

Introduction au Domaine de Recherche

Processus de Hawkes et applications aux Neurosciences

Théo Leblanc

May 21, 2024

Le but de ce document est de présenter une classe de processus particulier : les processus des Hawkes, et d'en discuter les différentes applications possibles, notamment en neurosciences. L'approche de ce document reste majoritairement mathématiques, les applications et connections avec les neurosciences sont évoquées sans rentrer dans trop de détails.

1 Préliminaires

1.1 Mesures aléatoires de Poisson

Comment tirer *aléatoirement* un point de \mathbb{R}^n ? Puisque \mathbb{R}^n est de cardinal infini, il n'existe pas de mesure de probabilité uniforme sur cet ensemble, il faut donc ruser pour obtenir un tirage uniforme sur \mathbb{R}^n . L'astuce c'est non pas de ne tirer qu'un seul point mais d'en tirer un nombre infini dénombrable. Les objets mathématiques permettant cela sont les mesures aléatoires de Poisson.

Une mesure aléatoire de Poisson π sur \mathbb{R}^n et d'intensité Leb, la mesure de Lebesgue, est une mesure aléatoire et atomique (somme de masses de dirac en des points deux à deux distincts) telle que

- Pour tout boréliens $B \subset \mathbb{R}^n$, $\pi(B)$ est mesurable et suit une loi $\mathcal{Pois}(\text{Leb}(B))$,
- Pour tout $k \in \mathbb{N}^*$, pour tout boréliens B_1, \dots, B_k deux à deux disjoints, les variables aléatoires $\pi(B_1), \dots, \pi(B_k)$ sont indépendantes.

1.2 Processus ponctuels sur \mathbb{R}

Un processus ponctuel univarié dans \mathbb{R} (par la suite, simplement processus ponctuel) est un ensemble aléatoire dénombrable de points $S \subset \mathbb{R}$ tel que pour tout borélien B , la quantité $S(B) := \text{Card}(S \cap B)$ est mesurable. Les processus ponctuels permettent de modéliser les instants aléatoires où des événements booléens se produisent. Ces événements booléens peuvent par exemple être chaque nouveau cas d'une maladie en épidémiologie, le déclenchement d'un nouveau tremblement de terre ou bien comme il sera question dans la suite de ce document, l'émission d'un signal électrique par un neurone (du cerveau).

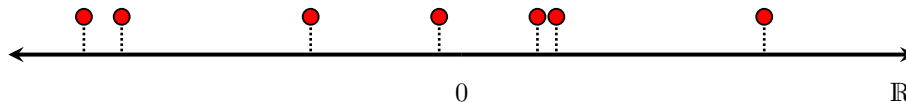


Figure 1: Processus ponctuel univarié.

Étant donné un ensemble fini \mathcal{M} , un processus ponctuel multivarié est un \mathcal{M} -couple de processus ponctuels univariés $(S^m)_{m \in \mathcal{M}}$ tel que les S^m sont disjoints (cette condition est souvent vérifiée en

pratique et c'est aussi le cadre utilisé dans [1]). L'ensemble \mathbf{M} est un ensemble de types, par exemple différentes maladies pour reprendre le premier exemple, les continents sur lesquels interviennent les tremblements de terres du deuxième exemple et les différents neurones du cerveau dans le troisième exemple.

2 Processus de Hawkes

Les processus de Hawkes, introduits par Alan Hawkes en 1971 [8], sont des processus ponctuels avec une structure de dépendance particulière : l'arrivée ou non au temps $t \in \mathbb{R}$ d'un évènement dépend du passé avant t . Avant de plonger dans l'étude des processus de Hawkes, il n'est pas inutile de faire une petite digression par les processus ponctuel sans structure de dépendance : les processus ponctuels de Poisson.

Les processus ponctuel de Poisson (PPP) sont les processus ponctuels les plus simples et les plus naturels. Ils sont l'équivalent des suites iid pour les suites de variables aléatoires. Ils ne diffèrent des mesures aléatoires de Poisson sur \mathbb{R} que par la forme (ensemble aléatoire et non mesure aléatoire) comme le montre le point 1 du résultat suivant. Le point 2 fait le lien avec les suites iid de variables aléatoires.

Proposition 1. *Soit S un processus ponctuel et $\lambda > 0$. Les assertions suivantes sont équivalentes.*

1. *Pour tout $n \geq 1$ et pour tout borélien B_1, \dots, B_n disjoints, les variables aléatoires $S(B_k)$ sont indépendantes et distribuées selon une loi $\mathcal{Poi}(\lambda|B_k|)$ où $|B_k|$ signifie la mesure de Lebesgue de l'ensemble B_k .*
2. *$S = \{\dots, T_{-1}, T_0, T_1, \dots\}$ avec $T_0 = \min\{t \in S \mid t \geq 0\}$ et $\dots < T_{-1} < T_0 < T_1 < \dots$ et les variables aléatoires $T_0, -T_{-1}$ et $I_k = T_k - T_{k-1}$ pour $k \in \mathbb{Z}^*$ sont iid de loi $\mathcal{Exp}(\lambda)$.*

Un tel processus est appelé processus ponctuel de Poisson d'intensité λ .

La terminologie "intensité λ " provient du fait que la probabilité infinitésimale qu'un point tombe dans un intervalle $[t, t + dt]$ est λdt .

Il existe une méthode à la fois très élégante et très utile en pratique pour réaliser des PPP. Cette méthode, dite de *thinning*, voir fig. 2, consiste à réaliser un PPP S d'intensité λ en ne conservant que les points d'une mesure aléatoire de Poisson sur \mathbb{R}^2 dont les ordonnées sont comprises entre 0 et λ . La proposition suivante formalise ce résultat.

Proposition 2. *Soit π une mesure aléatoire de Poisson sur \mathbb{R}^2 d'intensité $dt \times dx$ et une constante $\lambda > 0$. Alors le processus ponctuel définit par*

$$S = \{t \in \mathbb{R} \mid \exists x \in [0, \lambda] \text{ tel que } (t, x) \in \pi\}$$

est un processus ponctuel de Poisson d'intensité λ , voir fig. 2.

Cette représentation via une mesure aléatoire de Poisson ouvre la porte à des processus ponctuels plus généraux que les processus ponctuels de Poisson. Il est en effet naturel de libérer la contrainte sur le paramètre λ en l'autorisant à fluctuer, de manière déterministe ou stochastique. Quand l'intensité constante λ d'un processus ponctuel de Poisson est remplacée par une fonction déterministe $t \mapsto \lambda_t$ on parle de processus de Poisson inhomogène.

Les processus de Hawkes rentrent dans le formalisme où l'intensité λ est stochastique, et plus précisément *prévisible* c'est-à-dire qu'elle ne dépend que du passé : λ_t est une fonction mesurable de $\pi|_{(-\infty, t) \times \mathbb{R}}$ dans le cas de la méthode de thinning. Quantitativement, pour un processus de Hawkes linéaire univarié,

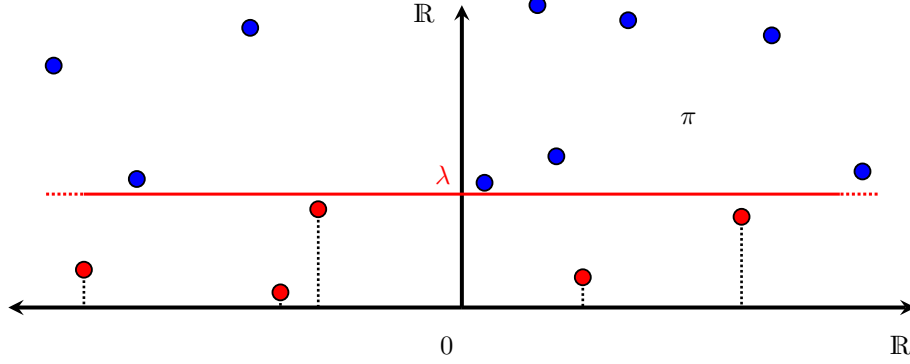


Figure 2: Méthode de thinning.

l'intensité stochastique s'écrit comme suit :

$$\lambda_t = \mu + \int_{-\infty}^{t-} h(t-s) dS_s = \mu + \sum_{T \in S, T < t} h(t-T) \quad (1)$$

où $\mu \geq 0$ est une constante et $h : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$ est une fonction positive. La constante μ est un taux spontané d'apparition de points, et chaque point $s \in S$ augmente le taux d'arrivée d'un nouveau point au temps $t > s$ de la quantité $h(t-s)$, la fonction h étant nommée fonction d'interaction.

Rien ne nous limite au cas linéaire et au cas univarié, ce qui donne la définition suivante pour un processus de Hawkes de le cas général.

Définition 1 (Processus de Hawkes). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et une filtration $(\mathcal{F}_t)_{t \in \mathbb{R}}$. Soit \mathbf{M} un ensemble fini, $\mathbf{h} = (h_{m'}^m)_{m, m' \in \mathbf{M}}$ des fonctions réelles et $\Phi = (\Phi^m)_{m \in \mathbf{M}}$ des fonctions réelles positives. Un processus de Hawkes de paramètres $\mathbf{M}, \mathbf{h}, \Phi$ est un processus ponctuel multivarié $\mathbf{S} := (S^m)_{m \in \mathbf{M}}$ tel que*

- \mathbf{S} est adapté à $(\mathcal{F}_t)_{t \in \mathbb{R}}$, ie $\mathbf{S}|_{(-\infty, t]}$ est mesurable pour \mathcal{F}_t ,
- presque sûrement, pour $m \neq m'$ dans \mathbf{M} , S^m n'a pas de points communs avec $S^{m'}$,
- \mathbf{S} a pour intensité stochastique $\lambda := (\lambda^m)_{m \in \mathbf{M}}$ donnée par

$$\lambda_t^m = \Phi^m \left(\sum_{m' \in \mathbf{M}} \int_{-\infty}^{t-} h_{m'}^m(t-s) dS_s^{m'} \right) \quad m \in \mathbf{M}, t \in \mathbb{R}. \quad (2)$$

Les Φ^m sont appelées les *fonctions de lien*, elles sont généralement croissantes, cela permet de relier le signe des *fonctions d'interactions* $h_{m'}^m$, à leur caractère excitateur ou inhibiteur. On retrouve le cas linéaire en prenant des fonctions d'interaction positives et $\Phi^m = \mu_m + \text{id}$.

L'existence d'un tel processus n'est pas clair a priori, le concept d'intensité stochastique devrait aussi être précisé. Intuitivement, l'intensité stochastique donne le taux d'apparition d'un point au temps t sachant le passé,

$$\mathbb{P}(S^m \text{ a 1 point dans } [t, t+dt] \mid \mathcal{F}_t) = \lambda_t^m dt, \quad m \in \mathbf{M}, t \in \mathbb{R}.$$

La *bonne* définition de l'intensité est fonctionnelle [1] : étant donnée une filtration, le processus ponctuel

adapté S a pour intensité stochastique prévisible λ si pour tout processus prévisible positif X on a

$$\mathbb{E} \left[\int_{\mathbb{R}} X_t^m dS_t^m \right] = \mathbb{E} \left[\int_{\mathbb{R}} X_t^m \lambda_t^m dt \right] \quad m \in \mathbf{M}. \quad (3)$$

La méthode de thinning contourne les difficultés théoriques de définition de l'intensité stochastique. En effet, un processus ponctuel S défini comme les points d'une mesure aléatoire de Poisson π sous les trajectoires du processus prévisible λ (voir fig. 3), où la filtration est donnée par $\mathcal{F}_t = \sigma(\pi|_{(-\infty, t) \times \mathbb{R}})$, vérifie (3) et admet donc λ comme intensité prévisible. Cela provient du fait que " $d\pi - dxdt$ " est une martingale.

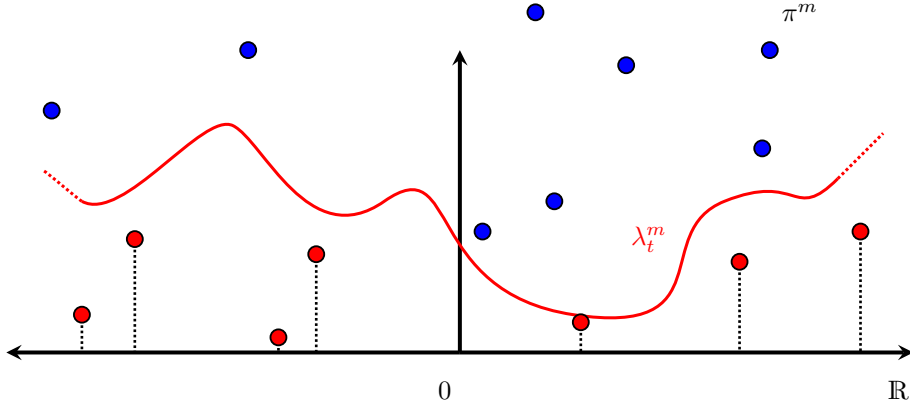


Figure 3: Méthode de thinning pour le processus de Hawkes. Chaque S^m est construit dans une mesure aléatoire de Poisson π^m (qui sont toutes indépendantes) en ne conservant que les points de π^m sous l'intensité λ^m . L'intensité ne dépendant que du passé, si le processus nous est donné jusqu'au temps t on peut poursuivre sa construction jusqu'à $+\infty$ (s'il n'y a pas d'explosion en temps fini).

Brémaud et Massoulié ont prouvé dans [2] le résultat suivant pour l'existence des processus de Hawkes sur \mathbb{R} tout entier.

Hypothèse 1. *Le rayon spectral de la matrice $\mathbf{H} = (L_m \|h_{m'}^m\|_1)_{m, m' \in \mathbf{M}}$, où $L_m := \text{Lip}(\Phi^m)$ est le coefficient de lipschitz de Φ^m , est strictement plus petit que 1, ie $\text{RSp}(\mathbf{H}) < 1$.*

Théorème 1 (Brémaud et Massoulié 1996). *Sous l'Hypothèse 1, il existe un unique (en loi) processus de Hawkes au sens de Définition 1 qui a des moments d'ordre 1 finis (ie $\sup_{m \in \mathbf{M}, t \in \mathbb{R}} \mathbb{E}[\lambda_t^m] < \infty$). De plus, ce processus est stationnaire.*

Idée de preuve. La preuve est basée sur une itération de Picard. Commençons avec des PPP d'intensité $\Phi^m(0)$ réalisés dans les mesures aléatoires de Poisson indépendantes π^m . Pour la première itération calculons la valeur du membre de droite de (2) avec ces PPP. On utilise ensuite ces processus $\lambda_t^{m,1}$ que l'on vient de calculer pour obtenir de nouveaux processus ponctuels $S^{m,1}$ définis comme les points des π^m sous les $\lambda^{m,1}$. On recommence l'opération une infinité de fois. L'Hypothèse 1 garantit que l'itération va converger, la limite est donc un point fixe de l'itération, c'est-à-dire solution de (2). Pour l'unicité en loi, étant donné un processus de Hawkes pas nécessairement réalisé avec la méthode de thinning, on peut montrer que quitte à augmenter l'espace de probabilité, on peut toujours construire des mesures aléatoires de Poisson telles que le processus de Hawkes soit réalisé dans ces mesures aléatoires de Poisson avec la méthode de thinning [3]. L'unicité en loi découle. \square

3 Applications aux Neurosciences

Les neurosciences ont pour objet d'étude le système nerveux, que ce soit chez l'Homme ou chez d'autres organismes vivants. L'étude du système nerveux se passe à plusieurs échelles temporelles et spatiales, allant du niveau moléculaire et des nanosecondes, indispensable pour comprendre le fonctionnement de cellules comme les neurones, jusqu'au niveau macroscopique comme le cerveau dans son ensemble et des échelles de temps comme l'année pour expliquer les mécanismes à l'origine des actions, comportements et pensées d'un individu face aux stimuli venant du monde extérieur.

Un neurone, cellule emblématique du système nerveux, émet des potentiels d'actions, souvent dénommés *spikes*. Les potentiels d'actions de ce neurone, appelé pré-synaptique, vont activer les synapses des neurones post-synaptiques auxquels il est lié. L'activation de ces synapses va modifier le potentiel de membrane des neurones post-synaptiques qui, s'ils sont ainsi assez excités, pourront décharger à leur tour des spikes.

Il est possible de décrire très précisément et de manière déterministe la formation et l'émission d'un spike avec le modèle de Hodgkin-Huxley (1952) qui donne des équations décrivant l'évolution du potentiel de membrane ainsi que l'ouverture des différentes portes ioniques en réponse à un courant d'entrée qui rend compte des spikes reçus par le neurone au niveau de ses synapses. Cependant, la complexité de ces équations rend impossible l'utilisation de ce modèle pour décrire l'ensemble des neurones du cerveau en même temps. L'aléatoire entre alors naturellement en jeu : on oublie les mécanismes complexes de la création d'un spike et de sa propagation à travers l'axone pour ne retenir qu'une valeur booléenne, 0 ou 1, indiquant s'il y a eu ou non un spike. Étant donné la brièveté d'un spike, cet événement booléen est considéré comme ponctuel. Ainsi, chaque neurone émet des spikes avec une probabilité dépendant de l'activité passée des autres neurones. Les processus de Hawkes sont donc parfaitement adaptés pour modéliser un tel phénomène, on peut notamment citer les travaux suivants : [10, 4, 5]. La recherche mathématiques sur les processus de Hawkes, notamment appliquée aux neurosciences, est très active depuis une quinzaine d'années. L'ensemble \mathbf{M} correspond à l'ensemble des neurones du cerveau (ou ceux observés), pour chaque $m \in \mathbf{M}$, le processus ponctuel S^m représente les temps auquel le neurone m a émis des spikes. Les paramètres, Φ et \mathbf{h} donnent de la souplesse à la modélisation, incluant des interactions excitatrices et/ou inhibitrices par exemple.

Les expériences en neurosciences portent souvent sur des rongeurs. Des données sur le cerveau de ces rongeurs sont récoltées tout du long de l'expérience. Pour le rongeur l'expérience peut consister par exemple à explorer un nouvel espace, choisir entre deux alternatives et se rappeler de quel côté la récompense se trouve ou même localiser une plateforme très légèrement immergée et invisible depuis la surface de l'eau pour ne citer que quelques exemples. Ces différentes situations permettent de mieux comprendre le fonctionnement du cerveau du rat et de déterminer par exemple si le rat possède une mémoire épisodique, de comprendre comment le rat apprend etc. Les données récoltées sont les trains de spikes d'une dizaine de neurones, c'est-à-dire les processus ponctuels S^m pour les neurones m observés grâce à des électrodes. Les neurones observés sont répartis dans différentes zones du cerveau (hippocampe, cortex pré-frontal...). L'inférence statistique des paramètres d'un processus de Hawkes joue donc un rôle majeur dans l'analyse des données en neurosciences en révélant les interactions entre les différents neurones des différentes zones du cerveau.

4 Représentation en clusters et applications

4.1 Représentation en clusters

La représentation en clusters des processus de Hawkes, introduite par Hawkes et Oakes en 1974 [7], est un des outils les plus puissants pour mener des calculs sur ces derniers. Cette représentation est uniquement valable pour les processus de Hawkes linéaires, mais puisque dans le cas lipschitz (le plus

classique) on peut dominer un processus de Hawkes non linéaire par un processus de Hawkes linéaire, cela permet aussi de dériver des résultats sur les processus de Hawkes non linéaires.

On considère donc dans le reste de cette partie un processus de Hawkes défini par l'équation suivante

$$\lambda_t^m = \mu_m + \sum_{m' \in \mathbf{M}} \int_{-\infty}^{t-} h_{m'}^m(t-s) dS_s^{m'}, \quad t \in \mathbb{R}, m \in \mathbf{M}. \quad (4)$$

La représentation en clusters met en lumière la structure de branchement du processus de Hawkes qui se ramène à des processus de branchements poissonniens, les clusters, eux-mêmes répartis au rythme de PPP. Intuitivement, dans (4), les PPP correspondent aux taux de bases μ_m . Chaque point, dits *immigrants*, générés par les μ peut possiblement générer de nouveaux points, ses *enfants*, du fait des termes auto-excitant de (4). Ensuite chaque enfant peut à son tour générer d'autres enfants etc. Ces arrivées successives de points, avec la structure généalogique, est appelé un cluster.

Rentrons un peu plus dans les détails. Soit $\mathcal{U} = \cup_{n \in \mathbb{N}} (\mathbb{N}^*)^n$ l'arbre d'Ulam avec $(\mathbb{N}^*)^0 = \{\emptyset\}$. Un $\mathcal{Pois}(\mathbf{h})$ -cluster avec une racine de type $m_0 \in \mathbf{M}$ et né au temps $t_0 \in \mathbb{R}$ est une variable aléatoire dont la loi est donnée par la loi de $((u, \text{tp}(u), \text{bd}(u)))_{u \in \mathcal{T}}$ où

- $\mathcal{T} \subset \mathcal{U}$ est un arbre aléatoire (avec les notation Ulam Harris Neveu pour étiqueter l'arbre). Pour $u \in \mathcal{T}$, $\text{tp}(u) \in \mathbf{M}$ est le type de u et $\text{bd}(u) \in \mathbb{R}$ est la date de naissance (**birth date**) de u .
- $((u, \text{tp}(u)))_{u \in \mathcal{T}}$ est un $\mathcal{Pois}(\mathbf{H})$ -Galton Watson de racine \emptyset de type m_0 (ie $\text{tp}(\emptyset) = m_0$). Cela signifie que indépendamment, pour chaque point $u \in \mathcal{T}$, si on appelle $X(u, \text{tp}(u), m)$ le nombre d'enfants de u qui sont de type $m \in \mathbf{M}$, alors les variables aléatoires $X(u, \text{tp}(u), m)$ pour $m \in \mathbf{M}$ sont indépendantes et $X(u, \text{tp}(u), m) \sim \mathcal{Pois}(\mathbf{H}_{\text{tp}(u), m})$ pour tout $m \in \mathbf{M}$.
- Conditionnellement à $((u, \text{tp}(u)))_{u \in \mathcal{T}}$, les dates de naissances sont distribuées ainsi : pour la racine de \mathcal{T} on a $\text{bd}(\emptyset) = t_0$, ensuite, indépendamment, pour tous $u, v \in \mathcal{T}$ avec v un enfant de u , on a

$$\text{bd}(v) - \text{bd}(u) \sim \frac{h_{\text{tp}(u)}^{\text{tp}(v)}(s)}{\int h_{\text{tp}(u)}^{\text{tp}(v)}(t) dt} \mathbf{1}_{s>0} ds.$$

Pour résumer, un cluster est un arbre de Galton Watson avec une loi de reproduction $\mathcal{Pois}(\mathbf{H})$ sur lequel on a rajouté un label temporel, les dates de naissances.

Soient \mathcal{I}^m des PPP d'intensité μ_m pour $m \in \mathbf{M}$ et indépendants. Ils sont les *points immigrants*. Conditionnellement aux $(\mathcal{I}^m)_{m \in \mathbf{M}}$, on se donne des $\mathcal{Pois}(\mathbf{h})$ -clusters indépendants G_x^m avec une racine de type m et né au temps x pour $m \in \mathbf{M}$ et $x \in \mathcal{I}^m$. Le résultat suivant est la représentation en cluster des processus de Hawkes.

Proposition 3. *Sous Hypothèse 1, si $(N^m)_{m \in \mathbf{M}}$ est un processus de Hawkes selon la Définition 1, alors on a l'égalité en loi suivante*

$$(N^m)_{m \in \mathbf{M}} = \left(\{s \in \mathbb{R} \mid \exists m' \in \mathbf{M}, \exists x \in \mathcal{I}^{m'}, \exists u \in \mathcal{U} \text{ tels que } (u, m, s) \in G_x^{m'}\} \right)_{m \in \mathbf{M}} \quad (5)$$

où les immigrants $(\mathcal{I}^m)_{m \in \mathbf{M}}$ sont des PPP indépendants d'intensité $\boldsymbol{\mu}$. De manière équivalente, en loi, N^m est l'ensemble des dates de naissances des points de types m des clusters générés par les immigrants.

La méthode de thinning permet de donner une preuve visuelle de ce résultat qui se résume au graphique suivant fig. 4 dans le cas univarié.

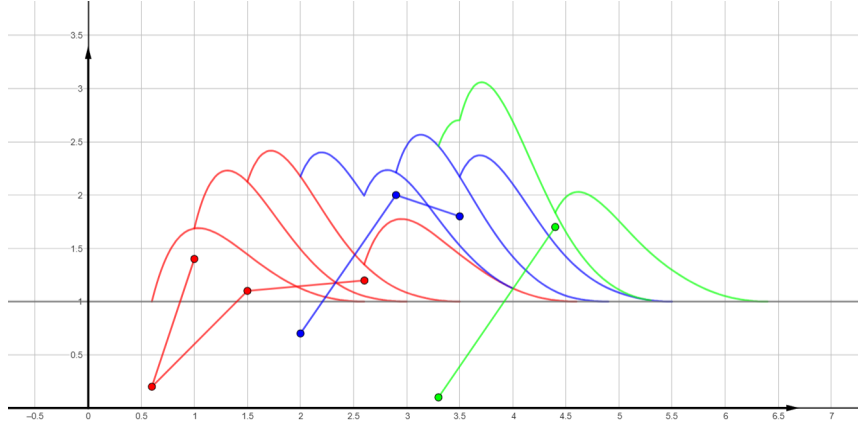


Figure 4: Illustration de la représentation en clusters pour les processus de Hawkes. Ici $\mu = 1$, les immigrants sont les points d'ordonnées inférieure à 1. On construit ensuite chaque clusters un par un, de manière à ce qu'ils soient conditionnellement indépendants les uns par rapports aux autres, et donc indépendants.

4.2 Moments exponentiels pour les processus de Hawkes

Illustrons la force de la décomposition en cluster sur un exemple : l'existence des moments exponentiels pour les processus de Hawkes. Cette partie est un résumé de la petite note que j'ai soumise à ECP [11]. Bien que pas très difficile techniquement, ce résultat n'est pas évident au premier abord et n'avait jamais été écrit par le passé.

Théorème 2 (T.L. 2024). *Soit un processus de Hawkes dont l'intensité s'écrit*

$$\lambda_t^m = \mu_m + \sum_{m' \in \mathcal{M}} \int_{-\infty}^{t-} h_{m'}^m(t-s) dS_s^{m'}, \quad t \in \mathbb{R}, m \in \mathcal{M}.$$

Supposons que l'Hypothèse 1 est satisfaite, ie le rayon spectral de $(\|h_{m'}^m\|)_{m,m' \in \mathcal{M}}$ est strictement inférieur à 1. Alors il existe $\xi > 0$ tel que pour tout ensemble borné $B \subset \mathbb{R}$ on ait

$$\mathbb{E} \left[e^{\xi S(B)} \right] < \infty$$

où $S(B) = \sum_{m \in \mathcal{M}} S^m(B)$.

Preuve. Pour éviter de se noyer dans les notations prenons un processus de Hawkes univarié et linéaire (taux de base μ et fonction d'interaction h) et l'intervalle $B = [0, 1)$. On note \mathcal{I} les immigrants et G_x les clusters pour $x \in \mathcal{I}$. la représentation en clusters nous permet d'écrire

$$\begin{aligned} \mathbb{E} \left[e^{\xi S(B)} \right] &= \mathbb{E} \left[\prod_{x \in \mathcal{I}} \mathbb{E} \left[e^{\xi G_x(B)} \mid \mathcal{I} \right] \right] \\ &= \mathbb{E} \left[\prod_{x \in \mathcal{I}} \mathbb{E} \left[e^{\xi G(B+x)} \right] \right] \end{aligned}$$

où G est un cluster né en $t = 0$. Le passage de la première à la deuxième ligne provient de l'invariante des clusters par translation temporelle. On peut voir entre ces deux premières lignes une sorte d'égalité à la *transport de masse* qui comme souvent clarifie ce type de situations. Notons $g(\xi, x) = \mathbb{E} [e^{\xi G(B+x)}]$. La formule exponentielle de Campbell donne

$$\mathbb{E} [e^{\xi S(B)}] = \exp \left(\mu \int_{\mathbb{R}} [g(\xi, x) - 1] dx \right).$$

Il reste à prouver que l'intégrale dans l'exponentielle du membre de droite est finie. Pour cela écrivons

$$\begin{aligned} \int_{\mathbb{R}} [g(\xi, x) - 1] dx &= \int_0^1 \sum_{n \in \mathbb{Z}} [g(\xi, x+n) - 1] dx \\ &= \int_0^1 \mathbb{E} \left[\sum_{n \in \mathbb{Z}} [e^{\xi G([x+n, x+n+1])} - 1] \right] dx \\ &\leq \int_0^1 \mathbb{E} \left[\prod_{n \in \mathbb{Z}} e^{\xi G([x+n, x+n+1])} \right] dx \\ &= \int_0^1 \mathbb{E} [e^{\xi G(\mathbb{R})}] dx \end{aligned}$$

pour le passage de la deuxième à la troisième ligne on a utilisé le fait que $\sum_n (y_i - 1) \leq \prod_i y_i$ pour des $y_i \geq 1$. Finalement puisque $G(\mathbb{R})$ est le cardinal d'un Galton Watson sous critique, il existe un $\xi > 0$ tel que $\mathbb{E} [e^{\xi G(\mathbb{R})}] < \infty$ et donc pour ξ assez petit, $\mathbb{E} [e^{\xi S([0,1])}] < \infty$. \square

5 Inférence statistique avec une procédure LASSO

L'inférence statistique des paramètres d'un processus de Hawkes (taux de bases et fonctions d'interactions) est un élément clé de l'analyse des données en neurosciences. En effet cela permet la reconstruction de la connectivité fonctionnelle du cerveau : les dépendances entre événements neurophysiologiques (par exemple l'émission d'un spike) de régions cérébrales distinctes. Il est ensuite possible d'interpréter les données et de tirer des conclusions : mise en évidence d'un apprentissage, activation préférentielle d'un type de mémoire, excitation d'une zone du cerveau par une autre en réponse à un stimuli de l'environnement etc.

Comment, à partir d'une observation du processus de Hawkes, donner une estimation de ses paramètres ? Pour estimer une grandeur, le/la statisticien/ne possède un plan général très efficace : répéter une même expérience un grand nombre de fois et faire concentrer une fonctionnelle de ces répétitions vers la grandeur inconnue. Dans le cas des processus ponctuel, la stationnarité joue le rôle de *même expérience* et l'observation du processus sur une grande plage de temps joue le rôle des *répétitions*.

Cette section suit les pas de [6] où les auteurs présentent une méthode robuste d'estimation des paramètres d'un processus de Hawkes linéaire. On se place dans le cas univarié par simplicité, on suppose aussi que le processus de Hawkes observé est linéaire et que sa fonction d'interaction est à support compact, son intensité s'écrit donc

$$\lambda_t = \mu + \int_{t-A}^{t-} h(t-s) dS_s \quad (6)$$

avec μ et h les vrais paramètres inconnus que l'on cherche à estimer et $A > 0$ telle que le support de h est inclus dans $[0, A]$ (A est supposée connue, ou a minima une majoration de A). Supposons que l'on dispose d'une observation du processus de Hawkes S sur $[-A, T]$. Étant donné des potentiels paramètres

(ν, g) , on peut calculer la quantité $\psi(\nu, g)$, qui correspond à l'intensité si jamais les vrais paramètres sont (ν, g) , définie par

$$\psi(\nu, g)_t = \nu + \int_{t-A}^{t-} g(t-s) dS_s, \quad t \in [0, T]. \quad (7)$$

On voudrait maintenant comparer $\psi(\nu, g)$ avec $\psi(\mu, h) = \lambda$ pour tester la plausibilité de (ν, g) comme estimateur de (μ, h) . Malheureusement λ n'est pas une quantité calculable à partir de l'observation du processus de Hawkes : on ne dispose que de S et (μ, h) sont inconnus. On en vient donc naturellement au contraste suivant :

$$\begin{aligned} \mathcal{C}(\nu, g) &= -2 \int_0^T \psi(\nu, g)_t dS_t + \int_0^T \psi(\nu, g)_t^2 dt. \\ &:= -2 \langle \psi(\nu, g), S \rangle + \|\psi(\nu, g)\|_{L^2(0, T)}^2. \end{aligned} \quad (8)$$

L'intuition derrière ce contraste est que son compensateur $\tilde{\mathcal{C}}$ s'écrit

$$\tilde{\mathcal{C}}(\nu, g) = -2 \int_0^T \psi(\nu, g)_t \lambda_t dt + \int_0^T \psi(\nu, g)_t^2 dt = \|\psi(\nu, g) - \lambda\|_{L^2(0, T)}^2 - \|\lambda\|_{L^2(0, T)}^2. \quad (9)$$

Le compensateur $\tilde{\mathcal{C}}$ mesure donc vraiment la distance entre $\psi(\nu, g)$ et λ . Il faudra donc contrôler les fluctuations entre \mathcal{C} et $\tilde{\mathcal{C}}$. Avant cela, le cadre, non paramétrique, est encore trop large : il y a trop de choix possibles pour g . On va donc utiliser un dictionnaire fini $\mathcal{D} \subset \mathbb{R} \times L^2([0, A])$ et rechercher l'estimateur des paramètres (μ, h) parmi les combinaisons linéaires du dictionnaire. Étant donnée une combinaison linéaire $a \in \mathbb{R}^{\mathcal{D}}$, les paramètres $(\nu_a, g_a) := \sum_{\varphi \in \mathcal{D}} a_\varphi \varphi$ lui sont associée et on note $\mathcal{C}_a := \mathcal{C}(\nu_a, g_a)$. La linéarité du contraste mène à l'écriture suivante

$$\mathcal{C}_a = -2 \langle a, b \rangle + \langle a, Ga \rangle \quad (10)$$

avec b et G , un vecteur et une matrice, dépendant uniquement du dictionnaire et de l'observation du processus de Hawkes.

Le contraste ci-dessus est, à son compensateur près, un contraste des moindres carrés. Ce qui caractérise les procédures d'estimations via le LASSO (Least Absolute Shrinkage and Selection Operator) est l'ajout d'une pénalisation L^1 pour optimiser le compromis biais/variance. L'estimateur des paramètres est donc donné par la combinaison linéaire \hat{a} où

$$\hat{a} = \operatorname{argmin}_{a \in \mathbb{R}^{\mathcal{D}}} \left\{ -2 \langle a, b \rangle + \langle a, Ga \rangle + 2 \langle d, |a| \rangle \right\} \quad (11)$$

avec $d \in \mathbb{R}_+^{\mathcal{D}}$ des poids.

Les auteurs de [6] ont démontré que cet estimateur satisfait une inégalité oracle sous réserve de bien choisir le dictionnaire et les poids :

- Il faut que la matrice de Gram G soit inversible de plus petite valeur propre $\gamma > 0$ avec grande probabilité, le choix du dictionnaire est important pour montrer ceci. Cela requiert aussi des propriétés de concentration du processus de Hawkes.
- Il faut aussi correctement calibrer les poids (qui peuvent dépendre de l'observation) pour optimiser le compromis biais/variance. Cela demande un contrôle des fluctuations de martingales liées au processus de Hawkes ainsi que les moments exponentiels des processus de Hawkes.

Au final on peut montrer que $(\hat{\nu}, \hat{g}) = (\nu_{\hat{a}}, g_{\hat{a}})$ vérifie sous un évènement de grande probabilité

$$\|\psi(\hat{\nu}, \hat{g}) - \lambda\|_{L^2(0, T)}^2 \leq C \inf_{a \in \mathbb{R}^{\mathcal{D}}} \left\{ \|\psi(\nu_a, g_a) - \lambda\|_{L^2(0, T)}^2 + \gamma^{-1} \sum_{\varphi \in \operatorname{supp}(a)} d_\varphi^2 \right\}. \quad (12)$$

Cette inégalité oracle traduit donc que l'estimateur est à une constante multiplicative près le meilleur possible. Cette procédure sélectionne puisque le support de a apparaît en indice de la somme des poids. On peut aussi remonter aux *vraies normes* grâce à l'inversibilité de la matrice de Gram.

Dans le cas linéaire avec interactions à supports bornés l'inférence des interactions d'un processus de Hawkes est donc bien comprise. Le critère (11) est convexe et s'implémente donc avec un faible coût calculatoire. Toutefois, l'estimation dans le cas non linéaire n'est pas encore résolue, des problèmes d'identifiabilité peuvent aussi se poser, la convexité du critère se perd aussi.

6 Pour aller plus loin : Processus HAR

Les données de neurosciences sont de nature très diverses et ne se limitent pas aux trains de spikes des neurones. Les électrodes internes qui collectent les trains de spikes récupèrent aussi un autre signal : le champ électrique local (abrégié LFP pour local field potential). Ce signal est une sorte de moyenne de l'activité globale d'une petite zone ou région du cerveau (quelques milliers de neurones). De ce signal LFP on peut extraire les rythmes cérébraux, qui correspondent aux différentes fréquences présentes dans le signal (coefficients de la décomposition du signal sur une base d'ondelettes). Les fréquences de ces oscillations peuvent aller du Hertz pour les ondes Delta à une centaine de Hertz pour les ondes Gamma.

Chez l'humain, il a été montré par exemple que les ondes Delta (1-4Hz) sont caractéristiques du sommeil profond alors que les ondes Thêta (4-8Hz) sont présentes pendant les phases de sommeil paradoxal. De manière général, les rythmes cérébraux traduisent toutes sortes d'états du cerveau liés à la mémorisation, à la réalisation de calculs mathématiques, à l'anxiété etc.

Il est connu que les rythmes interagissent entre eux [12] et peuvent aussi avoir une influence sur les trains de spikes des neurones [9]. Des expériences ont mis en évidence une interaction réciproque des spikes sur les rythmes cérébraux [14]. Cependant il n'existe pas de modèle global prenant en compte toutes les interactions possibles entre spike et rythmes cérébraux. C'est la première motivation à l'introduction des processus HAR, pour Hawkes AutoRéggressif. Les trains de spikes sont classiquement modélisés par des processus de Hawkes et l'évolution des rythmes cérébraux peut se modéliser via des processus AutoRéggressifs (AR) [15]. Une première version de ce modèle global, avec un processus de Hawkes à temps discret, a été introduite par [13], je vais présenter ici le modèle à temps continu, plus satisfaisant d'un point de vu mathématique.

Par souci de simplicité on se place dans le cas où on observe un seul neurone et un seul rythme d'une seule zone du cerveau. Si le rythme en question est une onde Delta sa fréquence est 1Hz, un nouveau rythme est généré toutes les secondes. L'objet mathématique correspondant est donc une suite $(W_n)_{n \in \mathbb{Z}}$ où W_n est la valeur du rythme au temps n . Le train de spike du neurone est lui un processus ponctuel S .

La modélisation naturelle pour S est celle des processus de Hawkes,

$$\lambda_t = \Phi_S \left(\int_{-\infty}^{t-} h_{S \rightarrow S}(t-s) dS_s \right) \quad (13)$$

où $h_{S \rightarrow S}$ est la fonction d'interaction spike sur spike et Φ_S la fonction de lien. Pour les rythmes, comme expliqué précédemment, il est classique d'utiliser une modélisation par un processus AR,

$$W_n = \Phi_W \left(\sum_{k < n} h_{W \rightarrow W}(n-k) \kappa_W(W_k) \right) + \varepsilon_n \quad (14)$$

où les ε_n sont des bruits, $h_{W \rightarrow W}$ est la fonction d'interaction rythme sur rythme et les fonctions Φ_W et κ_W donnent de la flexibilité au modèle en introduisant de la non linéarité, on retrouve la forme classique linéaire d'un processus AR en les prenant égales à l'identité.

La manière la plus naturelle de coupler ces deux modèles est de rajouter un terme AR dans l'équation de l'intensité et réciproquement de rajouter un terme Hawkes dans l'équation des rythmes. Cela donne

$$\begin{cases} \lambda_t = \Phi_S \left(\int_{-\infty}^{t-} h_{S \rightarrow S}(t-s) dS_s + \sum_{k < t} h_{W \rightarrow S}(t-k) \kappa_S(W_k) \right) \\ W_n = \Phi_W \left(\sum_{k < n} h_{W \rightarrow W}(n-k) \kappa_W(W_k) + \int_{-\infty}^{n-} h_{S \rightarrow W}(k-s) dS_s \right) + \varepsilon_n. \end{cases} \quad (15)$$

On obtient un système couplé avec les quatre interactions possibles, représentées par les quatre fonctions d'interactions. Les fonctions κ_S et κ_W ont, comme indiquées précédemment, pour rôle d'augmenter les capacités de modélisations : elles peuvent par exemple être prises égales à la valeur absolue si seule la puissance d'un rythme est la quantité pertinente dans la modélisation.

Plusieurs résultats sur les processus HAR ont été établis lors de mon stage de M2 et de ma 4ème année de l'ENS et donneront lieu à une future publication. Il sont brièvement présentés ci-dessous, rentrer dans les détails sortirait du cadre de cet IDR.

Résultats probabilistes

- **Existence** : on peut montrer l'existence des processus de HAR sous une hypothèse type rayon spectral qui étend l'[hypothèse 1](#).
- **Représentation en clusters** : A l'instar des processus de Hawkes linéaire, on peut montrer que les processus HAR *complètement linéaires* (ie Φ_S , Φ_W , κ_S , κ_W affines) ont une représentation en clusters. Dans ce cas l'intensité s'écrit

$$\lambda_t = F(t, \varepsilon_k : k < t) + \int_{-\infty}^{t-} \mathfrak{h}(s, t) dS_s \quad (16)$$

avec F une fonction déterministe et \mathfrak{h} telle que $\mathfrak{h}(s+1, t+1) = \mathfrak{h}(s, t)$. On a aussi l'écriture suivante pour les rythmes

$$W_n = c_0 + \sum_{k \leq n} C(n-k) \varepsilon_k + \int_{-\infty}^{n-} J(n, s) dS_s \quad (17)$$

avec c_0 une constante, $C(i) : i \in \mathbb{N}$ des coefficients et J une fonction telle que $J(n+1, s+1) = J(n, s)$.

- **Moments exponentiels** : En utilisant la représentation en clusters ci-dessus, qui est en fait un découplage des équations, on montre que les processus de HAR ont des moments exponentiels sous la même hypothèse que pour leur existence.

Résultats statistiques

- **Inférence des interactions** : On peut suivre la même méthode pour que dans [6] pour l'inférence des paramètres. C'est aussi ce qui a été fait dans [13]. Toutefois, dans [13], les auteurs ne sont pas parvenus à contrôler la matrice de Gram du LASSO, le résultat de [6] ne se transpose pas directement.
- **Contrôle de la matrice de Gram** : Sous l'hypothèse que l'espérance de la matrice de Gram du LASSO est inversible (ce qui ne dépend que du dictionnaire et des vrais paramètres) on est capable de montrer que la matrice l'est aussi avec grande probabilité. La clé de ce nouveau résultat est la concentration pour les processus de HAR. La concentration s'obtient en étudiant la vitesses de convergence d'un processus de HAR démarré en $t = 0$ vers le processus stationnaire démarré en $t = -\infty$ et en utilisant ensuite un couplage à la *Berbee*.

References

- [1] Pierre Brémaud. *Point Processes and Queues*. Springer New York, 1981.
- [2] Pierre Brémaud and Massoulié Laurent. Stability of nonlinear Hawkes processes. *Annals of Probability*, 1996.
- [3] Sylvain Delattre, Nicolas Fournier, and Marc Hoffmann. Hawkes processes on large networks. *The Annals of Applied Probability*, 26(1):216 – 261, 2016.
- [4] Antonio Galves and Eva Locherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *Journal of Statistical Physics*, 151:896–921, 2012.
- [5] Sonja Grün and Stefan Rotter. *Analysis of Parallel Spike Trains*. Springer, 01 2010.
- [6] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 2015.
- [7] Alan Hawkes and D.A. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 1974.
- [8] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.
- [9] Ole Jensen. Reading the hippocampal code by theta phase-locking. *Trends in Cognitive Sciences*, 9(12):551–553, 2005.
- [10] Régis C. Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, Nathalie Leresche, and Patricia Reynaud-Bouret. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of Neuroscience Methods*, 297:9–21, 2018.
- [11] Théo Leblanc. Exponential moments for Hawkes processes under minimal assumptions. working paper or preprint, March 2024.
- [12] John E. Lisman and Ole Jensen. The theta-gamma neural code. *Neuron*, 77(6):1002–1016, 2013.
- [13] Stefano Spaziani, Gabrielle Girardeau, Ingrid Bethus, and Patricia Reynaud-Bouret. Heterogeneous multiscale multivariate autoregressive model: existence, sparse estimation and application to functional connectivity in neuroscience. *Annals of Statistics*, *In press*, 2023.
- [14] Bartosz Teleńczuk, Nima Dehghani, Michel Le Van Quyen, Sydney S. Cash, Eric Halgren, Nicholas G. Hatsopoulos, and Alain Destexhe. Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Scientific Reports*, 7(1):40211, Jan 2017.
- [15] Yong Zhang, Bo Liu, Xiaomin Ji, and Dan Huang. Classification of eeg signals based on autoregressive model and wavelet packet decomposition. *Neural Processing Letters*, 45(2):365–378, Apr 2017.