FaDIn : Inférence discrétisée efficace pour les processus de Hawkes avec noyaux paramétriques génériques

Guillaume STAERMAN Cédric ALLAIN Alexandre GRAMFORT Thomas MOREAU
Mind, INRIA Saclay, Université Paris-Saclay, 1 rue Honoré Estienne d'Orves, 91120 Saclay, France

Résumé – Les processus de Hawkes se sont avérés être le modèle de processus ponctuels temporels (TPP) le plus largement utilisé, principalement en raison de leur modélisation adéquate pour diverses applications, en particulier lorsque l'on considère des noyaux exponentiels ou non paramétriques. Cependant, ces deux méthodes sont mal adaptées aux applications où des latences doivent être estimées, ou encore dans le cas où peu de données sont disponibles, comme c'est le cas en neurosciences. Ainsi, ce travail vise à offrir une méthode efficace d'inférence de TPP basée sur la discrétisation, en utilisant des noyaux paramétriques génériques à support fini.

Abstract – Hawkes processes have proven to be the most widely used temporal point processes model (TPP), mainly due to their adequate modeling for various applications, particularly when considering exponential or non-parametric kernels. However, these two methods are ill-suited for applications where latencies need to be estimated, or in the case of few available data, such as in neuroscience. Thus, this work aims to offer an efficient discretization-based solution to TPP inference using general parametric kernels with finite support.

1 Introduction

Les Processus ponctuels temporels (TPP ; 4) offrent un cadre statistique adapté pour modéliser des données basées sur des événements. Ils permettent de prédire de manière rigoureuse la probabilité d'apparition d'un événement en fonction du temps et des événements passés. Les TPP trouvent notamment de nombreuses applications en neurosciences, particulièrement pour modéliser les enregistrements de cellules individuelles et les trains d'impulsions neurales, parfois associés à des statistiques spatiales ou des modèles de réseau. Les processus de Hawkes multivariés (MHP ; 7) sont probablement les modèles de TPP les plus populaires, car ils peuvent modéliser les interactions entre de multiples processus univariés, ainsi que le comportement d'auto-excitation. La méthode d'inférence la plus couramment utilisée pour estimer les paramètres du modèle est le maximum de vraisemblance (MLE ; 4). Un critère d'estimation alternatif, et souvent négligé, est l'erreur des moindres carrés ℓ_2, inspiré par la théorie de la minimisation du risque empirique (ERM ; 12).

L'une des caractéristiques principales de la modélisation des MHP est le choix des noyaux représentant la manière dont les processus s'influencent mutuellement, qui peuvent être paramétriques ou non. Le cadre non paramétrique permet une grande flexibilité pour la forme du noyau, mais au risque d'une mauvaise estimation de celui-ci lorsque seule une petite quantité de données est disponible [13]. La méthode paramétrique, bien que pouvant introduire un biais en supposant une forme de noyau définie, présente plusieurs avantages : i) la réduction de la complexité d'inférence - le paramètre étant généralement de dimension inférieure à celle des noyaux non paramétriques -, ainsi que, ii) pour les noyaux satisfaisant la propriété de Markov [2], une intensité conditionnelle dont le calcul est linéaire avec le nombre total d'événements. Pour les noyaux paramétriques plus généraux qui ne vérifient pas la propriété de Markov, la procédure d'inférence avec la perte MLE ou ℓ_2 s'échelonne mal, car quadratique avec le nombre d'événements, ce qui rend leur utilisation limitée en pratique (voir 3, Chapter 1).

Cet article propose une nouvelle méthode d'inférence - appelée FaDIn - pour estimer n'importe quel noyau paramétrique dans le cadre des processus de Hawkes. Notre approche repose sur deux caractéristiques essentielles. Premièrement, nous utilisons des noyaux à support fini et une discrétisation appliquée à la perte des moindres carrés inspirée de l'ERM. Deuxièmement, nous proposons d'utiliser des pré-calculs qui réduisent considérablement le coût de calcul. Appliquée aux neurosciences, la flexibilité de FaDIn permet de modéliser la réponse neuronale consécutive à des stimuli externes à l'aide d'un noyau bien mieux adapté que la méthode existante dérivée de [1]. La méthode est évaluée sur des données simulées ainsi que sur deux jeux de données magnétoencéphalographie (MEG).

2 Inférence discrétisée rapide pour les processus de Hawkes (FaDIn)

Un processus ponctuel temporel est un processus stochastique dont la réalisation consiste en des événements discrets $\{t_n\}$ se produisant en temps continu, $t_n \in \mathbb{R}_+$ [4]. Dans le cas où la probabilité qu'un événement se produise au moment t dépend uniquement des événements passés $\mathscr{F}_t := \{t_n,\ t_n < t\}$, les PP sont généralement caractérisés par la fonction d'intensité conditionnelle $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$:

$$\lambda\left(t|\mathscr{F}_t\right) := \lim_{\mathrm{d}t \to 0} \frac{\mathbb{P}\left(N_{t+\mathrm{d}t} - N_t = 1|\mathscr{F}_t\right)}{\mathrm{d}t} \ , \qquad (1)$$

où $N_t := \sum_{i \geq 1} \mathbb{1}_{\{t_n \leq t\}}$ désigne le processus de comptage associé au PP. Cette fonction correspond, en espérance, au taux infinitésimal auquel les événements se produisent au temps t étant donné les temps d'arrivée des événements passés [4].

Les processus de Hawkes multivariés (MHP ; 7) permettent de modéliser les interactions de $p \in \mathbb{N}_*$ processus ponctuels temporels auto-excités. Étant donné p ensembles d'événements, $\mathscr{F}_T^i = \{t_n^i,\ t_n^i \in [0, T[\}_{n=1}^{N_T^i}, i = 1, \dots, p$, chaque processus i est décrit par la fonction d'intensité suivante :

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{p} \int_0^t \phi_{ij}(t-s)\, \mathrm{d}N_s^j, \tag{2}$$

où μ_i est le paramètre de l'intensité de base (*baseline intensity*), $N_t = [N_t^1, \dots, N_t^p]$ le processus de comptage multivarié associé et $\phi_{ij} : [0, T] \to \mathbb{R}_+$ les fonctions noyaux, représentant l'influence des événements passés du processus j sur les événements futurs du processus i. Dans le cas d'une classe de noyaux paramétriques paramétrés par η, l'objectif est de trouver les paramètres qui minimisent la fonction de perte des moindres carrés inspirée de l'ERM (Eq. (I.2) in 3, Chapter 1) :

$$\mathcal{L}\left(\theta, \mathscr{F}_T\right) = \frac{1}{N_T} \sum_{i=1}^{p} \left(\int_0^T \lambda_i(s)^2\, \mathrm{d}s - 2 \sum_{t_n^i \in \mathscr{F}_T^i} \lambda_i\left(t_n^i\right) \right), \tag{3}$$

où $N_T = \sum_{i=1}^p N_T^i$ est le nombre total d'événements, et où $\theta \coloneqq (\mu, \eta)$.

2.1 FaDIn

Noyaux à supports finis L'un des principaux obstacles pour l'estimation MLE ou ℓ_2 des noyaux paramétriques est la nécessité de calculer la fonction d'intensité pour tous les événements. Pour les noyaux généraux, la fonction d'intensité nécessite généralement $O((N_T)^2)$ opérations, ce qui la rend inabordable pour les processus de longue durée qui comportent un grand nombre d'événements. Pour rendre ce calcul plus efficace, nous considérons des noyaux à support fini. L'utilisation d'un noyau à support fini revient à fixer une limite dans le temps pour l'influence d'un événement passé sur l'intensité, *i.e.*, $\forall t \notin [0\,; W]\,, \phi_{ij}(t) = 0$, où W désigne la longueur du support du noyau. Cette hypothèse correspond à des applications dans lesquelles un événement ne peut pas avoir d'influence loin dans le futur, comme en neurosciences [1, 5, 10] ou dans le trading haute fréquence [2]. La fonction d'intensité équation 2 peut alors être reformulée comme une convolution entre le noyau ϕ_{ij} et la somme de fonctions de Dirac $z_i \coloneqq \sum_{t_n^i \in \mathscr{F}_T^i} \delta_{t_n^i}$ situées aux occurrences d'événements t_n^i :

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{p} \phi_{ij} * z_j(t), \quad t \in [0\,; T]\,.$$

Comme ϕ_{ij} a un support fini, l'intensité peut être calculée efficacement avec cette formule. En effet, seuls les événements dans l'intervalle $[t - W\,; t]$ doivent être pris en compte.

Discrétisation Pour rendre ces calculs encore plus efficaces, nous proposons de nous appuyer sur des processus discrétisés. La plupart des procédures d'estimation de processus de Hawkes impliquent un paradigme continu pour minimiser (3) ou son équivalent en log-vraisemblance. La discrétisation a été étudiée jusqu'à présent pour les noyaux non paramétriques [8, 9, 11]. La discrétisation d'un TPP consiste à projeter chaque événement t_n^i sur une grille régulière $\mathcal{G} = \{0, \Delta, 2\Delta, \dots, G\Delta\}$, où $G = \left\lfloor \frac{T}{\Delta} \right\rfloor$, avec Δ désignant la taille du pas de discrétisation et $\lfloor \cdot \rfloor$ la fonction partie entière. Soit $\widetilde{\mathscr{F}_T^i}$ l'ensemble des événements projetés de $\mathscr{F}_T^i$ sur la grille $\mathcal{G}$. La fonction d'intensité du i-ième processus de notre MHP discrétisé est alors définie comme suit :

$$\begin{aligned} \tilde{\lambda}_i[s] &= \mu_i + \sum_{j=1}^{p} \sum_{\tilde{t}_m^j \in \widetilde{\mathscr{F}}_{s\Delta}^j} \phi_{ij}(s\Delta - \tilde{t}_m^j) \\ &= \mu_i + \sum_{j=1}^{p} \underbrace{\sum_{\tau=1}^{L} \phi_{ij}^{\Delta}[\tau] z_j[s-\tau]}_{(\phi_{ij}^{\Delta} * z_j)[s]}, \quad s \in [\![0\,; G]\!]\,, \end{aligned} \tag{4}$$

où $L = \left\lfloor \frac{W}{\Delta} \right\rfloor$ désigne le nombre de points sur le support discrétisé, $\phi_{ij}^{\Delta}[s] = \phi_{ij}(s\Delta)$ est la valeur du noyau sur la grille et $z_i[s] = \#\left\{ |t_n^i - s\Delta| \leq \frac{\Delta}{2} \right\}$ désigne le nombre d'événements projetés sur la s-ième « case » de la grille. Dans le reste de l'article, on note $\phi_{ij}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ comme une fonction tandis que $\phi_{ij}^{\Delta}[\cdot]$ représente le vecteur discret $\phi_{ij}^{\Delta} \in \mathbb{R}_+^L$. La discrétisation améliore grandement l'efficacité de calcul, car on peut s'appuyer sur des convolutions discrètes, et, pour les noyaux dont les valeurs sont coûteuses à calculer, au plus L valeurs doivent être calculées. !

FaDIn vise à minimiser la perte ℓ_2 discrétisée, qui approxime l'intégrale de la partie gauche de (3) par une somme sur la grille $\mathcal{G}$ après projection des événements de $\mathscr{F}_T$ sur celle-ci. Cela revient à optimiser, à l'aide d'un algorithme basé sur le gradient du premier ordre, la perte suivante $\mathcal{L}_{\mathcal{G}}\left(\theta, \widetilde{\mathscr{F}_T}\right)$ définie par :

$$\frac{1}{N_T} \sum_{i=1}^{p} \left(\Delta \sum_{s \in [\![0\,; G]\!]} \left(\tilde{\lambda}_i[s]\right)^2 - 2 \sum_{\tilde{t}_n^i \in \widetilde{\mathscr{F}}_T^i} \tilde{\lambda}_i\left[\frac{\tilde{t}_n^i}{\Delta}\right] \right). \tag{5}$$

En développant cette équation, on peut faire apparaitre des termes qui apparaissent également dans les équations du gradient, et qui ne dépendent par de θ. Ainsi, ces termes peuvent être pré-calculés afin de réduire la complexité du calcul de la perte ℓ_2, alors égale à $O(p^3 L^2)$ et indépendante du nombre total d'événements N_T.

3 Expériences numériques sur l'impact de la discrétisation

Bien que la discrétisation permette des calculs efficaces, elle introduit également une perturbation dans la valeur de la perte. On cherche alors à quantifier l'impact de cette perturbation sur l'estimation des paramètres lorsque $\Delta \to 0$. Dans cette section, on observe un processus $\mathscr{F}_T$ dont la fonction d'intensité est donnée par la forme paramétrique $\lambda(\cdot\,; \theta^*)$. Si l'intensité du processus $\mathscr{F}_T$ n'est pas dans la famille paramétrique $\lambda(\cdot\,; \theta)$, θ^* est défini comme la meilleure approximation de sa fonction d'intensité dans le sens de ℓ_2. L'objectif est alors d'estimer les paramètres θ^*.

Lorsque l'on travaille avec le processus discret $\widetilde{\mathscr{F}_T}$, les événements t_n^i du processus original sont remplacés par leur projection $\tilde{t}_n^i$ sur la grille $\mathcal{G}$: $\tilde{t}_n^i \coloneqq t_n^i + \delta_n^i$. Ici, δ_n^i est uniformément distribué sur $[-\Delta/2, \Delta/2]$. On considère l'estimateur discret FaDIn $\widehat{\theta}_\Delta$ défini comme $\widehat{\theta}_\Delta \coloneqq \arg\min_\theta \mathcal{L}_{\mathcal{G}}(\theta)$. On

peut alors majorer l'erreur commise par $\widehat{\theta}_\Delta$ grâce à la décomposition suivante :

$$\left\|\widehat{\theta}_\Delta - \theta^*\right\|_2 \leq \underbrace{\left\|\widehat{\theta}_c - \theta^*\right\|_2}_{(*)} + \underbrace{\left\|\widehat{\theta}_\Delta - \widehat{\theta}_c\right\|_2}_{(**)}, \tag{6}$$

où $\widehat{\theta}_c \coloneqq \arg\min_\theta \mathcal{L}(\theta)$ est l'estimateur de référence pour θ^* basé sur l'estimateur ℓ_2 standard pour les processus ponctuels continus. Cette décomposition implique une erreur statistique $(*)$ et un biais $(**)$ induit par la discrétisation. Le terme statistique mesure à quel point les paramètres obtenus en minimisant la perte continue ℓ_2 avec un accès fini aux données sont éloignés des vrais paramètres. En revanche, le terme $(**)$ représente le biais de discrétisation induit par la minimisation de la perte discrète équation 5 plutôt que de la perte continue (équation 3).

Il peut être mathématiquement démontré que lorsque le pas de discrétisation Δ tend vers 0, l'intensité perturbée et la perte ℓ_2 sont de bonnes estimations de leurs homologues continus, mais aussi que l'estimateur $\widehat{\theta}_\Delta$ est équivalent à $\widehat{\theta}_c$.

Afin d'étudier le biais d'estimation dû à la discrétisation, nous avons mené deux expériences et rapporté les résultats dans la figure 1. Le paradigme général est un TPP à une dimension avec une intensité paramétrisée comme dans l'équation 2 avec un noyau gaussien tronqué de moyenne $m \in \mathbb{R}$ et d'écart-type $\sigma > 0$, avec un support fixe $[0\,;W] \subset \mathbb{R}_+$, $W > 0$. Il correspond à $\phi(\cdot) = \alpha\kappa(\cdot), \alpha \geq 0$ avec

$$\kappa(\cdot) \coloneqq \kappa(\cdot; m, \sigma, W) = \frac{1}{\sigma} \frac{f\left(\frac{\cdot - m}{\sigma}\right)}{F\left(\frac{W-m}{\sigma}\right) - F\left(\frac{-m}{\sigma}\right)} \mathbb{1}_{\{0 \leq \cdot \leq W\}},$$

où f (resp. F) est la fonction de densité de probabilité (resp. la fonction de répartition) de la distribution normale standard. Ainsi, les paramètres à estimer sont μ et $\eta = (\alpha, m, \sigma)$.

Dans les deux expériences, pour plusieurs longueurs T de processus, les estimations discrètes sont calculées pour des pas de grille variables, Δ allant de 10^{-1} à 10^{-3}. Le paramètre W est fixé à 1. La norme ℓ_2 de la différence entre les estimations et les vraies valeurs de paramètres – celles utilisées pour la simulation des données – est calculée et rapportée. Nous avons d'abord calculé les estimations de paramètres avec notre méthode FaDIn pour $T \in \{10^3, 10^5, 10^4, 10^6\}$, pour 100 simulations à chaque fois. Deuxièmement, afin de séparer le biais de discrétisation du biais statistique, nous avons calculé les estimations avec un algorithme d'espérance-maximisation (EM), à la fois de manière continue et discrète, et ce, pour 50 simulations de données aléatoires. En effet, la forme particulière du noyau choisi permet d'avoir une forme close de la fonction de perte, et donc, à l'aide d'un algorithme basé sur l'EM, on peut effectuer l'estimation des paramètres dans un cadre continu, on obtient alors $\widehat{\theta}_c$ [1].

On peut observer que les erreurs ℓ_2 entre les estimations discrètes et les vrais paramètres tendent vers zéro lorsque T augmente. Pour une valeur fixe de T, on peut observer des plateaux qui commencent pour des valeurs de taille de pas qui ne sont pas particulièrement petites, ce qui indique que le biais de discrétisation est limité. La deuxième expérience avec l'algorithme EM montre que lorsque le plateau est atteint, cela correspond à une certaine erreur statistique. En d'autres termes, même pour une taille de pas raisonnablement grossière, le biais induit par la discrétisation est faible par rapport à l'erreur statistique.

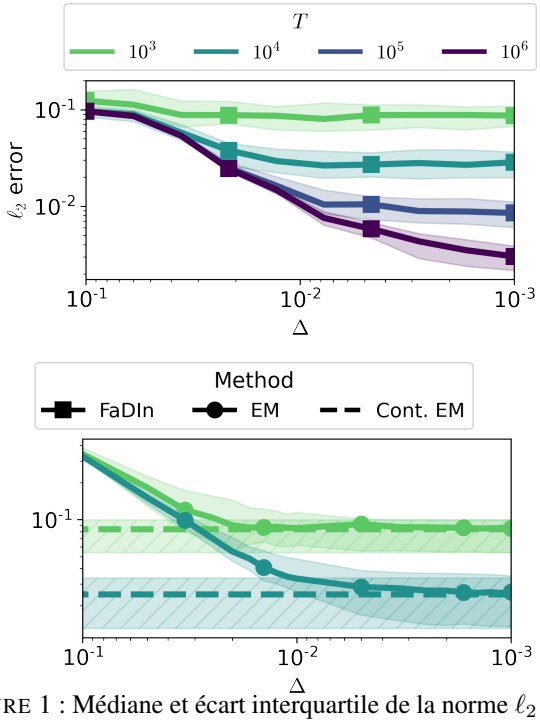

FIGURE 1 : Médiane et écart interquartile de la norme ℓ_2 entre les vrais paramètres et les estimations de paramètres calculées avec FaDIn (en haut) et avec l'algorithme EM (en bas), de manière continue et discrète, en fonction du pas de la grille Δ.

D'autres expériences qui comparent FaDIn avec différentes méthodes paramétriques et non paramétriques ont permis de montrer que FaDIn a l'avantage sur les autres méthodes, lorsque T varie, tant d'un point de vue statistique (l'erreur d'estimation est plus faible) que computationnel (le temps de calcul est bien plus faible grâce aux pré-calculs).

4 Application à la MEG

Les expériences sur les données MEG ont été réalisées sur deux jeux de données provenant du package `Python MNE` [6] : *sample* obtenu lors de stimulations audio-visuelles et *somato* obtenu lors de stimulation du nerf médian[1]. Ces jeux de données ont été sélectionnés car ils présentent deux types distincts d'activations neuronales liées à des événements : des réponses évoquées dont la latence est quasi-fixe et des réponses induites qui présentent des latences plus variables.

Pareillement à [1], pour chaque jeu de données, un prétraitement des données brutes est effectué, similaire, puis un modèle de *Convolutional dictionary learning* (CDL) – méthode non supervisée permettant d'extraire des motifs (atomes) récurrents et invariants au temps dans des séries temporelles – est appliqué : 40 atomes de durée 1 s chacun sont extraits pour *sample*, et 20 atomes de durée 0.53 s pour *somato*. De cette façon, chaque ensemble de données est représenté par deux ensembles de processus ponctuels temporels : un ensemble de processus stochastiques représentant les activations des atomes, et un ensemble de processus déterministes codant pour les événements de stimuli externes.

L'objectif principal d'utiliser des TPP à de telles données est de caractériser directement quand et comment chaque stimu-

[1]Tous deux disponibles sur https://mne.tools/stable/overview/datasets_index.html

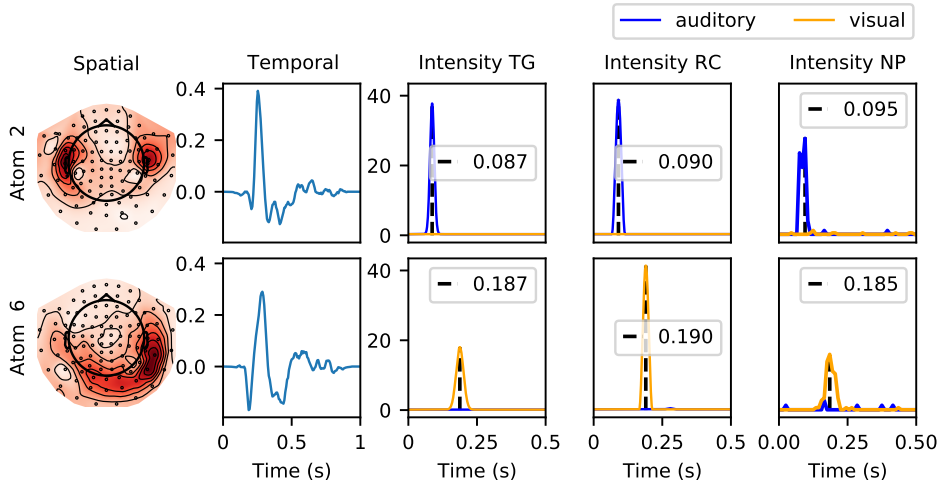

FIGURE 2 : Représentations spatiales et temporelles de 4 atomes extraits de MNE sample, et leurs fonctions d'intensité estimées respectives après un stimulus auditif ou visuel au temps t = 0 s, avec des noyaux non-paramétrique (NP), gaussienne tronquée (TG) et cosinus surélevé (RC).

lus est responsable de l'apparition de réponses neuronales, en particulier en estimant la distribution des latences. Nous nous intéressons au paradigme de *Driven Point Process* (DriPP ; 1) et, pour chaque atome extrait, sa fonction d'intensité liée aux stimuli correspondants est estimée de trois façons différentes : à l'aide d'un noyau non paramétrique (NP) et de deux paramétrisations de noyau : Truncated Gaussian (TG) et Raised Cosine (RC).

Les résultats montrent que les trois noyaux sont d'accord sur un pic de latence autour de 90 ms pour la condition auditive et de 190 ms pour la condition visuelle. En raison du nombre limité d'événements, on peut observer que le noyau non paramétrique estimé est moins lisse, avec des pics parasites plus tard dans l'intervalle. Dans l'ensemble, ces résultats sur des données MEG réelles démontrent que notre approche avec une paramétrisation de noyau RC peut récupérer des estimations de latence correctes même dans le cadre d'une discrétisation avec un pas de 0,02. De plus, l'utilisation de RC nous permet d'avoir des pics plus nets en intensité par rapport à TG, renforçant le lien entre le stimulus externe et l'activation de l'atome. Cette différence vient principalement du fait qu'à la différence du noyau TG, le noyau RC n'a pas besoin de support prédéterminé, puisqu'il dépend de ses paramètres de moyenne et de variance. Cet avantage est encore plus prononcé dans le cas de réponses induites, comme dans l'ensemble de données *somato* (figure non présentée ici), où la plage de valeurs de latence possibles est plus difficile à déterminer à l'avance.

5 Conclusion

FaDIn est une approche efficace pour inférer des noyaux paramétriques généraux pour les processus de Hawkes multivariés, notamment à l'aide de l'hypothèse de support fini et grâce à de possibles pré-calculs permis par la discrétisation. Ces caractéristiques permettent une approche basée sur les gradients efficace d'un point de vue computationnel, améliorant les méthodes de l'état de l'art tout en offrant une utilisation flexible de noyaux bien adaptés aux applications considérées. De plus, ce travail montre que le biais induit par la discrétisation est négligeable, tant sur le plan théorique que numérique. En permettant l'utilisation d'un noyau paramétrique général dans les processus de Hawkes, cette contribution ouvre de nouvelles possibilités pour de nombreuses applications. C'est le cas avec les données MEG, où l'estimation d'informations sur la fréquence et la latence des occurrences de motifs de signaux cérébraux est au cœur des questions en neurosciences.

Références

[1] Cédric ALLAIN, Alexandre GRAMFORT et Thomas MOREAU : DriPP : Driven point processes to model stimuli induced patterns in M/EEG signals. *In International Conference on Learning Representations*, 2021.

[2] Emmanuel BACRY, Iacopo MASTROMATTEO et Jean-François MUZY : Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[3] Martin BOMPAIRE : *Machine learning based on Hawkes processes and stochastic optimization*. Thèse de doctorat, Université Paris-Saclay, 2019.

[4] Daryl J. DALEY et David VERE-JONES : *An introduction to the theory of point processes. Volume I : Elementary theory and methods*. Probability and Its Applications. Springer-Verlag New York, 2003.

[5] Michael EICHLER, Rainer DAHLHAUS et Johannes DUECK : Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 38(2):225–242, 2017.

[6] Alexandre GRAMFORT, Martin LUESSI, Eric LARSON, Denis A. ENGEMANN, Daniel STROHMEIER, Christian BRODBECK, Lauri PARKKONEN et Matti S. HÄMÄLÄINEN : MNE software for processing MEG and EEG data. *Neuroimage*, 86:446–460, 2014.

[7] Alan G. HAWKES : Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society : Series B (Methodological)*, 33(3):438–443, 1971.

[8] Matthias KIRCHNER : Hawkes and INAR(∞) processes. *Stochastic Processes and their Applications*, 126(8):2494–2525, août 2016.

[9] Matthias KIRCHNER et A BERCHER : A nonparametric estimation procedure for the hawkes process : comparison with maximum likelihood estimation. *Journal of Statistical Computation and Simulation*, 88(6):1106–1116, 2018.

[10] Michael KRUMIN, Inna REUTSKY et Shy SHOHAM : Correlation-based analysis and generation of multiple spike trains using hawkes models with an exogenous input. *Frontiers in computational neuroscience*, 4:147, 2010.

[11] Daisuke KURISU : Discretization of self-exciting peaks over threshold models. 2016.

[12] Patricia REYNAUD-BOURET et Vincent RIVOIRARD : Near optimal thresholding estimation of a poisson intensity on the real line. *Electronic journal of statistics*, 4:172–238, 2010.

[13] Hongteng XU, Dixin LUO et Hongyuan ZHA : Learning hawkes processes from short doubly-censored event sequences. *In International Conference on Machine Learning*, pages 3831–3840. PMLR, 2017.