Étude des processus de Hawkes non-linéaires et applications en neurosciences

Offre de stage de M2 avec possibilité de thèse

LPSM – Sorbonne Université

1 Sujet d'étude

Le sujet d'étude porte sur des événements temporels dont la probabilité d'apparition dépend des observations passées. Un modèle couramment utilisé pour ce type de phénomènes est le processus de Hawkes, dont les domaines d'application sont très variés et incluent la génétique, la sismologie et les neurosciences. Nous nous intéressons ici à différentes façons de modéliser la dépendance au passé dans un processus de type Hawkes et cherchons à analyser et à comparer ces approches.

Le travail proposé dans ce stage vise donc à étudier les modélisations présentées ci-dessous en se concentrant sur les aspects théoriques et numériques des méthodes d'estimation, ainsi que sur leur mise en pratique sur des jeux de données issus des neurosciences.

Processus ponctuel Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On considère un processus ponctuel multivarié $N = (N^1, \dots, N^d)$ (d étant un entier non-nul fixé) où, pour tout $i \in [\![1, d]\!]$, $N^i : \mathcal{F} \to \mathbb{N}$ est un processus univarié associé à une suite croissante de temps $(T_k^i)_{k\geq 1}$:

$$\forall A \in \mathcal{F}, \quad N^i(A) = \sum_{k\geq 1} \mathbf{1}_{T_k^i \in A}.$$

Autrement dit, $N^i(A)$ compte le nombre d'événements T_k^i tombant dans l'ensemble A (conférant parfois à N^i le nom de « mesure de comptage »).

Modèle de Hawkes multivarié Un processus de Hawkes est un processus ponctuel pour lequel on modélise explicitement l'apparition des événements en fonction du passé. Pour ce faire, on caractérise la dépendance des temps d'événements $\left\{T_k^i : k \geq 1, i \in [\![1, d]\!]\right\}$ en spécifiant la loi du processus N à travers les intensités conditionnelles définies, pour tout $i \in [\![1, d]\!]$, par :

$$\forall t \in \mathbb{R}_+, \quad \lambda^i\left(t \mid \mathcal{H}_t\right) = \lim_{h\to 0} \frac{\mathbb{P}(N^i(]t, t+h]) > 0 \mid \mathcal{H}_t)}{h},$$

où $\mathcal{H}_t$ est l'histoire du processus N jusqu'au temps t. Intuitivement, $\lambda^i\left(t \mid \mathcal{H}_t\right)$ représente la probabilité instantanée qu'un événement de type i apparaisse à l'instant t.

Pour un processus de Hawkes [Hawkes, 1971], chaque intensité conditionnelle est modélisée par :

$$\forall t \in \mathbb{R}_+, \quad \lambda^i\left(t \mid \mathcal{H}_t\right) = \mu_i + \sum_{j=1}^{d} \sum_{k\geq 1 : T_k^i \leq t} \varphi_{ij}(t - T_k^i).$$

L'apparition d'événements est donc contrôlée par un mécanisme de fond, paramétré par le terme constant $\mu_i > 0$ (appelé intensité de base), et par une dépendance au passé, représentée par le noyau $\varphi_{ij} : \mathbb{R} \to \mathbb{R}_+$, appelée alternativement fonction d'interaction ou fonction de mémoire. Dans ce contexte, φ_{ii} capture l'auto-excitation du sous-processus N^i sur lui-même tandis que φ_{ij} $(j \neq i)$ encode l'excitation mutuelle du sous-processus N^j sur le sous-processus N^i.

Modèle de Hawkes non-linéaire Le modèle excitant présenté ci-dessus est bien adapté à certaines applications, telles que l'analyse d'événements sismiques. Il est cependant trop restrictif pour modéliser des interactions neuronales, pour lesquelles on observe fréquemment des effets d'inhibition [Reynaud-Bouret et al., 2014]. Afin d'intégrer ces effets inhibiteurs dans la modélisation, on considère donc des processus de Hawkes non-linéaires :

$$\forall t \in \mathbb{R}_+, \quad \lambda^i\left(t \mid \mathcal{H}_t\right) = \max\left(0, \mu_i + \sum_{j=1}^{d} \sum_{k \geq 1 : T_k^i \leq t} \varphi_{ij}(t - T_k^i)\right),$$

pour lesquels $\varphi_{ij} : \mathbb{R} \to \mathbb{R}$ est une fonction à valeurs potentiellement négatives, modélisant ainsi l'auto ou l'inter-inhibition.

L'existence et la stabilité de tels processus ont été étudiées [Brémaud and Massoulié, 1996]. On compte de plus quelques méthodes d'inférence bayésiennes [Sulem et al., 2021, 2022, Deutsch and Ross, 2022] et fréquentistes [Bonnet et al., 2023]. Cependant les propriétés asymptotiques de ces estimateurs sont peu connues, en particulier dans le cas fréquentiste.

Ce point pourra faire l'objet d'une première partie du travail de stage/thèse.

Mémoire de longueur variable Dans la modélisation précédente, s'il est possible de considérer que l'intensité ne dépend que d'événements « récents » (en utilisant des fonctions de mémoire φ_{ij} à support borné), plusieurs travaux suggèrent qu'une modélisation avec une mémoire de longueur variable (i.e. dépendant des événements passés, donc aléatoire) est plus adaptée pour décrire l'activité neuronale [Galves and Löcherbach, 2013, Duarte et al., 2019, De Santis et al., 2022]. Plus précisément, dans ce modèle, l'activation (ou « spike ») d'un neurone ne dépend que de l'activité des autres neurones depuis son dernier « spike ». Les travaux de Galves and Löcherbach [2013] proposent notamment une nouvelle modélisation d'un système de particules en interaction qui intègre cette hypothèse de mémoire variable, qui a ensuite donné lieu à des méthodes d'estimation du graphe d'interaction entre les neurones [Duarte et al., 2019, De Santis et al., 2022]. Il est cependant difficile de comparer ces deux approches (mémoire fixe ou mémoire à longueur variable) car les travaux existants font le choix initial de se placer dans l'un ou l'autre de ces cadres.

Nous proposons ici une nouvelle modélisation décomposant la fonction d'intensité λ^i en deux termes : l'un décrivant l'influence du passé récent (i.e. depuis le dernier événement de type i), l'autre du passé plus ancien. Formellement, en notant $N^i(t) = N^i([0,t])$ le nombre d'événements de type i se produisant avant t, on suppose que :

$$\forall t \in \mathbb{R}_+, \qquad \lambda^i\left(t \mid \mathcal{H}_t\right) = \max\left(0, \mu_i + \sum_{j=1}^{d} \left[\sum_{\substack{k \geq 1 : \\ T_k^j \in]T^i_{N^i(t)}, t]}} \varphi_{ij}(t - T_k^j) + \sum_{\substack{k \geq 1 : \\ T_k^j \in [0, T^i_{N^i(t)}]}} \widetilde{\varphi}_{ij}(t - T_k^j) \right]\right),$$

pour certains paramètres μ_i et fonctions $\varphi_{ij} : \mathbb{R} \to \mathbb{R}$ et $\widetilde{\varphi}_{ij} : \mathbb{R} \to \mathbb{R}$. Cette modélisation a l'intérêt d'unifier trois situations d'intérêt en neurosciences puisque le processus ainsi défini est :

— un processus de Hawkes non-linéaire usuel lorsque $\varphi_{ij} = \widetilde{\varphi}_{ij}$ pour tout couple (i,j) ;

— un processus de Hawkes non-linéaire à mémoire fixe lorsque $\varphi_{ij} = \widetilde{\varphi}_{ij}$ et φ_{ij} est à support borné pour tout couple (i,j) ;

— un processus de Hawkes non-linéaire à mémoire de longueur variable lorsque $\widetilde{\varphi}_{ij} = 0$ pour tout couple (i,j).

L'étude de ce nouveau modèle pourra faire l'objet d'une deuxième partie du travail de stage/thèse.

2 Objectifs

L'objectif de ce travail est d'avancer dans l'une ou les deux directions décrites précédemment en étudiant les propriétés des méthodes d'estimation fréquentistes des processus des Hawkes multivariés non-linéaires usuels ou à mémoire de longueur variable. On cherchera en particulier à établir :

— l'identifiabilité des modèles statistiques proposés (en partie étudiée dans [Sulem et al., 2022, Bonnet et al., 2023], uniquement pour le modèle usuel ou à mémoire fixe) ;

— l'estimation des fonctions d'interaction dans le modèle non-linéaire (connue uniquement dans le cas usuel [Bonnet et al., 2023])

— la convergence et la loi asymptotique des estimateurs (connues pour les processus de Hawkes usuels et uniquement excitants [Ogata, 1978, Guo et al., 2018]) ;

— l'intérêt d'une mémoire à longueur variable vis-à-vis du modèle de Hawkes usuel, d'un point de vue théorique, numérique et sur des données neuronales. Il s'agit notamment de mettre en place des tests statistiques pour les hypothèses nulles que l'un ou les sous-processus ont une mémoire de longueur variable (i.e. $\widetilde{\varphi}_{ij} = 0$).

3 Profil souhaité

Le ou la candidate sélectionné·e sera en cours d'obtention (ou titulaire) d'un M2 de mathématiques appliquées, ou en dernière année d'école d'ingénieur·e·s, option mathématiques appliquées. Il ou elle disposera d'un socle fondamental en probabilités, statistique ou apprentissage statistique et devra manipuler avec aisance un langage de calcul scientifique tel que R ou Python. En fonction du profil et des souhaits du candidat ou de la candidate, les aspects théoriques, numériques et/ou applicatifs pourront être privilégiés.

4 Environnement de travail

Le stage se déroulera à partir d'avril/mai 2024, pour une durée de 5 à 6 mois, au Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR 8001) de Sorbonne Université. Il sera encadré par Anna Bonnet et Maxime Sangnier, et gratifié à hauteur de $4,05$€/h, soit environ 560€/mois.

5 Candidature

Le dossier de candidature est constitué :

— d'un curriculum vitæ ;

— d'une lettre de motivation ;

— des relevés de notes de niveau L3, M1 et M2 (correspondant à la première, deuxième et troisième année d'école pour les étudiant·e·s ingénieur·e·s) dans la mesure des disponibilités.

Il est à envoyer par mail à Anna Bonnet et à Maxime Sangnier.

Références

A. Bonnet, M. Martinez Herrera, and M. Sangnier. Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity. *Statistics and Computing*, 33(91) : 1573–1375, 2023. doi : 10.1007/s11222-023-10264-w.

P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24 (3) :1563–1588, 1996. doi : 10.1214/aop/1065725193.

E. De Santis, A. Galves, G. Nappo, and M. Piccioni. Estimating the interaction graph of stochastic neuronal dynamics by observing only pairs of neurons. *Stochastic Processes and their Applications*, 149 :224–247, 2022. ISSN 0304-4149. doi : https://doi.org/10.1016/j.spa.2022.03.016. URL https://www.sciencedirect.com/science/article/pii/S0304414922000801.

I. Deutsch and G.J. Ross. Bayesian estimation of multivariate Hawkes processes with inhibition and sparsity, 2022. Preprint at https://arxiv.org/abs/2201.05009.

Aline Duarte, Antonio Galves, Eva Löcherbach, and Guilherme Ost. Estimating the interaction graph of stochastic neural dynamics. *Bernoulli*, 25(1), February 2019. ISSN 1350-7265. doi : 10.3150/17-bej1006. URL http://dx.doi.org/10.3150/17-BEJ1006.

A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *Journal of Statistical Physics*, 151(5) :896–921, March 2013. ISSN 1572-9613. doi : 10.1007/s10955-013-0733-9. URL http://dx.doi.org/10.1007/s10955-013-0733-9.

X. Guo, A. Hu, R. Xu, and J. Zhang. Consistency and computation of regularized mles for multivariate Hawkes processes, 2018. Preprint at https://arxiv.org/abs/1810.02955.

A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1) : 83–90, 1971. doi : 10.2307/2334319.

Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30 :243–261, 1978. doi : 10.1007/BF02480216.

P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and non-parametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4(3) :3, 2014. doi : 10.1186/2190-8567-4-3.

D. Sulem, V. Rivoirard, and J. Rousseau. Bayesian estimation of nonlinear Hawkes process, 2021. Preprint at https://arxiv.org/abs/2103.17164.

D. Sulem, V. Rivoirard, and J. Rousseau. Scalable variational bayes methods for Hawkes process. *arXiv :2212.00293*, 2022.