(\mathbb{R}_+)$  is clearly a sub-algebra of  $\hat{C}(\mathbb{R}_+)$ . We prove its density using the locally

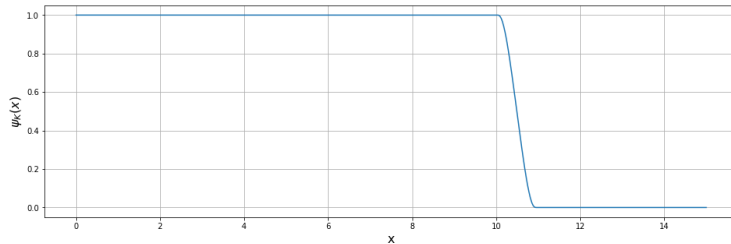


Figure 3:  $\psi_K$  is constructed using  $\phi_K(x) = \exp\left(\frac{-1}{1-(2x-2K+1)^2}\right)$  and  $K = 10$ .

compact version of the Stone-Weierstrass Theorem.

- Let  $c \neq c'$  be two elements of  $\mathbb{R}_+^2$ . Assume, without loss of generality that  $c < c'$ . Set  $f(x) = \psi_{cn}(xn)$  where  $n$  is such that  $\frac{1}{n} < c' - c$ . Clearly  $f \in \hat{C}_c^2(\mathbb{R}_+)$  and  $f(c) = 1$  whereas  $f(c') = 0$ .

- For any  $c \in \mathbb{R}_+$ , the last function guarantees that  $f(c) \neq 0$  thus  $\hat{C}_c^2(\mathbb{R}_+)$  vanishes nowhere.

We conclude that  $\hat{C}_c^2(\mathbb{R}_+)$  is dense in  $\hat{C}(\mathbb{R}_+)$  for the norm  $\|\cdot\|$ .  $\square$

**Lemma 2.4.** *The set of twice differentiable functions with compact support  $\hat{C}_c^2(\mathbb{R}_+^2)$  is dense in  $\hat{C}(\mathbb{R}_+^2)$  for the norm  $\|\cdot\|$ .*

*Proof.* The proof of this lemma is an extension of the previous one. Set

$$\mathcal{B} = \{(x, y) \rightarrow \sum_{k=1}^n f_k(x)g_k(y), n \in \mathbb{N}^*, (f_k, g_k) \in \hat{C}_2^c(\mathbb{R}_+)^2\}$$

a sub-algebra of  $\hat{C}(\mathbb{R}_+^2)$  (that is stable by sum, product as well as scalar multiplication). In order to apply the Stone-Weierstrass Theorem one must make sure that  $\mathcal{B}$  separates points and vanishes nowhere.

- Let  $X \neq X'$  be two vectors in  $\mathbb{R}_+^2$ . Assume, without loss of generality that their first components  $c$  and  $c'$  are such that  $c < c'$ . Set  $f(x, y) = \psi_{cn}(xn)$  where  $n$  is such that  $\frac{1}{n} < c' - c$ . Clearly  $f \in \mathcal{B}$  and  $f(X) = 1$  where as  $f(X') = 0$ .
- For any  $X \in \mathbb{R}_+^2$ , the last function guarantees that  $f(X) \neq 0$  thus  $\mathcal{B}$  vanishes nowhere.

We conclude that  $\mathcal{B}$  and a fortiori  $\hat{C}_c^2(\mathbb{R}_+^2)$  is dense in  $\hat{C}(\mathbb{R}_+^2)$  for the norm  $\|\cdot\|$ .  $\square$

## 2.2 General Results on Continuous Time Markov Processes

**Definition 2.5.** *A family of bounded linear operators  $(\mathcal{T}(t))_{t \geq 0}$  on  $\hat{C}(E)$  is called a semigroup if for each  $s, t \geq 0$ :*

- $\mathcal{T}(t + s) = \mathcal{T}(t) \cdot \mathcal{T}(s)$ ,
- $\mathcal{T}(0) = Id$ .

*A semigroup is called:*

- *A contraction semigroup if  $\forall f \in \hat{C}(E)$  and  $\forall t \geq 0, \|\mathcal{T}(t)f\| \leq \|f\|$ .*
- *Strongly continuous if  $\forall f \in \hat{C}(E), \lim_{t \rightarrow 0} \|\mathcal{T}(t)f - f\| = 0$ .*
- *Conservative if  $\mathcal{T}(t)\mathbb{1}_E = \mathbb{1}_E$ . Where  $\mathbb{1}_E$  is the function that takes the value 1 everywhere.*
- *Positive if  $\forall t \geq 0$  and  $\forall f \in \hat{C}(E)$  such that  $f \geq 0, \mathcal{T}(t)f \geq 0$ .*

*If a semigroup has all the previous properties then it is called a Feller semigroup.*

**Definition 2.6.** *The infinitesimal generator  $\mathcal{A}$  of a semigroup  $(\mathcal{T}(t))_{t \geq 0}$  on  $\hat{C}(E)$  is the linear operator defined by:*

$$\mathcal{A}f = \lim_{t \rightarrow 0} \frac{\mathcal{T}(t)f - f}{t},$$

*whenever the limit exists in  $\hat{C}(E)$ .*

*The domain  $\mathcal{D}(\mathcal{A})$  is the subset of the functions  $f \in \hat{C}(E)$  for which the limit exists.*

**Definition 2.7.** *Let  $\mathcal{A}$  be the generator of a Feller semigroup  $\mathcal{T}(t)$  on  $\hat{C}(E)$ . Let  $D$  be a dense subspace of  $\hat{C}(E)$  with  $D \subset \mathcal{D}(\mathcal{A})$ . If  $\mathcal{T}(t) : D \rightarrow D$  for all  $t \geq 0$ , then we say that  $D$  is a core for  $\mathcal{A}$ .*

**Remark 2.8.** *The actual definition of a core is different (cf [5], page 17), what we have just introduced above is merely a sufficient condition for a subset to be a core. It is sufficient for our application nevertheless.*

## 2.3 Known Results on the Continuous Time Intensity

**Theorem 2.9.** 1. Let  $(H_t)_{t \in [0, +\infty)}$  be a Hawkes process whose intensity  $(\lambda_t)_{t \in [0, +\infty)}$  follows the Markov dynamics of equation (1). Then  $(\lambda_t)_{t \in [0, +\infty)}$  is a Markov process whose semigroup

$$\mathcal{T}_e(t)f(x) = \mathbb{E}[f(\lambda_t) | \lambda_0 = x]$$

is a well defined Feller semigroup that satisfies  $\mathcal{T}_e(t) : \hat{C}(\mathbb{R}_+) \rightarrow \hat{C}(\mathbb{R}_+)$ .

The domain of the generator is  $\mathcal{D}(\mathcal{A}_e) = C^1(\mathbb{R}_+)$ . Moreover, the generator -defined on the set of continuously differentiable functions  $C^1(\mathbb{R}_+)$ - is

$$\mathcal{A}_e f(\lambda) = \beta(\lambda_\infty - \lambda)f'(\lambda) + \lambda \int (f(\lambda + \alpha z) - f(\lambda))d\nu(z).$$

2. Let  $(H_t)_{t \in [0, +\infty)}$  be a Hawkes process whose intensity  $(\lambda_t)_{t \in [0, +\infty)}$  follows the Erlang dynamics of equation (2). Then  $(\lambda_t, \xi_t)_{t \in [0, +\infty)}$  (where  $\xi$  is the auxiliary process) is a Markov process whose semigroup

$$\mathcal{T}_E(t)f(x, y) = \mathbb{E}[f(\lambda_t, \xi_t) | \lambda_0 = x, \xi_0 = y]$$

is a well defined Feller semigroup that satisfies  $\mathcal{T}_E(t) : \hat{C}(\mathbb{R}_+^2) \rightarrow \hat{C}(\mathbb{R}_+^2)$ .

The domain of the generator is  $\mathcal{D}(\mathcal{A}_E) = C^1(\mathbb{R}_+^2)$ . Moreover, the generator -defined on the set of continuously differentiable functions  $C^1(\mathbb{R}_+^2)$ - is

$$\mathcal{A}_E f(\lambda, \xi) = (\xi + \beta(\lambda_\infty - \lambda))\partial_\lambda f(\lambda, \xi) - \beta\xi\partial_\xi f(\lambda, \xi) + \lambda \int (f(\lambda, \xi + \alpha z) - f(\lambda, \xi))d\nu(z).$$

*Proof.* 1. If the kernel is an exponential function For the proof that  $(\lambda_t)_{t \in [0, +\infty)}$  is a Markov process and the expression of its generator we refer to [4], section 2.3.

However, we prove that  $\mathcal{T}(t) : \hat{C}(\mathbb{R}_+) \rightarrow \hat{C}(\mathbb{R}_+)$  is a Feller semigroup.

Let  $t \geq 0$  and  $f \in \hat{C}(\mathbb{R}_+)$ . Start by showing that  $\mathcal{T}(t)f$  is continuous. To do so, let  $x \in \mathbb{R}_+$  and a sequence  $\epsilon_n \rightarrow 0$  ( $\epsilon_n$  must be positive if  $x = 0$ ).

$$\begin{aligned} \mathcal{T}(t)f(x + \epsilon_n) &= \mathbb{E}[f(\lambda_t) | \lambda_0 = x + \epsilon_n], \\ &= \mathbb{E}\left[f(\lambda_\infty + (x + \epsilon_n - \lambda_\infty)e^{-\beta t} + \int_{[0, t]} e^{-\beta(t-s)} dL_s)\right]. \end{aligned}$$

Since  $f$  is continuous and  $x + \epsilon_n \rightarrow x$ , and given that  $f(\lambda_\infty + (x + \epsilon_n - \lambda_\infty)e^{-\beta t} + \int_{[0, t]} e^{-\beta(t-s)} dL_s) \leq \|f\| \in L^1$ , one can apply the Dominated Convergence Theorem to conclude that:

$$\mathcal{T}f(x + \epsilon_n) \rightarrow_{n \rightarrow +\infty} \mathcal{T}f(x).$$

To prove that  $\mathcal{T}(t)f$  vanishes at infinity we start by setting  $\epsilon > 0$  and we take  $K$  such that  $x > K$  implies  $|f(x)| \leq \epsilon$ . If  $x > (K - \lambda_\infty)e^{\beta t} + \lambda_\infty$  one has

$$|f(\lambda_\infty + (x + \epsilon_n - \lambda_\infty)e^{-\beta t} + \int_{[0, t]} e^{-\beta(t-s)} dL_s)| \leq \epsilon.$$

Thus  $|\mathcal{T}(t)f(x)| \leq \epsilon$  and  $\mathcal{T}(t) : \hat{C}(\mathbb{R}_+) \rightarrow \hat{C}(\mathbb{R}_+)$ .

Now we prove that the semigroup is Feller.

- Let  $f \in \hat{C}(\mathbb{R}_+)$ , for all  $x \in \mathbb{R}_+$  we have  $|\mathcal{T}(t)f(x)| \leq \mathbb{E}[|f(\lambda_t)| | \lambda_0 = x] \leq \mathbb{E}[|f| | \lambda_0 = x] \leq \|f\|$ . Thus  $T$  is a contraction.
- $\mathcal{T}(t)\mathbf{1}_{\mathbb{R}_+}(x) = \mathbb{E}[\mathbf{1}_{\mathbb{R}_+}(\lambda_\infty + (x - \lambda_\infty)e^{-\beta t} + \int_{[0, t]} e^{-\beta(t-s)} dL_s) | \lambda_0 = x] = \mathbf{1}_{\mathbb{R}_+}(x)$ . Thus  $\mathcal{T}$  is conservative.
- If  $f \geq 0$  then clearly  $\mathbb{E}[f(\lambda_t) | \lambda_0 = x] \geq 0$ . Thus  $\mathcal{T}$  is positive.
- To prove strong continuity, we start by taking  $\epsilon > 0$ ,  $f \in \hat{C}(\mathbb{R}_+)$  and  $f_n \in \hat{C}_c^2(\mathbb{R}_+)$  such that  $\|f_n - f\| \leq \epsilon$  (cf lemma 2.3). Since we will make  $t \rightarrow 0$  it is possible to assume  $t < 1$ . Using Jensen's inequality:

$$\|\mathcal{T}(t)f_n - f_n\| = \sup_{x \in \mathbb{R}_+} \left| \mathbb{E}[f_n(\lambda_\infty + (x - \lambda_\infty)e^{-\beta t} + \int_{[0, t]} e^{-\beta(t-s)} dL_s) - f_n(x)] \right|,$$



$$\begin{aligned}
&\leq \sup_{x \in \mathbb{R}_+} \mathbb{E} \left[ \left| f_n(\lambda_\infty + (x - \lambda_\infty)e^{-\beta t} + \int_{[0,t)} e^{-\beta(t-s)} dL_s) - f_n(x) \right| \right], \\
&\leq \mathbb{E} \left[ \sup_{x \in \mathbb{R}_+} \left| f_n(\lambda_\infty + (x - \lambda_\infty)e^{-\beta t} + \int_{[0,t)} e^{-\beta(t-s)} dL_s) - f_n(x) \right| \right],
\end{aligned}$$

Now set  $\alpha_x = \lambda_\infty + (x - \lambda_\infty)e^{-\beta t} + \int_{[0,t)} e^{-\beta(t-s)} dL_s$ , since we assumed that  $t < 1$  we have  $\lambda_\infty + (x - \lambda_\infty)e^{-\beta} \leq \inf(x, \alpha_x)$ .

Using the mean value theorem, there is  $\lambda_\infty + (x - \lambda_\infty)e^{-\beta t} \leq \theta_x$  (random) such that

$$\begin{aligned}
f_n(\alpha_x) - f_n(x) &= (\alpha_x - x) \cdot f'_n(\theta_x), \\
&= ((x - \lambda_\infty)(e^{-\beta t} - 1) + \int_{[0,t)} e^{-\beta(t-s)} dL_s) \cdot f'_n(\theta_x).
\end{aligned}$$

the function  $f_n$  is  $C^1$  with compact support. Therefore the following inequalities are obtained:

- (a)  $|f'_n(x)| \leq M$  where  $M \in [0, +\infty)$  (deterministic), for any  $x \in \mathbb{R}_+$ .
- (b)  $|(x - \lambda_\infty)f'_n(\theta_x)| \leq M'$  where  $M' \in [0, +\infty)$  (deterministic), for any  $x \in \mathbb{R}_+$ . This is due to the fact that  $\lambda_\infty + (x - \lambda_\infty)e^{-\frac{1}{\tau}} \leq \theta_x$  which imposes that if  $x$  is too large, then  $f'(\theta_x) = 0$ .

Moreover, since  $e^{-\beta(t-s)} \leq 1$  for  $s \in [0, t)$  one has

$$\begin{aligned}
\int_{[0,t)} e^{-\beta(t-s)} dL_s &\leq \int_{[0,t)} 1 dL_s, \\
&= L_{t-}, \\
&\leq L_t.
\end{aligned}$$

Combining all these elements yields:

$$\begin{aligned}
\|\mathcal{T}(t)f_n - f_n\| &\leq \mathbb{E} \left[ \sup_{x \in \mathbb{R}_+} \left| ((x - \lambda_\infty)(e^{-\beta t} - 1) + \int_{[0,t)} e^{-\beta(t-s)} dL_s) \cdot f'_n(\theta_x) \right| \right], \\
&\leq \mathbb{E} [M' \cdot (1 - e^{-\beta t}) + M \cdot H_t].
\end{aligned}$$

From [2] we have an explicit expression for  $\mathbb{E}[H_t]$  and we know that  $\lim_{t \rightarrow 0} \mathbb{E}[H_t] = 0$ , thus the result for  $f_n$ . Now we extend it by density for the norm  $\|\cdot\|$ :

$$\begin{aligned}
\|\mathcal{T}(t)f - f\| &= \|\mathcal{T}(t)(f - f_n) + \mathcal{T}(t)f_n - (f - f_n) - f_n\|, \\
&\leq \|\mathcal{T}(t)(f - f_n) - (f - f_n)\| + \|\mathcal{T}(t)f_n - f_n\|, \\
&\leq \|\mathcal{T}(t)(f - f_n)\| + \|(f - f_n)\| + \|\mathcal{T}(t)f_n - f_n\|.
\end{aligned}$$

Finally, since  $\mathcal{T}$  is a contraction,  $\|\mathcal{T}(t)(f - f_n)\| \leq \|(f - f_n)\|$  and we conclude that

$$\|\mathcal{T}(t)f - f\| \rightarrow 0.$$

2. If the kernel is an Erlang function:

The generator and its domain can be found in [3]. All the other computations are identical to those of the exponential kernel case. □

### 3 Proof of the Main Result

The main result (Theorem 1.6) is an immediate corollary of the following theorem (Theorem 2.7 from [5] page 168):

**Theorem 3.1.** *Let  $E$  be locally compact and separable. For  $N = 1, 2, \dots$  let  $\mu_N(x, \Gamma)$  be a transition function on  $E \times \mathcal{B}(E)$  such that  $\mathcal{T}_N$  defined by*

$$\mathcal{T}_N f(x) = \int f(y) \mu_N(x, dy),$$

satisfies  $\mathcal{T}_N : \hat{C}(E) \rightarrow \hat{C}(E)$ . Suppose that  $(\mathcal{T}(t))_{t \geq 0}$  is a Feller semigroup on  $\hat{C}(E)$ . Let  $h_N > 0$  satisfy  $\lim_{N \rightarrow +\infty} h_N = 0$  and suppose that for every  $f \in \hat{C}(E)$ ,

$$\lim_{N \rightarrow +\infty} \mathcal{T}_N^{\lfloor t/h_N \rfloor} f = \mathcal{T}(t)f, \quad t \geq 0.$$

For each  $N \geq 1$ , let  $(Y_k^N)_{k \geq 0}$  be a Markov chain in  $E$  with transition function  $\mu_N(x, \Gamma)$  and suppose that  $Y_0^N$  has a limiting distribution  $\nu$ . Define  $X^N$  by  $X_t^N = Y_{\lfloor t/h_N \rfloor}^N$ . Then there is a Markov process  $X$  corresponding to  $(\mathcal{T}(t))_{t \geq 0}$  with initial distribution  $\nu$  and sample paths in  $D_E[0, +\infty[$  and  $X^N \Rightarrow X$ .

In this section we prove that the process  $(\tilde{\lambda}_t^N)_{t \in [0, +\infty)}$  satisfies the conditions of Theorem 3.1. In the context of this paper,  $(Y_k^N)_{k \geq 0} = (l_k^N)_{k \geq 0}$  (or  $(l_k^N, a_k^N)_{k \geq 0}$ ),  $X^N = \tilde{\lambda}^N$  (or  $(\tilde{\lambda}^N, \tilde{\xi}^N)$ ).

### 3.1 Initial Condition

First of all, we fix  $l_0^N = x$  (and  $a_0^N = 0$  if the kernel is an Erlang function) for some  $x \in \mathbb{R}_+$  independently from  $N$ , thus  $l_0^N$  does have a limiting distribution  $\delta_x$ .

### 3.2 Convergence of the Operators

Now the trickier part to prove is the convergence of the discrete one-step operator to the Feller semigroup associated with the Hawkes intensity. Unfortunately, we do not know that much about the semigroup nor about the composition of one-step operator with itself. That is why using generators is indispensable. We start this part by mentioning the lemmas (from [5]) that will be used:

**Lemma 3.2.** *Let  $L$  be a Banach functional space on  $E$ .*

*For  $N = 1, 2, \dots$  let  $\mathcal{T}_N$  be a linear contraction on  $L$ , let  $h_N$  be a positive number and put  $\mathcal{A}_N = h_N^{-1}(\mathcal{T}_N - Id)$ . Assume that  $\lim_{N \rightarrow +\infty} h_N = 0$ . Let  $((\mathcal{T}(t))_{t \geq 0})$  be a strongly continuous contraction semigroup on  $L$  with generator  $\mathcal{A}$  and let  $D$  be a core for  $\mathcal{A}$ . Then the following are equivalent:*

1. *For each  $f \in L$ ,  $\mathcal{T}_N^{\lfloor t/h_N \rfloor} f \rightarrow \mathcal{T}(t)f$  for all  $t \geq 0$ .*
2. *For each  $f \in D$  there exists  $f_N \in L$  such that  $f_N \rightarrow f$  and  $\mathcal{A}_N f_N \rightarrow \mathcal{A}f$ .*

*Proof.* Cf [5] page 31. □

**Lemma 3.3.**  *$\hat{C}_c^2(E)$  is a core for  $\mathcal{A}_j$ ,  $j \in \{e, E\}$ .*

*Proof.* Let us start with the exponential kernel case. According to Definition 2.7 one must show that  $\mathcal{T}(t) : \hat{C}_c^2(\mathbb{R}_+) \rightarrow \hat{C}_c^2(\mathbb{R}_+)$ , the density has been proven in Lemma 2.3.

Let  $t \geq 0$  and  $f \in C_c^2(\mathbb{R}_+)$ . There exists  $B > 0$  such that  $x \geq B \Rightarrow f(x) = 0$ .

If  $x \geq (B - \lambda_\infty)e^{\beta t} + \lambda_\infty$ , then  $\lambda_t \geq \lambda_\infty + (x - \lambda_\infty)e^{-\beta t} \geq B$ .

It follows that  $\mathcal{T}(t)f(x) = \mathbb{E}(f(\lambda_t) | \lambda_0 = x) = 0$ .

The interchangeability of the derivative and the expectation is possible because  $\|f'\| < +\infty$  and  $\|f''\| < +\infty$ , thus the (twice) differentiability of  $\mathcal{T}(t)f$ .

If the kernel is an Erlang function, the computations are similar. □

**Proposition 3.4.**  *$E = \mathbb{R}_+$  or  $\mathbb{R}_+^2$ .*

1. *We assume that the kernel is exponential. The one-step transition operator  $\mathcal{T}_e^N$  associated to  $l^N$  and evaluated at a function  $f \in \hat{C}(E)$  is:*

$$\begin{aligned} \mathcal{T}_e^N f(y) &:= \mathbb{E}[f(l_{k+1}^N) | l_k^N = y], \\ &= f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h})(1 - yh)\mathbb{1}_{yh < 1} \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h})d\nu(z)(yh\mathbb{1}_{yh < 1} + \mathbb{1}_{yh \geq 1}). \end{aligned}$$

2. If the kernel is an Erlang function, then the one-step transition operator  $\mathcal{T}_E^N$  associated to  $(l^N, a^N)$  and evaluated at a function  $f \in \hat{C}(E)$  is:

$$\begin{aligned}\mathcal{T}_E^N f(y, v) &:= \mathbb{E}[f(l_{k+1}^N, a_{k+1}^N) | l_k^N = y, a_k^N = v], \\ &= f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + vhe^{-\beta h}, ve^{-\beta h})(1 - yh)\mathbb{1}_{yh < 1} \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h})d\nu(z)(yh\mathbb{1}_{yh < 1} + \mathbb{1}_{yh \geq 1}).\end{aligned}$$

*Proof.* Let  $N \in \mathbb{N}^*$  and  $T > 0$ .

1. Set  $F(l, u, \zeta) = \lambda_\infty(1 - e^{-\beta h}) + (l + \alpha\zeta\mathbb{1}_{u < l \cdot h})e^{-\beta h}$ .  $F$  is clearly measurable and  $l_{k+1}^N = F(l_k^N, U_{k+1}^N, \zeta_{k+1}^N)$ , where  $l_k^N$  is  $\sigma(U_i^N, \xi_i^N, i \in \llbracket 0, k \rrbracket)$  measurable, for any  $k \in \mathbb{N}$ . Thus,  $(l_k^N)_{k \in \mathbb{N}}$  is a Markov chain. When it comes to the one-step transition operator, computing the expected value yields:

$$\begin{aligned}\mathcal{T}_e^N f(y) &:= \mathbb{E}[f(l_{k+1}^N) | l_k^N = y], \\ &= \mathbb{E}[f(\lambda_\infty(1 - e^{-\beta h}) + (l_k^N + \alpha\xi_{k+1}^N\mathbb{1}_{U_{k+1}^N < l_k^N \cdot h})e^{-\beta h}) | l_k^N = y], \\ &= \mathbb{E}[f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha\xi_{k+1}^N\mathbb{1}_{U_{k+1}^N < y \cdot h})e^{-\beta h})],\end{aligned}$$

and since  $U_{k+1}^N$  and  $\xi_{k+1}^N$  are independent from  $l_k^N \in \sigma(U_i^N, \xi_i^N, i \in \llbracket 0, k \rrbracket)$  and since  $\mathbb{1}_{U_{k+1}^N < y \cdot h}$  is a Bernoulli variable with parameter  $y \cdot h$  independent from  $\xi_{k+1}^N$ :

$$\begin{aligned}\mathcal{T}_e^N f(y) &= f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h})(1 - yh)\mathbb{1}_{yh < 1} \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h})d\nu(z)(yh\mathbb{1}_{yh < 1} + \mathbb{1}_{yh \geq 1}).\end{aligned}$$

2. Set

$$F(l, a, u, \zeta) = \begin{pmatrix} \lambda_\infty(1 - e^{-\beta h}) + le^{-\beta h} + (a + \alpha\zeta\mathbb{1}_{u < l \cdot h})he^{-\beta h} \\ (a + \alpha\zeta\mathbb{1}_{u < l \cdot h})e^{-\beta h} \end{pmatrix}$$

a measurable function. Clearly  $(l_{k+1}^N, a_{k+1}^N) = F(l_k^N, a_k^N, U_{k+1}^N, \zeta_{k+1}^N)$  where  $(l_k^N, a_k^N)$  is  $\sigma(U_i^N, \xi_i^N, i \in \llbracket 0, k \rrbracket)$  measurable, for any  $k \in \mathbb{N}$ . Thus,  $(l_k^N, a_k^N)_{k \in \mathbb{N}}$  is a Markov chain.

The Erlang one-step generator can be obtained just like exponential one. □

**Theorem 3.5.**  $E = \mathbb{R}_+$  or  $\mathbb{R}_+^2$ .

Let  $f \in \hat{C}_c^2(E)$ ,  $\mathcal{A}$  be the generator of a Hawkes intensity and  $\mathcal{T}_j^N$  where  $j \in \{e, E\}$  the operator described in Proposition 3.4. Then

$$\left\| \frac{\mathcal{T}_j^N f - f}{h_N} - \mathcal{A}_j f \right\| \longrightarrow 0.$$

*Proof.* First we remind that  $h = h_N$ .

**If the kernel is an exponential function**

$E = \mathbb{R}_+$ .

Let  $f \in \hat{C}_c^2(E)$  be a fixed function. We start by giving an alternative expression for  $\frac{\mathcal{T}_e^N f(y) - f(y)}{h_N}$  for a fixed  $y \in E$ :

$$\begin{aligned}\frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} &= [f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h})(1 - yh)\mathbb{1}_{yh < 1} \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h})d\nu(z)(yh\mathbb{1}_{yh < 1} + \mathbb{1}_{yh \geq 1}) - f(y)]h_N^{-1}, \\ &= [f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h})(1 - yh) + \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h})d\nu(z)yh - f(y)]h_N^{-1}\mathbb{1}_{yh < 1} \\ &\quad + \left[ \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h}) - f(y)d\nu(z) \right]h_N^{-1}\mathbb{1}_{yh \geq 1},\end{aligned}$$

since  $\int d\nu(z) = 1$ .

Now using the fact that  $f$  is twice differentiable, we use Taylor expansion with a Lagrange remainder:

$$\begin{aligned} f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h}) &= f(y) + f'(y)[(\lambda_\infty - y)(1 - e^{-\beta h})] \\ &\quad + \frac{1}{2}f''(\theta_y)[(\lambda_\infty - y)(1 - e^{-\beta h})]^2, \end{aligned}$$

where  $\theta_y \in [\inf(y, \lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h}), \sup(y, \lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h})] \subset [ye^{-\beta T}, \lambda_\infty(1 - e^{-\beta T}) + y]$ . The last inclusion will be used later.

We apply another Taylor expansion to obtain:

$$f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h}) = f(y + \alpha z) + f'(\gamma_y)[(\lambda_\infty - y - \alpha z)(1 - e^{-\beta h})]$$

where  $\gamma_y \in [(y + \alpha z)e^{-\beta T}, \lambda_\infty(1 - e^{-\beta T}) + y + \alpha z]$ .

Thus we have:

$$\begin{aligned} \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} \mathbb{1}_{yh < 1} &= (f(y) + f'(y)[(\lambda_\infty - y)(1 - e^{-\beta h})] + \frac{1}{2}f''(\theta_y)[(\lambda_\infty - y)(1 - e^{-\beta h})]^2 \\ &\quad - yh f(y) - yh f'(y)[(\lambda_\infty - y)(1 - e^{-\beta h})] - yh \frac{1}{2}f''(\theta_y)[(\lambda_\infty - y)(1 - e^{-\beta h})]^2 \\ &\quad + yh \int f(y + \alpha z) + f'(\gamma_y)[(\lambda_\infty - y - \alpha z)(1 - e^{-\beta h})] d\nu(z) - f(y) h^{-1} \mathbb{1}_{yh < 1}, \end{aligned}$$

a Taylor expansion for small  $h$  yields  $1 - e^{-\beta h} = \beta h + O(h^2) = O(h)$ , thus:

$$\begin{aligned} \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} \mathbb{1}_{yh < 1} &= (f'(y)(\lambda_\infty - y)(\beta + O(h)) + \frac{1}{2}f''(\theta_y)(\lambda_\infty - y)^2 \frac{1 - e^{-\beta h}}{h}) O(h), \\ &\quad - y f(y) - y f'(y)(\lambda_\infty - y) O(h) - y \frac{1}{2}f''(\theta_y)(\lambda_\infty - y)(1 - e^{-\beta h}) O(h) \\ &\quad + y \int f(y + \alpha z) d\nu(z) + y \int f'(\gamma_y)(\lambda_\infty - y - \alpha z) d\nu(z) O(h) \mathbb{1}_{yh < 1}, \\ &= (\beta(\lambda_\infty - y)f'(y) + y \int (f(y + \alpha z) - f(y)) d\nu(z)) \mathbb{1}_{yh < 1} + R(y) O(h) \mathbb{1}_{yh < 1}, \\ &= \mathcal{A}_e f(y) \mathbb{1}_{yh < 1} + R(y) O(h) \mathbb{1}_{yh < 1}, \end{aligned}$$

The remainder  $R$  has the expression:

$$\begin{aligned} R(y) &= f'(y)(\lambda_\infty - y) + \frac{1}{2}f''(\theta_y)(\lambda_\infty - y)^2 \frac{1 - e^{-\beta h}}{h} \\ &\quad + y f'(y)(\lambda_\infty - y) - y \frac{1}{2}f''(\theta_y)(\lambda_\infty - y)^2 (1 - e^{-\beta h}) + y \int f'(\gamma_y)(\lambda_\infty - y - \alpha z) d\nu(z). \end{aligned}$$

Note that the remainder is bounded. Remember that  $f$  is continuous with compact support (so are its derivatives), so the function  $y \rightarrow \sup_{x \in [ye^{-\beta T}, \lambda_\infty(1 - e^{-\beta T}) + y]} |f''(x)|$  is also continuous with compact support, which in turn means that  $y \rightarrow y^k \sup_{x \in [ye^{-\beta T}, \lambda_\infty(1 - e^{-\beta T}) + y]} |f''(x)|$  for  $k = 2, 3$  is also compact support, thus bounded by a positive constant independently from  $y$ . Applying the same logic to the terms involving  $f'$  yields:

$$|R(y)| \leq C,$$

where  $C$  is a positive constant independent from  $y$ . After all these computations, we prove the uniform convergence for.

Let  $\epsilon > 0$  and  $B$  a constant such that  $y > B$  implies  $f(y) = f'(y) = 0$ .

For all  $y \in \mathbb{R}_+$  we have:

$$\left| \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right| = \left| \left( \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right) \mathbb{1}_{yh < 1} + \left( \frac{\mathcal{T}_e^N f(y) - f(y)}{y} - \mathcal{A}_e f(y) \right) \mathbb{1}_{yh \geq 1} \right|$$

$$\leq \left| \left( \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right) \mathbb{1}_{yh < 1} \right| + \left| \left( \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right) \mathbb{1}_{yh \geq 1} \right|.$$

Now we plug in the inequality obtained previously for the term in front of  $\mathbb{1}_{yh < 1}$  and expand the one in front of  $\mathbb{1}_{yh \geq 1}$ :

$$\left| \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right| \leq Ch \mathbb{1}_{yh < 1} + \left| \left[ \int f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h}) - f(y) d\nu(z) \right] h_N^{-1} - \mathcal{A}_e f(y) \right| \mathbb{1}_{yh \geq 1}$$

where  $O(h)$  has been absorbed by the constant  $C$ .

Let  $N_0$  be the integer such that if  $N \geq N_0$ , then  $\frac{N}{T} e^{-\beta \frac{T}{N}} \geq B$ . Such integer exists because  $\frac{N}{T} e^{-\beta \frac{T}{N}} \geq \frac{N}{T} e^{-\beta T} \rightarrow +\infty$  as  $N \rightarrow +\infty$ .

Set  $N_1 = \lfloor \frac{CT}{\epsilon} \rfloor + TB + N_0$ .

For every  $N \geq N_1$  and every  $y \in \mathbb{R}_+$  only one of these two scenarios is possible:

- $\mathbb{1}_{yh < 1} = 1$  and  $\mathbb{1}_{yh \geq 1} = 0$ , so  $\left| \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right| \leq \epsilon$  because  $N \geq N_1 \geq \lfloor \frac{CT}{\epsilon} \rfloor$ .
- $\mathbb{1}_{yh < 1} = 0$  and  $\mathbb{1}_{yh \geq 1} = 1$  which means  $y \geq \frac{N}{T}$ , thus  $y \geq B$  and  $(y + \alpha z)e^{-\beta h} \geq \frac{N}{T} e^{-\beta \frac{T}{N}} \geq B$ . Therefore  $f(\lambda_\infty(1 - e^{-\beta h}) + (y + \alpha z)e^{-\beta h}) = f(y) = f'(y) = 0 \forall z \in \mathbb{R}_+$ . Which leads to  $\left| \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right| \leq 0$ .

Each scenario leads to the same result:  $\left| \frac{\mathcal{T}_e^N f(y) - f(y)}{h_N} - \mathcal{A}_e f(y) \right| \leq \epsilon$ .

In conclusion, since the rank  $N_1$  is independent from the choice of  $y$ , one can deduce that  $\forall N \geq N_1$

$$\left\| \frac{\mathcal{T}_e^N f - f}{h_N} - \mathcal{A}_e f \right\| \leq \epsilon.$$

### If the kernel is an Erlang function

$E = \mathbb{R}_+^2$ . Let  $f \in \hat{C}_c^2(E)$  be a fixed function. We start by giving an alternative expression for  $\frac{\mathcal{T}_E^N f(y) - f(y)}{h_N}$  for a fixed  $(y, v) \in E$ :

$$\begin{aligned} \frac{\mathcal{T}_E^N f(y, v) - f(y, v)}{h_N} &= [f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + vhe^{-\beta h}, ve^{-\beta h})(1 - yh) \mathbb{1}_{yh < 1} \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) d\nu(z) \\ &\quad (yh \mathbb{1}_{yh < 1} + \mathbb{1}_{yh \geq 1}) - f(y, v)] h_N^{-1}, \\ &= [f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + vhe^{-\beta h}, ve^{-\beta h})(1 - yh) \\ &\quad + \int f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) d\nu(z) yh - f(y, v)] h_N^{-1} \mathbb{1}_{yh < 1} \\ &\quad + \left[ \int f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) - f(y, v) d\nu(z) \right] h_N^{-1} \mathbb{1}_{yh \geq 1} \end{aligned}$$

since  $\int d\nu(z) = 1$ .

Now using the fact that  $f$  is twice differentiable, we use Taylor expansion with a Lagrange remainder:

$$\begin{aligned} f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + vhe^{-\beta h}, ve^{-\beta h}) &= f(y, v) + ((\lambda_\infty - y)(1 - e^{-\beta h}) + vhe^{-\beta h}) \partial_\lambda f(y, v) \\ &\quad + v(e^{-\beta h} - 1) \partial_\xi f(y, v) + \frac{1}{2} v^2 (e^{-\beta h} - 1)^2 \partial_{\xi\xi} f(\theta_y, \theta_v) \\ &\quad + \frac{1}{2} ((\lambda_\infty - y)(1 - e^{-\beta h}) + vhe^{-\beta h})^2 \partial_{\lambda\lambda} f(\theta_y, \theta_v) \\ &\quad + ((\lambda_\infty - y)(1 - e^{-\beta h}) + vhe^{-\beta h}) v (e^{-\beta h} - 1) \partial_{\lambda\xi} f(\theta_y, \theta_v), \end{aligned}$$

where  $\theta_y \in [ye^{-\beta T}, \lambda_\infty(1 - e^{-\beta h}) + y + vh]$  and  $\theta_v \in [ve^{-\beta T}, v]$ .

Just like the exponential case, it is possible to bound all the second order terms by a constant, thus:

$$f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + vhe^{-\beta h}, ve^{-\beta h}) = f(y, v) + ((\lambda_\infty - y)(1 - e^{-\beta h}) + vhe^{-\beta h}) \partial_\lambda f(y, v)$$

$$+ v(e^{-\beta h} - 1)\partial_\xi f(y, v) + R_1(y, v)h^2,$$

where  $R_1$  is a compact support function that contains all the second derivatives. We apply another Taylor expansion to obtain:

$$f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) = f(y, v + \alpha z) + (v + \alpha z)(e^{-\beta h} - 1)\partial_\xi f(\gamma_y, \gamma_v) \\ + ((\lambda_\infty - y)(1 - e^{-\beta h}) + h(v + \alpha z)e^{-\beta h})\partial_\xi f(\gamma_y, \gamma_v),$$

where  $\gamma_y \in [ye^{-\beta T}, \lambda_\infty(1 - e^{-\beta h}) + y + (v + \alpha z)h]$  and  $\gamma_v \in [(v + \alpha z)e^{-\beta T}, (v + \alpha z)]$ . And it is possible to write it under the form:

$$f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) = f(y, v + \alpha z) + R_2(y, v, z)O(h),$$

where  $R_2$  is not necessarily bounded as  $z$  goes to infinity but it is not a problem since  $\int z d\nu(z)$  is bounded. Hence:

$$\begin{aligned} \frac{\mathcal{T}_E^N f(y, v) - f(y, v)}{h_N} \mathbb{1}_{yh < 1} &= [(\cancel{f(y, v)} + ((\lambda_\infty - y)(1 - e^{-\beta h}) + vhe^{-\beta h})\partial_\lambda f(y, v) \\ &\quad + v(e^{-\beta h} - 1)\partial_\xi f(y, v) + R_1(y, v)O(h^2) - \cancel{f(y, v)})(1 - yh)]h^{-1} \mathbb{1}_{yh < 1} \\ &\quad + [yh \int f(y, v + \alpha z) + R_2(y, v, z)O(h) - f(y, v)d\nu(z)]h^{-1} \mathbb{1}_{yh < 1}, \\ &= [((\lambda_\infty - y)(\beta h + O(h^2)) + vO(h))\partial_\lambda f(y, v) \\ &\quad + v(-\beta h + O(h^2))\partial_\xi f(y, v) + R_1(y, v)O(h^2)](1 - yh)h^{-1} \mathbb{1}_{yh < 1} \\ &\quad + [yh \int f(y, v + \alpha z) - f(y, v)d\nu(z) + h \int yR_2(y, v, z)d\nu(z)O(h)]h^{-1} \mathbb{1}_{yh < 1}, \\ &= \mathcal{A}_E f(y, v) \mathbb{1}_{yh < 1} + R(y, v)O(h) \mathbb{1}_{yh < 1}. \end{aligned}$$

The remainder  $R(y, v)$  contains  $R_1(y, v)$  and  $\int R_2(y, v, z)d\nu(z)$  as well as their products with  $y$  and  $v$ . It is a compact support functions thus it is bounded by a constant  $C$  independent from  $y, v$  and  $h$ .

Let  $\epsilon > 0$  and  $B$  a constant such that  $y > B$  or  $v > B$  implies  $f(y, v) = f'(y, v) = 0$ .

For all  $y \in \mathbb{R}_+$  we have:

$$\begin{aligned} \left| \frac{\mathcal{T}_E^N f(y) - f(y)}{h_N} - \mathcal{A}_E f(y) \right| &= \left| \left( \frac{\mathcal{T}_E^N f(y) - f(y)}{h_N} - \mathcal{A}_E f(y) \right) \mathbb{1}_{yh < 1} + \left( \frac{\mathcal{T}_E^N f(y) - f(y)}{y} - \mathcal{A}_E f(y) \right) \mathbb{1}_{yh \geq 1} \right| \\ &\leq \left| \left( \frac{\mathcal{T}_E^N f(y) - f(y)}{h_N} - \mathcal{A}_E f(y) \right) \mathbb{1}_{yh < 1} \right| + \left| \left( \frac{\mathcal{T}_E^N f(y) - f(y)}{h_N} - \mathcal{A}_E f(y) \right) \mathbb{1}_{yh \geq 1} \right|. \end{aligned}$$

Now we plug in the inequality obtained previously for the term in front of  $\mathbb{1}_{yh < 1}$  and expand the one in front of  $\mathbb{1}_{yh \geq 1}$ :

$$\begin{aligned} \left| \frac{\mathcal{T}_E^N f(y, v) - f(y, v)}{h_N} - \mathcal{A}_E f(y, v) \right| &\leq Ch \mathbb{1}_{yh < 1} \\ &\quad + \left| \left[ \int f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + h(v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) - f(y, v)d\nu(z) \right] h_N^{-1} \right. \\ &\quad \left. - \mathcal{A}_E f(y, v) \right| \mathbb{1}_{yh \geq 1} \end{aligned}$$

where  $O(h)$  has been absorbed by the constant  $C$ .

Let  $N_0$  be the integer such that if  $N \geq N_0$ , then  $\frac{N}{T}e^{-\beta \frac{T}{N}} \geq B$ . Such integer exists because  $\frac{N}{T}e^{-\beta \frac{T}{N}} \geq \frac{N}{T}e^{-\beta T} \rightarrow +\infty$  as  $N \rightarrow +\infty$ .

Set  $N_1 = \lfloor \frac{CT}{\epsilon} \rfloor + TB + N_0$ .

For every  $N \geq N_1$  and every  $(y, v) \in E$  only one of these two scenarios is possible:

- $\mathbb{1}_{yh < 1} = 1$  and  $\mathbb{1}_{yh \geq 1} = 0$ , so  $\left| \frac{\mathcal{T}_E^N f(y, v) - f(y, v)}{h_N} - \mathcal{A}_E f(y, v) \right| \leq \epsilon$  because  $N \geq N_1 \geq \lfloor \frac{CT}{\epsilon} \rfloor$ .

- $\mathbb{1}_{yh < 1} = 0$  and  $\mathbb{1}_{yh \geq 1} = 1$  which means  $y \geq \frac{N}{T}$ , thus  $y \geq B$  and  $ye^{-\beta h} \geq \frac{N}{T}e^{-\beta \frac{T}{N}} \geq B$ .  
Therefore  $f(\lambda_\infty(1 - e^{-\beta h}) + ye^{-\beta h} + (v + \alpha z)e^{-\beta h}, (v + \alpha z)e^{-\beta h}) = f(y, v) = \partial_\lambda f(y, v) = \partial_\xi f(y, v) = 0$   
 $\forall z \in \mathbb{R}_+$ .  
Which leads to  $|\frac{\mathcal{T}_E^N f(y, v) - f(y, v)}{h_N} - \mathcal{A}_E f(y, v)| \leq 0$ .

In conclusion

$$\|\frac{\mathcal{T}_E^N f - f}{h_N} - \mathcal{A}_E f\| \leq \epsilon.$$

□

## 4 Conclusion

We have proven that the DTHP converges weakly to a time continuous Hawkes process in the case the kernel is an exponential or an Erlang function. The following figure shows a trajectory of a DTHP with a small time step.

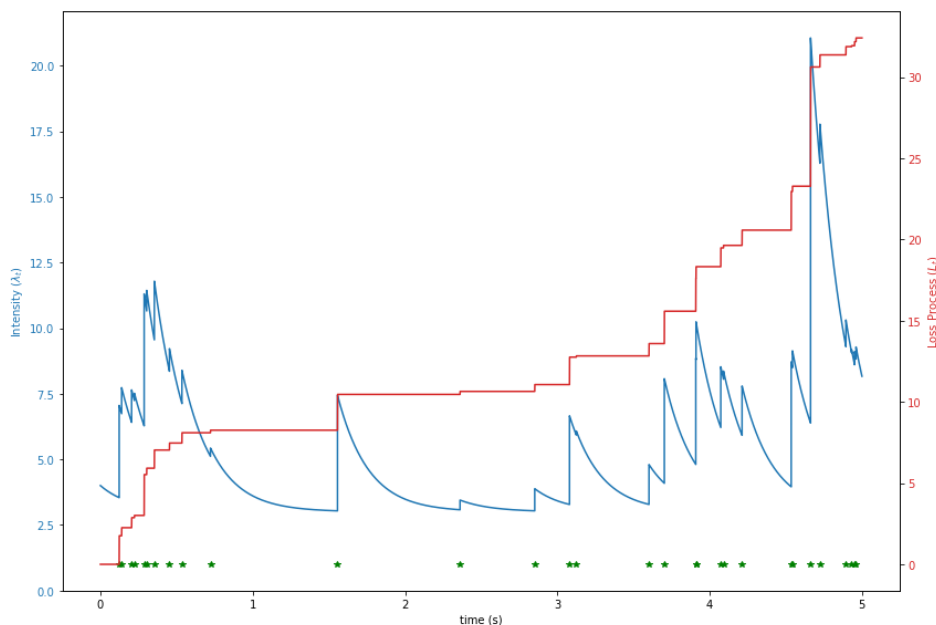


Figure 4: A trajectory of the loss process as well as the intensity in the case of an exponential kernel for  $\alpha = 2$ ,  $\beta = 5$ ,  $\lambda_\infty = 3$ ,  $\lambda_0 = 4$  and  $N = 100000$  points. The financial losses follow an exponential distribution of rate one. The green stars show the jumping times. Note how they are identical for  $L_t$  and  $\lambda_t$  and exhibit a clustering behaviour.

This result is generalisable to a wider class of Hawkes processes like the multivariate Hawkes process whose kernels are exponential/Erlang functions or in the case of a higher order Erlang kernel  $\phi(u) = \alpha u^n e^{-\beta u}$  with  $n \geq 2$ . However, despite being of the same nature, computations for these classes are way too heavy and repetitive to be included in this document.

It is also worth mentioning that this convergence does not have a quantified speed yet. It would be interesting to have an upper bound on the distance between the two processes as a function of the time step.

## References

- [1] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance, 2015.
- [2] Angelos Dassios and Hongbiao Zhao. Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Probab.*, 18:no. 62, 13, 2013.

- [3] Aline Duarte, Eva Löcherbach, and Guilherme Ost. Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. *ESAIM Probab. Stat.*, 23:770–796, 2019.
- [4] Eymen Errais, Kay Giesecke, and Lisa R. Goldberg. Affine point processes and portfolio credit risk. *SIAM J. Financial Math.*, 1(1):642–665, 2010.
- [5] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
- [6] Alan. G. Hawkes. Spectra of some mutually exciting point processes with associated variables. In *Stochastic point processes: statistical analysis, theory, and applications (Conf., IBM Res. Center, Yorktown Heights, N.Y., 1971)*, pages 261–271. 1972.
- [7] Thibault Jaisson and Mathieu Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.*, 25(2):600–631, 2015.
- [8] Jesper Møller and Jakob G. Rasmussen. Perfect simulation of Hawkes processes. *Adv. in Appl. Probab.*, 37(3):629–646, 2005.
- [9] Yoshihiko Ogata. On lewis’ simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.
- [10] Youngsoo Seol. Limit theorems for discrete Hawkes processes. *Statist. Probab. Lett.*, 99:223–229, 2015.