Maximum Likelihood Estimation for Hawkes Processes with self-excitation or inhibition

Anna Bonnet, Miguel Martinez Herrera, Maxime Sangnier

Maximum Likelihood Estimation for Hawkes Processes with self-excitation or inhibition

Anna Bonnet, Miguel Martinez Herrera, Maxime Sangnier*

March 8, 2021

Abstract

In this paper, we present a maximum likelihood method for estimating the parameters of a univariate Hawkes process with self-excitation or inhibition. Our work generalizes techniques and results that were restricted to the self-exciting scenario. The proposed estimator is implemented for the classical exponential kernel and we show that, in the inhibition context, our procedure provides more accurate estimations than current alternative approaches.

1 Introduction

The Hawkes model is a point process observed on the real line, which generally corresponds to the time, where any previously encountered event has a direct influence on the chances of future events occurring. This past-dependent mathematical model was introduced in [1] and its first application was to model earthquakes occurrences [2, 3]. Since then, Hawkes processes have been widely used in various fields, for instance finance [4], social media [5, 6], epidemiology [7], sociology [8] and neuroscience [9].

The main advantage of Hawkes processes is their ability to model different kinds of relationships between phenomena through an unknown kernel or transfer function. The Hawkes model was originally introduced as a self-exciting point process where the appearance of an event increases the chances of another one triggering. Several estimation procedures have been proposed for the kernel function, both in parametric [2, 10, 11] and nonparametric [9, 12] frameworks.

However, the inhibition setting, where the presence of an event decreases the chance of another occurring, has drawn less attention in the literature, although it can be of great interest in several fields, in particular in neuroscience [13]. In this inhibition context, the cluster representation [14] on which is based the construction of a self-exciting Hawkes process, is no longer valid. While the existence and the construction of such nonlinear processes can be found in recent

*Sorbonne Université, CNRS, LPSM, Paris, France

works for the univariate [15] and multivariate [16] cases, statistical estimation of the kernel function has been hardly addressed. A first approach consists in computing an approximation of the likelihood as if the intensity function could take negative values, and optimizing it to get a maximum likelihood estimator [17]. In an other way, the type of interaction (excitation or inhibition) can be considered as a hidden variable, giving rise to a very practical estimation method [18].

In this paper, we propose a maximum likelihood procedure that can handle both excitation and inhibition scenarios for a univariate Hawkes process. Our approach is based on an explicit computation of the likelihood for any type of kernel functions, which is facilitated by the introduction of the natural concept of restart points. The latters are the times when the intensity function, that can be null on some intervals, become strictly positive again. We show that these restart points have a closed-form expression when the kernel is exponential, which allows us to rewrite and maximize the likelihood without approximations that are proposed for instance in [17]. Our estimator is implemented in Python (the code is freely available online[1]). We also propose a numerical study which shows the good performance of our exact estimation procedure compared to approximated approaches, especially when the intensity function is frequently equal to zero.

To outline the paper, besides a quick introduction to self-regulating Hawkes processes (or Hawkes processes with inhibition), Section 2 introduces the concepts of underlying intensity function and restart points. General results concerning the compensator and the exact maximum likelihood estimation procedure are described in Section 3. At last, after a brief discussion about goodness-of-fit in Section 4, Section 5 concludes with a numerical study of the estimation error.

2 The Hawkes process

Let N be a point process on $\mathbb{R}_+^*$ and $(T_k)_{k\geq 1}$ its associated event times (with convention $T_0 = 0$). For any $t \geq 0$, let us note $N(t) = \sum_{k\geq 1} \mathbb{1}_{T_k \leq t}$ the number of events in $[0, t]$, and λ its conditional intensity function [19, 20]:

$$\lambda(t) = \lim_{h\to 0} \frac{\mathbb{P}(N(t+h) - N(t) > 0)}{h}.$$

A univariate Hawkes process is a point process defined by the conditional intensity function:

$$\lambda(t) = \left(\lambda_0 + \int_{-\infty}^{t} h(t-s)\, dN(s)\right)^+ = \left(\lambda_0 + \sum_{T_k \leq t} h(t - T_k)\right)^+, \qquad (1)$$

where $x^+ = \max(0, x)$ denotes the positive part of any real value x, $\lambda_0 \in \mathbb{R}_+^*$ is the baseline intensity and $h : \mathbb{R} \to \mathbb{R}$ is the kernel, which is assumed to be a

[1] `https://github.com/migmtz/hawkes-inhibition-expon`

monotone measurable function with $\lim_{t\to+\infty} h(t) = 0$. The kernel function h is the key component of a Hawkes process: it translates the influence (generally assumed to fade away over time) of a past event over the process. Here, h is allowed to take negative values, meaning that it can model both the self-exciting and self-regulating Hawkes processes.

Working with such Hawkes processes may prove to be difficult as the positive part function is non-linear. In particular, while computing the compensator function [19, 20]

$$\Lambda(t) = \int_0^t \lambda(t)\, dt, \quad \forall t \geq 0, \tag{2}$$

is very easy in the self-exciting case (by linearity of the intensity), it becomes more challenging for the self-regulating Hawkes process. As it is the keystone to derive the likelihood function (and then to obtain a parametric estimation method), our first contribution is to provide an exact expression of the compensator.

For this purpose, let us first introduce the *underlying intensity function* and the *restart time*, two quantities which will allow us to extend to monotone Hawkes processes the classical techniques that are used for self-exciting processes.

Definition 2.1. Let the *underlying intensity function* of N be:

$$\lambda^\star(t) = \lambda_0 + \int_{-\infty}^t h(t-s)\, dN(s).$$

In addition, let the *restart time* T_k^* be, for any positive integer k:

$$T_k^\star = \inf\{t \geq T_k \mid \lambda(t) > 0\},$$

along with its corresponding *cooldown interval* $\tau_k^* = T_k^* - T_k$.

As illustrated in Figure 1, $\lambda^\star$ corresponds to the intensity λ as if it were allowed to take negative values. Moreover, as the kernel is assumed to be monotone, the restart time associated to one occurrence can be interpreted as the first moment after this occurrence from which λ and $\lambda^\star$ become equal (in particular, the restart time and the occurrence time coincide if the intensity function is nonnegative at this time, see Figure 1):

$$T_k^\star = \inf\{t \geq T_i \mid \forall t \in (T_k^*, T_{k+1}),\, \lambda(t) = \lambda^\star(t)\}.$$

3 Maximum likelihood estimation and the exponential model

Assume a parametric model $\mathcal{P} = \{\lambda_\theta, \theta \in \Theta\}$ for the conditional intensity function λ, where θ contains unknown quantities such as the baseline λ_0 and the kernel h. Then, with convention $log(t) = -\infty$ for $t \leq 0$, the log-likelihood ℓ_t

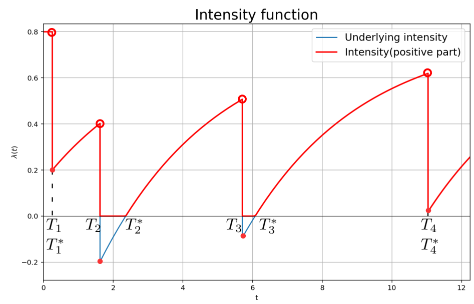

Figure 1: Example of the intensity (red curve) and underlying intensity (blue curve) for a self-regulating Hawkes process, with the associated restart times. We only see the negative values of the blue curve since they precisely correspond to the values for which the two intensity functions are not equal.

of any $\theta \in \Theta$ with respect to the observations $T_1, \dots, T_{N(t)}$ in the time interval $[0, t]$ is [19, Proposition 7.2.III.], [11]:

$$\ell_t(\theta) = \sum_{k=1}^{N(t)} \log\left(\lambda_\theta(T_k^-)\right) - \Lambda_\theta(t), \tag{3}$$

where the compensator Λ_θ is defined as in Equation (2) and $\lambda_\theta(T_k^-) = \lim_{t \to T_k^-} \lambda_\theta(t)$.

Equation (3) reveals the importance of being able to compute the compensator Λ (equivalently Λ_θ) in order to provide a practical implementation of the maximum likelihood estimator of λ. Thus, a first contribution of this paper lies in Proposition 3.1, which establishes a decomposition of the compensator Λ using the underlying intensity function $\lambda^\star$ and the restart times $T_1^\star, \dots, T_{N(t)}^\star$.

Proposition 3.1. *For any $t > 0$, the compensator Λ can be expressed as:*

$$\Lambda(t) = \begin{cases} \lambda_0 t & \text{if } t < T_1 \\ \lambda_0 T_1 + \sum_{k=2}^{N(t)} \int_{T_{k-1}^\star}^{T_k} \lambda^\star(u)\, du + \int_{T_{N(t)}^\star}^{t} \lambda^\star(u)\, du & \text{if } t \geq T_1, \end{cases} \tag{4}$$

with the conventions that the sum is equal to 0 if $N(t) = 1$ and the last integral is equal to 0 if $t < T_{N(t)}^*$.

Proof. This comes directly from splitting the integral of $\Lambda(t) = \int_0^t \lambda(t)\, dt$ on the intervals $[T_i, T_{i+1})$ ($i \in \{0, \dots, N(t) - 1\}$) and $[T_{N(t)}, t]$, and by remarking that, since h is monotone, $\forall t \in [T_i, T_{i+1})$, $\lambda(t) = \lambda^\star(t) \mathbb{1}_{[T_i^*, T_{i+1})}(t)$. $\square$

In order to give an explicit computation of the quantity $\int_{T_{k-1}^\star}^{T_k} \lambda^\star(u)\, du$ (equivalently $\int_{T_{N(t)}^\star}^{t} \lambda^\star(u)\, du$) which appears in Proposition 3.1, we focus on the classical scenario where we consider an exponential kernel $h(t) = \alpha e^{-\beta t}$, for some $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_+^*$. Let us notice that α can be either positive or negative, meaning that the process may be either self-exciting or self-regulating.

Then, the underlying intensity function can be written as:

$$\lambda^\star(t) = \lambda_0 + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)}\, dN(s). \tag{5}$$

The forthcoming proposition steps forward in computing the compensator for an exponential kernel.

Proposition 3.2 (Compensator for exponential kernel)**.** *Let $t > 0$ and $k \in \{1, \ldots, N(t)+1\}$. The restart times read:*

$$T_k^\star = T_k + \beta^{-1} \log\left(\frac{\lambda_0 - \lambda^\star(T_k)}{\lambda_0}\right),$$

and the compensator is expressed as in Equation (4), with, for any $\tau \in [T_{k-1}^*, T_k]$:

$$\int_{T_{k-1}^\star}^{\tau} \lambda^\star(u)\, du = \lambda_0(\tau - T_{k-1}^\star) + \beta^{-1}(\lambda^\star(T_{k-1}) - \lambda_0)(e^{-\beta(T_{k-1}^\star - T_{k-1})} - e^{-\beta(\tau - T_{k-1})}).$$

Proof. The proof is in A. □

Corollary 3.1 (Log-likelihood for exponential kernel)**.** *Let*

$$\mathcal{P} = \left\{\lambda_\theta = \bar{\lambda}_0 + \int_{-\infty}^{t} \bar{\alpha} e^{-\bar{\beta}(t-s)}\, dN(s) : \theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta}) \in \Theta\right\}, \tag{6}$$

be a parametric exponential model for the conditional intensity function λ with $\Theta = \mathbb{R}_+^* \times \mathbb{R} \times \mathbb{R}_+^*$, along with the candidate compensator Λ_θ, the underlying intensity function $\lambda_\theta^\star$ and the restart times $T_{\theta,1}^\star, \ldots, T_{\theta,N(t)}^\star$ associated to λ_θ (see Equation (2) and Definition 2.1).

For any $\theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta}) \in \Theta$, by denoting

$$\Lambda_{\theta,k} = \bar{\lambda}_0(T_k - T_{\theta,k-1}^\star) + \bar{\beta}^{-1}(\lambda_\theta^\star(T_{k-1}) - \bar{\lambda}_0)(e^{-\bar{\beta}(T_{\theta,k-1}^\star - T_{k-1})} - e^{-\bar{\beta}(T_k - T_{k-1})}),$$

the log-likelihood reads (with convention $\log(x) = -\infty$ for $x \leq 0$):

$$\begin{aligned}\ell_t(\theta) = {}& \log \bar{\lambda}_0 - \bar{\lambda}_0 T_1 + \sum_{k=2}^{N(t)} \left[\log\left(\bar{\lambda}_0 + (\lambda_\theta^\star(T_{k-1}) - \bar{\lambda}_0) e^{-\bar{\beta}(T_k - T_{k-1})}\right) - \Lambda_{\theta,k}\right] \\ & - \left[\bar{\lambda}_0(t - T_{\theta,N(t)}^\star) + \bar{\beta}^{-1}(\lambda_\theta^\star(T_{N(t)}) - \bar{\lambda}_0)\left(e^{-\bar{\beta}(T_{\theta,N(t)}^\star - T_{N(t)})} - e^{-\bar{\beta}(t - T_{N(t)})}\right)\right] \mathbb{1}_{t > T_{\theta,N(t)}^\star}.\end{aligned} \tag{7}$$

Proof. The proof of Proposition 3.2 reveals that

$$\lambda_\theta^\star(T_k^-) = \begin{cases} \bar{\lambda}_0 & \text{if } k = 1. \\ \bar{\lambda}_0 + (\lambda_\theta^\star(T_{k-1}) - \bar{\lambda}_0) e^{-\bar{\beta}(T_k - T_{k-1})} & \text{if } k \geq 2. \end{cases}$$

Combining this expression with Propositions 3.1 and 3.2 leads to the result. □

Corollary 3.1 exhibits that the log-likelihood for self-regulating Hawkes processes with an exponential kernel can be evaluated in $O(N(t))$ operations (by computing iteratively the quantities $T_{\theta,k}^\star$ and $\Lambda_{\theta,k}$ appearing in the summation of Equation (7)), as already known for self-exciting exponential Hawkes processes [21, Chapter 4.2]. For other kernels without the Markov property, evaluating the log-likelihood with the method proposed here requires $O(N(t)^2)$ operations, similarly to existing approaches for self-exciting Hawkes processes.

4 Goodness-of-fit

Even though computing the compensator Λ (equivalently Λ_θ) was clearly motivated by maximum likelihood estimation, it turns out that it is of great benefit to assess goodness-of-fit, and in particular to check the validity of a maximum likelihood estimators (as done for self-exciting Hawkes processes in [21, Chapter 5]). This is possible thanks to the Time Change Theorem, a result originally stated for inhomogeneous Poisson processes.

Theorem 4.1 ([19, Theorem 7.4.IV])**.** *Assume that Λ is continuous, monotone and $\Lambda(t) \xrightarrow[t\to+\infty]{} +\infty$ a.s. Then a.s., $(U_k)_{k\geq 1}$ is a realization of event times of N if and only if $(\Lambda(U_k))_{k\geq 1}$ is a realization of a homogeneous Poisson process with unit intensity $\lambda = 1$.*

5 Numerical Results

This section is aimed at assessing the maximum likelihood estimation method for self-regulating Hawkes processes, based on the exact computation of the compensator Λ_θ in the exponential model (6) (Corollary 3.1). This procedure is compared to the approximated maximum likelihood estimation proposed in [17], which consists in approximating Λ_θ by:

$$\Lambda_\theta^{LM}(t) = \int_0^t \lambda_\theta^\star(u)\, du.$$

The optimization procedure is performed with the Nelder-Mead simplex algorithm and we consider an ℓ_2-penalized version of the likelihood in addition to vanilla likelihood in order to avoid numerical instabilities. In other words, estimators are:

$$\hat{\theta} \in \arg\max_{\theta\in\Theta} \left\{ \ell_{T_{N_{max}}}(\theta) - C\|\theta\|_2^2 = \left(\sum_{k=1}^{N_{max}} \log\left(\lambda_\theta(T_k^-)\right) - \Lambda_\theta(T_{N_{max}}) \right) - C\|\theta\|_2^2 \right\},$$

where $C \in \{0, 0.1\}$ is the penalization coefficient, $N_{max} = 150$ is the total number of jumps and Λ_θ can be replaced by Λ_θ^{LM} to obtain the approximated likelihood proposed in [17]. The details for the numerical calibration of the regularization parameter C can be found in C.1.

The comparison between the exact and the approximated estimation procedures is based on simulated data sets coming from self-regulating Hawkes processes of the form (6) with 5 different values of $\theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta}) \in \Theta$ which have been chosen in order to explore different scenarii, in particular depending on whether the intensity function is frequently null or not. Observations are sets of time jumps generated with a sampling algorithm (see the algorithm in B and implementation online[2]), which is a particular case of Ogata's thinning simulation method [22] that can handle Hawkes processes with either self-excitation or inhibition. For each the 5 models, statistics are provided based on 100 random repetitions.

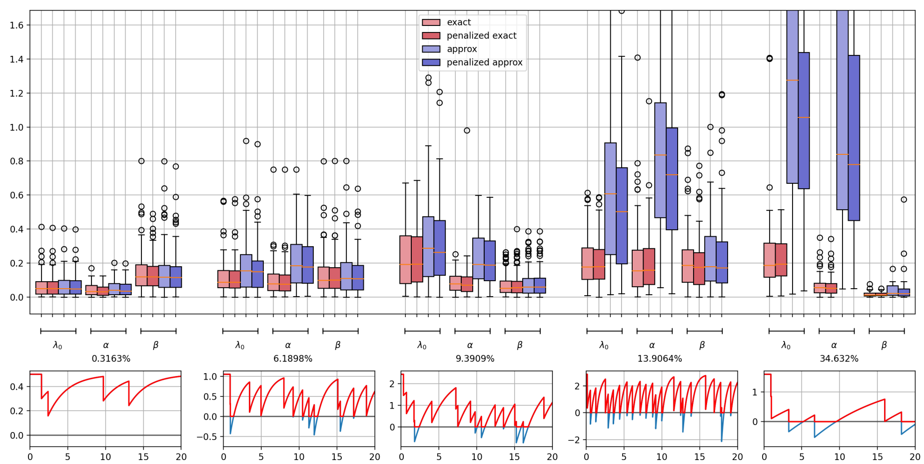

Figure 2: Top panel: absolute errors of estimations $\hat{\theta} = (\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$. Bottom panel: example of simulated intensities for each set of values $\theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta})$ with the corresponding average percentage of time when the intensities are equal to zero.

Figure 2 represents the absolute errors of estimations $\hat{\theta} = (\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$ for each of the 5 simulated models. We observe that the exact approach provides more accurate estimations than the approximated procedure. As expected, the more time the conditional intensity equals 0 (from left to right in Figure 2), the greater the differences between the two procedures.

Finally, we observe that penalized estimation allows to get rid of the few outlier estimations as we can see on Figure 3. Such estimations achieve high p-values when considering a Kolmogorov-Smirnov test associated to the null hypothesis "The transformed inter-arrival times are i.i.d. following an exponential

[2] `https://github.com/migmtz/hawkes-inhibition-expon`

distribution of parameter 1". For instance, after performing such goodness-of-fit tests, we obtain an average p-value among the five estimations of 0.712.

6 Discussion

In this paper we proposed a maximum likelihood approach for Hawkes processes that can handle both self-exciting and self-regulating scenarii, the first case being already covered in the literature and the latter being our main contribution. For this purpose, we define the concepts of underlying intensity function and restart times when working with monotone kernel functions. In particular we obtain exact expressions of the compensator for the exponential Hawkes process which is the key step of the estimation procedure. Our estimation method also contains a ℓ_2-penalization which ensures the numerical stability of the procedure. We present numerical results on synthetic data that show the efficiency of our procedure, with a substantial improvement compared to approximated approaches when the intensity function is frequently null.
From a theoretical point of view, future work will consist in adapting analytical results to study the convergence of our estimator in the self-regulating case. Regarding modeling, it would be of great interest to consider kernel functions outside the classical exponential scenario. Another important step is the extension of our concepts and algorithms to the multivariate version of the process, which is not straightforward since in the multivariate setting the expression of the restart times are no longer explicit. This last point is the object of a future work, with a further perspective to use our procedure in neuroscience applications in order to detect attraction and repulsion effects between neurons.

References

[1] A. G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1) (1971) 83–90.

[2] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, Journal of the American Statistical Association 83 (1988) 9–27.

[3] Y. Ogata, Space-time point-process models for earthquake occurrences, Annals of the Institute of Statistical Mathematics 50 (1998) 379–402.

[4] E. Bacry, S. Delattre, M. Hoffmann, J. Muzy, Scaling limits for hawkes processes and application to financial statistics, Stochastic Processes and Applications 123 (2013) 2475–2499.

[5] M. Rizoiu, Y. Lee, S. Mishra, L. Xie, A tutorial on hawkes processes for events in social media (2017). `arXiv:1708.06401`.

[6] S. Mishra, M. Rizoiu, L. Xie, Feature driven and point process approaches for popularity prediction (2016). `arXiv:1608.04862v2`.

[7] M. Rizoiu, S. Mishra, Q. Kong, M. Carman, L. Xie, Sir-hawkes: linking epidemic models and hawkes processes to model diffusions in finite populations, in: Proceedings of the 2018 World Wide Web Conference, International World Wide Web Conferences Steering Committee, 2018, pp. 419–428.

[8] S. Linderman, R. Adams, Discovering latent network structure in point process data, in: Proceedings of the 31st International Conference on Machine Learning, Proceedings of Machine Learning Research, 2014, pp. 1413–1421.

[9] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, C. Tuleau-Malot, Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis, The Journal of Mathematical Neuroscience 4 (2014) 3.

[10] J. Da Fonseca, R. Zaatour, Hawkes process: Fast calibration, application to trade clustering, and diffusive limit, Journal of Futures Markets 34 (2013) 548–579.

[11] T. Ozaki, Maximum likelihood estimation of hawkes' self-exciting point processes, Annals of the Institute of Statistical Mathematics 31 (1979) 145–155.

[12] E. Bacry, J. Muzy, Second order statistics characterization of hawkes processes and non-parametric estimation (2015). `arXiv:1401.0903v2`.

[13] P. Reynaud-Bouret, R. Lambert, C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, L. Leresche, Reconstructing the functional connectivity of multiple spike trains using hawkes models, Journal of Neuroscience Methods 297 (2018) 9–21.

[14] A. G. Hawkes, D. Oakes, A cluster process representation of a self-exciting process, Journal of Applied Probability 11 (1974) 493–503.

[15] M. Costa, C. Graham, L. Marsalle, V. Tran, Renewal in hawkes processes with self-excitation and inhibition (2018). `arXiv:1801.04645v2`.

[16] S. Chen, A. Shojaie, E. Shea-Brown, D. Witten, The multivariate hawkes process in high dimensions: Beyond mutual excitation (2017). `arXiv:1707.04928v2`.

[17] R. Lemonnier, N. Vayatis, Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate hawkes processes, in: Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, 2014, p. 161–176.

[18] H. Mei, J. Eisner, The neural hawkes process: A neurally self-modulating multivariate point process (2017). `arXiv:1612.09328v3`.

[19] D. J. Daley, D. Vere-Jones, An introduction to the theory of point processes. Vol. I, 2nd Edition, Probability and its Applications (New York), Springer-Verlag, 2003.

[20] D. J. Daley, D. Vere-Jones, An introduction to the theory of point processes. Vol. II, 2nd Edition, Probability and its Applications (New York), Springer-Verlag, 2008.

[21] P. Laub, Hawkes processes: Simulation, estimation, and validation (2014).

[22] Y. Ogata, On lewis' simulation method for point processes, IEEE Transactions on Information Theory 27 (1981) 23–30.

A Proof of Proposition 3.2

Let us begin by rewriting the underlying intensity function between two event times. The method is the same commonly used when working with the exponential kernel. Let us notice that, in our particular case:

$$\lambda^\star(t) = \lambda_0, \quad \text{for } t \in [0, T_1).$$

Let us work for any $k \in \mathbb{N}$ within the interval $[T_k, T_{k+1})$. In this interval, the underlying intensity is differentiable which allows us to obtain the following differential equation:

$$(\lambda^\star)'(t) = -\beta(\lambda^\star(t) - \lambda_0),$$

with a left condition which will note $\lambda_k^\star := \lambda^\star(T_k)$. By solving this problem, we obtain the following expression:

$$\lambda^\star(t) = \lambda_0 + (\lambda_k^\star - \lambda_0)e^{-\beta(t-T_k)}. \tag{8}$$

Now, by definition of the restart time $T_k^\star$, which we recall here:

$$T_k^\star = \inf\{t \geq T_k \mid \lambda(t) > 0\},$$

we can see that if $\lambda_k^\star \geq 0$, then $T_k^\star = T_k$. Otherwise, as $\lambda^\star$ is continuous on the interval $[T_k, T_{k+1})$, the restart time can be obtained by solving for t:

$$\lambda^\star(t) = 0.$$

We can then use Equation (8) to obtain:

$$\begin{aligned} & \lambda^\star(T_k^\star) = 0 \\ \iff \quad & \lambda_0 + (\lambda_k^\star - \lambda_0)e^{-\beta(T_k^\star - T_k)} = 0 \\ \iff \quad & T_k^\star = T_k + \beta^{-1}\log\left(\frac{\lambda_0 - \lambda_k^\star}{\lambda_0}\right). \end{aligned}$$

Each restart time can be expressed then as:

$$T_k^\star = \begin{cases} T_k & \text{if } \lambda_k^\star \geq 0. \\ T_k + \beta^{-1}\log\left(\frac{\lambda_0 - \lambda_k^\star}{\lambda_0}\right) & \text{if } \lambda_k^\star < 0. \end{cases}$$

With the expression of the restart times, we can then establish an explicit form of the (log-)likelihood with exponential kernel. For this, we need to estimate the value of $\lambda(T_k^-)$ and Λ_k for any $k \in \mathbb{N}$. For both, we will be using Equation (8). Firstly,

$$\lambda^\star(T_k^-) = \begin{cases} \lambda_0 & \text{if } k = 1. \\ \lambda_0 + (\lambda_{k-1}^\star - \lambda_0)e^{-\beta(T_k - T_{k-1})} & \text{if } k \geq 2. \end{cases} \tag{9}$$

The interest of using this form is that it provides an iterative algorithm to compute $\lambda(T_k^-)$, as $\lambda_k^\star = \lambda^\star(T_k^-) + \alpha$.

Secondly, the same Equation (8) is used for the integral. As we assume that $t > T_1$, it follows that $\Lambda_1 = \lambda_0 T_1$ (as for Theorem 3.1). For any $k \in \{2, \ldots, N(t)+1\}$ and for any $\tau \in [T_{k-1}^\star, T_k]$,

$$\begin{aligned}\int_{T_{k-1}^\star}^{\tau} \lambda^\star(u)\, du &= \int_{T_{k-1}^\star}^{\tau} \left(\lambda_0 + (\lambda_{k-1}^\star - \lambda_0) e^{-\beta(u - T_{k-1})}\right) du \\ &= \lambda_0(\tau - T_{k-1}^\star) + \beta^{-1}(\lambda_{k-1}^\star - \lambda_0)(e^{-\beta(T_{k-1}^\star - T_{k-1})} - e^{-\beta(\tau - T_{k-1})}),\end{aligned} \tag{10}$$

B Simulation algorithm

Algorithm 1 builds upon Ogata's thinning simulation method [22, Proposition 1] in order to handle Hawkes processes with either self-excitation or inhibition.

Algorithm 1: Thinning algorithm for monotone Hawkes process

Input Parameters λ_0, h a monotone function, and a stopping criteria (end-time T or maximal number of jumps N_{max});
Initialization Initialize $\lambda_k = \lambda_0$, $t_k = 0$ and list of times $\mathcal{T} = \emptyset$;
while Stopping criteria not fulfilled **do**
  Set $\lambda_{max} = \max(\lambda_0, \lambda_k)$;
  Generate candidate time $t_{cand} = t_k - \frac{\log(U_1)}{\lambda_{max}}$, $U_1 \sim U([0,1])$;
  Estimate intensity $\lambda_k = \lambda(t)$ using sequence of times $\mathcal{T}$;
  Sample $U_2 \sim U([0,1])$;
  if $U_2 \leq \frac{\lambda_k}{\lambda_{max}}$ **then**
    Add t to sequence of times $\mathcal{T}$;
  end
  Set $t_k = t_{cand}$
end
return the sequence of jumps $\mathcal{T}$

C Additional numerical results

C.1 Numerical study of the penalization

Let us assess the impact of penalization using the likelihood as presented in Equation (7). For this purpose, for each of the 5 sets of parameters $(\bar{\lambda}_0, \bar{\alpha}, \bar{\beta})$, we generated 100 realizations of Hawkes processes. Each realization has a total of 500 event times and we compare the non-penalized estimation and the penalized one with four different values for the constant C. For each set of parameters, we compute the mean relative error ε_θ over the 100 realizations defined as:

$$\varepsilon_{\hat{\theta}} = \frac{|\bar{\lambda}_0 - \hat{\lambda}_0|}{|\bar{\lambda}_0|} + \frac{|\bar{\alpha} - \hat{\alpha}|}{|\bar{\alpha}|} + \frac{|\bar{\beta} - \hat{\beta}|}{|\bar{\beta}|}.$$

$\bar{\lambda}_0$	$\bar{\alpha}$	$\bar{\beta}$	Non-Penalized	$C = 0.01$	$C = 0.1$	$C = 1$	$C = 10$
0.5	-0.2	0.4	0.137319	0.142888	0.143537	0.132592	0.122738
1.05	-0.75	0.8	0.067637	0.075153	0.074581	0.063276	0.093864
2.43	-0.98	0.4	0.068603	0.060913	0.060279	0.073221	0.306698
2.85	-2.5	1.8	0.073671	0.065809	0.064419	0.076797	0.419435
1.6	-0.75	0.1	10^{24}	0.074167	0.061362	0.058963	0.180468

Table 1: Mean relative error over 100 realizations of Hawkes processes generated for each set of parameters $(\bar{\lambda}_0, \bar{\alpha}, \bar{\beta})$.

We observe (see Table 1) a few abnormal estimation when using the non-penalized MLE, as shown by the relative squared error obtained with parameters $(1.6, -0.75, 0.1)$, which disappear with the penalization. Let us also notice that overall, each time the penalized version is able to obtain better results that the non-penalized one, even if the best value of the constant C is not always the same.

C.2 Comparison of estimations

This section provides an additional figure regarding the numerical comparison presented in Section 5. It is aimed at highlighting the presence of outlier estimations when considering the non-penalized estimators.

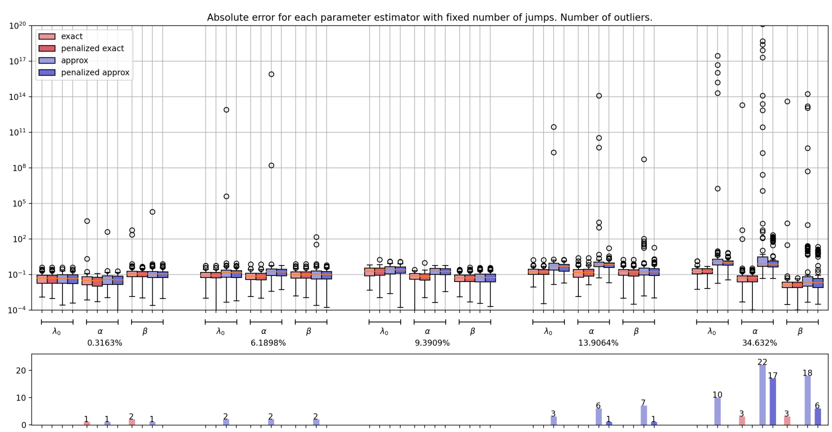

Figure 3: Top panel: absolute errors of estimations $\hat{\theta} = (\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$ in logarithmic scale. Bottom panel: number of abnormal estimations (absolute error greater than 10^2) and average percentage of time that the intensities equal 0.