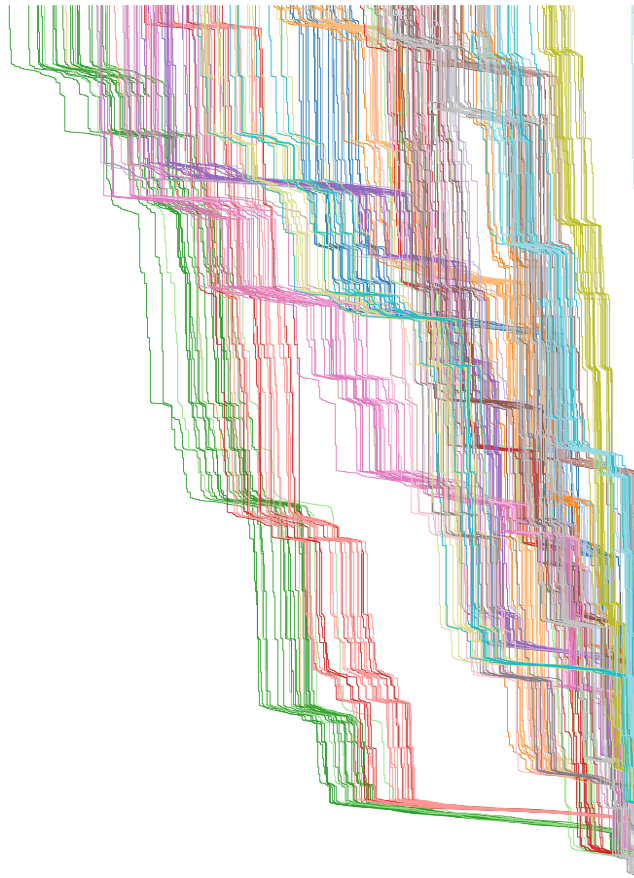


Stage de Master 2

Stage supervisé par Charlotte Dion-Blanc³
Dernière modification le 9 septembre 2026 à 15:53



Estimation L^2 de processus de Hawkes paramétriques en grande dimension



1. Département de Mathématiques et Applications de l'École Normale Supérieure - PSL, adrien.chivet@ens.psl.eu

2. Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay

3. Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, UMR 8001, Paris, F-75005, France

Table des matières

1 - Remerciements	1
I - Introduction	1
1 - État de l'art	2
2 - Contribution	3
3 - Plan	3
II - Généralités	3
1 - Intuition	4
2 - Processus de Hawkes	5
3 - Modèle statistique et notations	6
4 - Rappel du modèle linéaire	8
5 - Revue de la littérature	10
III - Résultats	12
1 - Hypothèses principales	13
2 - Estimateurs	18
3 - Inégalités de concentration	18
4 - Inégalité oracle	25
5 - Vitesse de convergence	27
IV - Unicité	28
1 - Résultats	29
2 - Preuve du Lemme	30
V - Données synthétiques	32
1 - Exemples choisis	33
2 - Détails sur l'optimisation	33
3 - Résultats expérimentaux	33
VI - Conclusion et ouverture	36
VII - Annexe	37
1 - Moyenne glissante	38
2 - Inégalité de Bernstein	38
3 - Continuité absolue par rapport à Lebesgue	39
VIII - Bibliographie	42

1 - Remerciements

Je tiens à remercier chaleureusement Charlotte Dion-Blanc pour son encadrement, sa disponibilité et ses conseils réguliers tout au long de mon stage. Son accompagnement m'a permis de prendre du recul sur mon travail, de remettre en question mes méthodes et de progresser avec davantage de confiance et de rigueur.

I – Introduction

Les processus de Hawkes constituent une famille de processus ponctuels introduite par Hawkes [18]. Ils permettent de modéliser des dynamiques temporelles complexes dans lesquelles l’occurrence d’un événement peut être influencée par les événements passés. La version multivariée de ces processus constitue une généralisation naturelle, offrant des possibilités de modélisation plus riches. En effet, en plus de l’auto-excitation propre au processus, elle permet de représenter les interactions entre différents individus connectés au sein d’un réseau, en intégrant ainsi des dépendances spatio-temporelles. Ces interactions peuvent être représentées par une matrice d’adjacence, dont chaque coefficient quantifie l’intensité de l’interaction entre deux individus du réseau. Initialement utilisés dans l’étude de la sismologie [29], les processus de Hawkes sont aujourd’hui étudiés dans de nombreux domaines, tels que la génomique [30], la finance [15], l’étude des crimes urbains [26], le football [4], les réseaux sociaux [34] ou encore les neurosciences [5]. Plus récemment, ils ont également été employés en écologie [28] afin de modéliser les interactions entre espèces, ainsi que dans [12] pour la surveillance des chauves-souris. Enfin, dans le cadre d’observations répétées d’un même processus de Hawkes sur un horizon temporel fixé, [23] propose un test du rapport de vraisemblance permettant de tester la présence d’interactions à partir de la somme des observations répétées.

Avant de considérer les problèmes d’estimation en grande dimension, il est utile de rappeler quelques résultats classiques concernant la bonne définition et la stabilité des processus de Hawkes multivariés. Dans leur article fondateur, Brémaud et Massoulié établissent des conditions générales sous lesquelles un processus de Hawkes multivarié est bien défini et admet une version stationnaire [6]. En particulier, lorsque le rayon spectral de la matrice dont les coefficients sont donnés par les normes L^1 des noyaux d’interaction est strictement inférieur à un, le processus satisfait une condition de stabilité garantissant que l’excitation générée par les événements passés n’entraîne pas de comportement explosif. Cette condition joue un rôle fondamental dans la théorie des processus de Hawkes, puisqu’elle garantit l’existence d’un régime stationnaire et permet de contrôler le comportement à long terme du processus. Ces résultats constituent ainsi un fondement théorique essentiel pour l’étude des processus de Hawkes multivariés et sont également étroitement liés aux hypothèses de stabilité utilisées en inférence et en estimation statistique.

Le contexte de la grande dimension correspond à une situation dans laquelle le nombre de paramètres M est important. Il est alors classique d’imposer une hypothèse de parcimonie sur les paramètres caractérisant le processus à partir duquel les observations sont obtenues. L’estimation de ces paramètres constitue un enjeu important pour comprendre au mieux le processus ponctuel régissant le phénomène étudié, notamment en quantifiant précisément les interactions et les phénomènes d’auto-excitation. L’étude des pénalisations adaptées à l’inférence de processus ponctuels fait notamment l’objet de la revue récente [22].

Nous allons considérer une classe de processus de Hawkes multivariés, dans un contexte de grande dimension, sur laquelle nous allons admettre certaines hypothèses classiques que l’on pourra retrouver dans la littérature. Ces hypothèses imposeront une structure sur les processus considérés et permettront alors d’obtenir un estimateur des paramètres recherchés. Nous étudierons cet estimateur pour en déduire des vitesses de convergence en norme L^2 .

1 - État de l’art

La procédure Lasso a initialement été introduite dans le cadre gaussien classique par [31] et [10]. Il s’agit aujourd’hui d’une méthode statistique largement utilisée pour les problèmes en grande dimension, notamment grâce au développement d’implémentations efficaces. Par ailleurs, ses propriétés théoriques ont été largement étudiées, comme en témoignent par exemple [24], [33] et [7]. Dans le cadre des processus de Hawkes multidimensionnels en grande dimension, l’étude d’inégalités oracle en norme 2 sur l’intensité des processus a notamment été menée par [8] et [2].

D’autres approches que la procédure Lasso ont également été proposées pour traiter ce problème. On peut notamment citer [14], qui adopte une approche non paramétrique et bayésienne. On peut également mentionner le développement du package Python `tick` par [3], permettant d’implémenter efficacement un estimateur Lasso dans le cadre des processus de Hawkes multivariés en grande dimension.

De nombreux travaux récents portent sur l’estimation de processus de Hawkes multivariés dans un régime de grande dimension. La plupart d’entre eux considèrent cependant un unique processus observé sur un horizon temporel long, contrairement à notre cadre, dans lequel nous considérons un horizon temporel fixé mais disposons de nombreuses observations i.i.d. du processus. L’article [2] étudie ainsi un estimateur Lasso dans le cadre d’un processus observé sur un horizon temporel long. Les auteurs établissent dans un premier temps une inégalité de concentration pour des opérateurs associés aux martingales compensées du processus, puis en déduisent, sous certaines hypothèses, une

inégalité oracle en norme 2. Ils obtiennent ainsi une vitesse de convergence de l'intensité de l'ordre de $\sqrt{\frac{\log(M)s^*}{T}}$, où M désigne la dimension, T l'horizon temporel du processus et s^* le paramètre de parcimonie. Dans un cadre similaire, les auteurs de [19] obtiennent notamment une convergence de leurs estimateurs Lasso en grande probabilité, à une vitesse de l'ordre de $\frac{\log^3(MT)s^*}{\sqrt{T}}$ pour la norme 1 moyennée sur les dimensions. Ces deux références considèrent donc toujours un modèle statistique reposant sur l'observation d'un processus de Hawkes sur un horizon temporel long.

Le modèle statistique considéré dans [13] diffère des précédents, puisque les auteurs étudient des répétitions i.i.d. de processus de Hawkes multivariés. Ils utilisent également une méthode de type Lasso pour construire leurs estimateurs et obtiennent, sous certaines hypothèses, une convergence en norme ∞ de l'ordre de $\frac{\log(nM^2)^4\sqrt{s^*}}{\sqrt{n}}$, où n désigne ici le nombre de processus i.i.d. observés. En outre, ils montrent que le support de leur estimateur coïncide avec le support des paramètres recherchés avec grande probabilité.

2 - Contribution

Une des caractéristiques principales d'un processus de Hawkes multivarié de dimension M est la matrice constituée des normes L^1 des noyaux d'interaction, communément appelée matrice d'adjacence. Cette matrice comporte M^2 coefficients, ce qui peut représenter un nombre de paramètres supérieur au nombre d'observations n . Nous proposons donc une hypothèse de parcimonie permettant de rendre le problème d'estimation accessible dans ce contexte de grande dimension. Nous supposons que le processus est observé n fois sur un intervalle de temps fixé. Nos hypothèses font l'objet des propriétés suivantes :

- stabilité des processus de Hawkes **(H1)**,
- parcimonie de nos paramètres **(H2)**,
- une condition de valeurs propres restreintes sur l'espérance d'un opérateur associé **(H3)**,
- et la majoration du noyau et du vecteur de baseline **(H4)**.

L'hypothèse des valeurs propres restreintes **(H3)** peut être retrouvée dans [17] dans la section 11.2.2, apparaissant classiquement dans une inégalité vraie presque sûrement, nous avons l'avons réduite à une condition vraie seulement en moyenne, c'est-à-dire en ne considérant plus l'opérateur aléatoire mais son espérance qui est donc déterministe. Nous ne supposons aucune relation vraie avec grande probabilité ou presque sûrement, contrairement aux hypothèses 6 et 7 présentes dans [13]. Ce dernier article considère aussi une hypothèse de moments exponentiels que nous évitons grâce à une utilisation du Théorème 1 de [21]. Nous montrons aussi que l'hypothèse **(H3)** assure l'identifiabilité du modèle statistique développé au cours de l'étude. Notre hypothèse de stabilité **(H1)** apparait dans [19] dans la définition de leur espace d'états en début de section 2. et aussi comme l'hypothèse 3 de [8], qui eux supposent de plus que l'intensité globale du processus est bornée uniformément et que le noyau ait un support compact, hypothèses non nécessaires dans notre cadre. Nous noterons que notre hypothèse **(H4)** est incluse dans les hypothèses 1 et 2 de [8].

Notre méthode d'estimation repose sur un risque empirique de moindres carrés associé à une pénalisation L^1 portant sur l'ensemble des paramètres, à savoir la matrice d'adjacence ainsi que le vecteur de baseline. Nous établissons des taux de convergence en norme L^2 pour les coefficients estimés du processus de Hawkes. Ces vitesses sont comparables à celles obtenues dans différents travaux portant sur des problèmes similaires, notamment [2], [8], [13] et [19]. Elles sont également comparables à celles obtenues dans le cadre linéaire gaussien classique, que l'on peut notamment retrouver dans [17].

3 - Plan

Le rapport suivante sera construit de la manière suivante. Nous commencerons par introduire dans la section II les différentes notions de bases considérées par la suite, leurs applications dans des cas simples et un résumé des résultats et méthodes apparaissant dans les articles [2], [19] et [13]. Nous passerons ensuite aux résultats présentés dans la section III qui exposera notamment le modèle et les hypothèses considérées qui permettront de déduire des inégalités de concentration grâce au Théorème 1 de [21]. Nous utiliserons aussi dans cette section les inégalités de concentration développées par [2] qui permettront de déduire une inégalité oracle en norme 2 sur l'intensité et 1 sur les paramètres. Nous pourrons ensuite conclure en utilisant une hypothèse de valeurs propres restreintes notamment exposée dans [17]. Ensuite, dans la section IV, nous discuterons de l'unicité de notre estimateur défini comme un minimiseur d'une fonction convexe. Pour finir, nous exposons des résultats expérimentaux sur des données synthétiques et sur des bases de données publiques dans la section V.

II – Généralités

1 - Intuition

Nous allons étudier une classe de processus ponctuels, c'est-à-dire de mesures aléatoires à valeurs dans $\mathbb{N}$, appelés *processus de Hawkes*, introduits par Alan G. Hawkes [18]. Ces processus interviennent aujourd'hui dans de nombreux domaines d'application, tels que la finance, les neurosciences, les sciences cognitives, l'écologie ou encore les réseaux sociaux, où ils permettent de modéliser des phénomènes d'auto-excitation.

Afin d'introduire intuitivement ces processus, considérons le problème de la modélisation des arrivées de clients dans un magasin. Supposons que chaque client potentiel soit associé à un temps d'arrivée aléatoire, indépendant de ceux des autres clients. Nous faisons en outre l'hypothèse que l'arrivée d'un client ne modifie pas les temps d'arrivée des autres. Cette propriété correspond à l'absence de mémoire du processus : la seule loi de probabilité sur $\mathbb{R}_+$ possédant cette propriété est la loi exponentielle. Plus précisément, si X suit une loi exponentielle, alors pour tous $t, s \geq 0$,

$$\mathbb{P}(X > t + s, |, X > s) = \mathbb{P}(X > t).$$

On note alors $(X_i)_{i \in \mathbb{N}}$ les temps d'arrivée des clients, supposés indépendants et identiquement distribués selon une loi exponentielle de paramètre $1/\lambda$, de moyenne λ . Le nombre de clients arrivés entre deux instants $s < t$ est alors décrit par le processus ponctuel

$$N(t) := \sum_{i \geq 0} \mathbf{1}_{0 \leq X_i \leq t}.$$

Dans ce cadre, N est un processus de Poisson homogène d'intensité λ . Nous introduisons également la filtration canonique associée au processus,

$$(\mathcal{F}_t)_{t \geq 0} := (\sigma(N_j(s), 1 \leq j \leq M, 0 \leq s \leq t))_{t \geq 0},$$

qui représente l'information disponible jusqu'à l'instant t .

L'intensité d'un processus ponctuel mesure le nombre moyen d'événements observés. Plus précisément, pour toute partie mesurable A , on définit

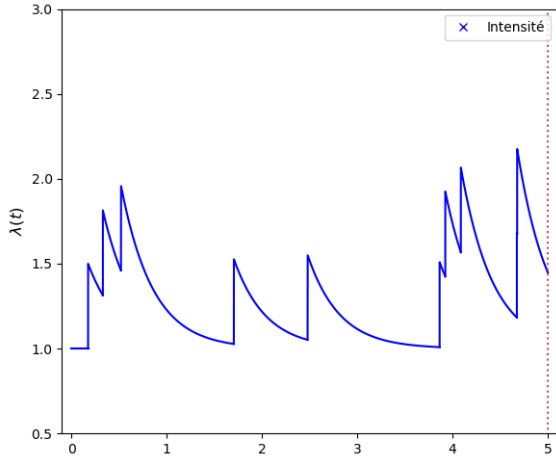
$$\lambda(A) = \mathbb{E}(N(A)).$$

Dans le cas du processus de Poisson, l'intensité est proportionnelle à la mesure de Lebesgue : un intervalle de longueur t contient en moyenne λt événements. De plus, si A et B sont deux ensembles mesurables disjoints, alors les variables aléatoires $N(A)$ et $N(B)$ sont indépendantes. Ainsi, le nombre de clients observé dans un intervalle de temps n'a aucune influence sur les arrivées futures. Cette hypothèse est cependant peu réaliste dans de nombreuses situations. Par exemple, un magasin très fréquenté peut attirer davantage de clients, tandis qu'un magasin vide peut paraître moins attractif.

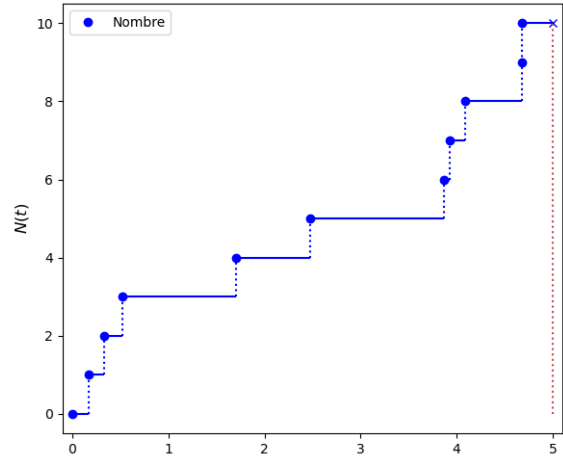
Pour prendre en compte ce phénomène, Hawkes propose de faire dépendre l'intensité du passé du processus, tout en conservant son caractère prévisible. On introduit pour cela un noyau d'interaction h , qui quantifie l'influence des événements passés sur l'intensité future :

$$\lambda(t) = \mu + \int_0^{t^-} h(t-s) dN(s),$$

où $\mu > 0$ représente une intensité de base que l'on appelle la *baseline*. Chaque nouvel événement augmente alors temporairement l'intensité, l'effet décroissant au cours du temps étant déterminé par le noyau h .



Intensité pour un noyau $t \mapsto \beta e^{-\alpha t}$



Trajectoire du processus $N(t)$

Cette construction se généralise naturellement au cas multidimensionnel. Imaginons plusieurs magasins situés à proximité les uns des autres. L'activité d'un magasin peut influencer celle de ses voisins : un magasin très fréquenté peut attirer des clients dans les commerces environnants (excitation), tandis qu'il peut également leur faire concurrence (inhibition). Pour modéliser ces interactions, on considère M processus ponctuels $(N_1, \dots, N_M)$ et l'on fait dépendre l'intensité de chaque composante des événements survenus dans toutes les autres. On obtient ainsi

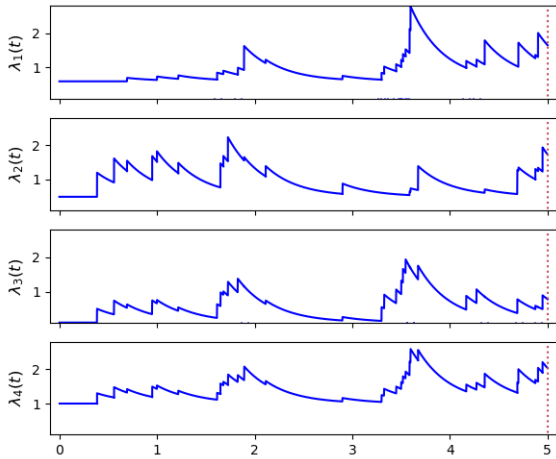
$$\lambda_j(t) = \mu_j + \sum_{i=1}^M \int_0^{t^-} h_{ji}(t-s) dN_i(s), \quad (1)$$

où le noyau h_{ji} décrit l'influence des événements de la composante i sur la composante j .

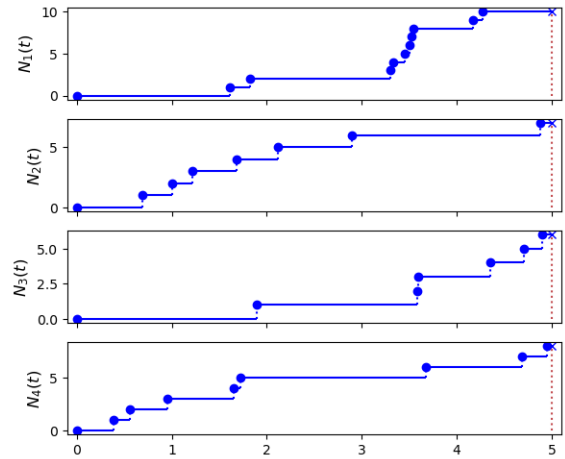
Lorsque certains noyaux sont négatifs, afin de modéliser des phénomènes d'inhibition, le terme situé à droite de (1) peut devenir négatif, ce qui est incompatible avec la définition d'une intensité. On introduit alors une fonction de transfert positive ϕ , et l'intensité est définie par

$$\lambda_j(t) = \phi \left(\mu_j + \sum_{i=1}^M \int_0^{t^-} h_{ji}(t-s) dN_i(s) \right).$$

Dans ce manuscrit, nous nous limiterons au cadre des processus de Hawkes purement excitateurs. Nous supposons donc que tous les noyaux h_{ij} sont positifs, ce qui garantit la positivité de l'intensité et permet de travailler directement avec l'équation (1).



Intensités avec des noyaux proportionnels à $t \mapsto e^{-\alpha t}$



Trajectoires du processus $N(t)$

2 - Processus de Hawkes

Pour introduire les processus de Hawkes que nous allons étudier par la suite, nous avons besoin de considérer une filtration $(\mathcal{F}_t)_{t \geq 0}$ d'un espace de probabilité implicite Ξ , où $\mathcal{F}_t$ représentera la tribu de l'information passée avant le temps t au sens large. Si les événements (ou temps de sauts) ont lieu aux temps $T_1 < T_2 < \dots$ alors on a $\mathcal{F}_t = \sigma(T_k \mid \forall k \geq 1, T_k \leq t)$ et on notera

$$N(t) = \sum_{k \geq 1} \mathbb{1}_{[0, t]}(T_k)$$

pour tout $t \geq 0$, N est ce que l'on appelle un *processus ponctuel*. Le processus $(N(t))_{t \geq 0}$ est adapté à la filtration $\mathcal{F}$ si pour tout $t \geq 0$,

$$\mathcal{F}_t = \sigma(N(s) \mid s \in [0, t]).$$

On aurait alors envie de définir une intensité conditionnelle $\lambda(t)$ qui représente le taux instantané auquel on s'attend à voir apparaître un événement, conditionnellement à tout ce que l'on sait à l'instant t . Plus précisément une relation semblable à

$$\mathbb{P}(N([t, t + dt]) = 1 \mid \mathcal{F}_{t-}) \simeq \lambda(t)dt.$$

Plus rigoureusement, un seul processus prévisible par rapport à $\mathcal{F}$ et positif λ peut vérifier la propriété que $(N_t - \int_0^t \lambda)_{t \geq 0}$ soit une $\mathcal{F}$ -martingale, et on le nomme, le cas échéant, l'intensité du processus ponctuel N . On dira finalement que N est un processus de Hawkes de *baseline* μ et de *noyau* h si pour tout $t \geq 0$,

$$\lambda(t) = \mu + \int_0^{t-} h(t-s)dN(s).$$

Pour définir un processus de Hawkes multivarié, on considère une collection $(N_j)_{1 \leq j \leq M}$ de processus ponctuels d'intensités $(\lambda_j)_{1 \leq j \leq M}$ et on dira que $(N_j)_{1 \leq j \leq M}$ est un processus de Hawkes multivarié de baseline $(\mu_j)_{1 \leq j \leq M}$ et de noyaux $(h_{j,j'})_{1 \leq j, j' \leq M}$ si pour tout $j \in [1, M]$ et $t \geq 0$,

$$\lambda_j(t) = \mu_j + \sum_{j'=1}^M \int_0^{t-} h_{j,j'}(t-s)dN_{j'}(s).$$

Le noyau $h_{j,j'}$ décrit donc l'excitation de la composante j due aux événements de la composante j' .

3 - Modèle statistique et notations

Nous allons considérer le modèle suivant de processus de Hawkes : un vecteur de baseline $\mu = (\mu_j)_{1 \leq j \leq M} \in \mathbb{R}_+^M$ et des noyaux $(h_{j,j'})_{1 \leq j, j' \leq M} = (a_{j,j'}h)_{1 \leq j, j' \leq M}$ où $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est une fonction positive sur $\mathbb{R}_+$ et $A = (a_{j,j'})_{1 \leq j, j' \leq M} \in \mathbb{R}_+^{M \times M}$ est la matrice d'interactions. Il est important de comprendre que plus μ_j est grand, plus la dimension j va avoir tendance à sauter toute seule tandis que plus $a_{j,j'}$ est élevé, plus la dimension j' va exciter la dimension j . Les paramètres d'un processus de Hawkes considéré sont alors les

$$(\mu, A) =: \theta \in \Theta := (\mathbb{R}_+)^{M \times (M+1)},$$

qui forment notre modèle statistique comme étant la loi d'un processus de Hawkes multivarié $(N_j)_{1 \leq j \leq M}$ de paramètres θ . Remarquons que l'élément θ est vu comme la matrice de taille $M \times (M+1)$ telle que pour tout $j \in [1, M]$, la ligne θ_j vaut :

$$\theta_j = (\mu_j, a_{j,1}, \dots, a_{j,M}).$$

Nous allons maintenant considérer n observations de processus de Hawkes, i.i.d. de paramètres $\theta^* \in \Theta$, que l'on va noter $(N^{(i)})_{i \geq 1}$, chacune mesurable par rapport à un même espace de probabilité Ξ . Nous étudierons toujours nos processus à horizon temporel fixé $T > 0$ contrairement à de nombreux articles sur le sujet, notamment [2] ou [8].

Nous allons imposer des hypothèses sur les paramètres θ^* .

⊙ Hypothèse (H1) :

1

Le noyau h du processus de Hawkes est positif sur son domaine de définition $\mathbb{R}_+$ et y est d'intégrale 1.
De plus les matrices A^* vérifient $\|A^*\|_{\infty \rightarrow \infty} \leq \gamma$ pour un certain $\gamma \in]0, 1[$ indépendant de n .

Ces conditions sont classiques dans l'étude des processus de Hawkes, la condition de positivité de h assure la positivité de l'intensité car nous restons dans le cadre d'un processus de Hawkes linéaire. La seconde condition sur la norme d'opérateur $\|A^*\|_{\infty \rightarrow \infty}$ implique l'inégalité de stationnarité du processus : $\rho(A^*) < 1$, elle peut se retrouver notamment dans l'article [19] dans la définition de leur espace de paramètres introduit au début de la section 2. On peut aussi voir cette hypothèse apparaître comme l'hypothèse 3 de l'article [11] portant sur des inégalités de concentration d'intégrale couplées de processus de Hawkes en temps long. On remarque alors que le processus de Hawkes associé aux paramètres $\theta \in \Theta$ est bien défini et $\theta^* \in \Theta$. Nous noterons λ_θ une intensité de processus de Hawkes de paramètres θ et en particulier $\lambda^* := \lambda_{\theta^*}$ une intensité d'un processus de Hawkes associé aux paramètres θ^* , seule sa loi est pertinente puisque λ ne peut pas être reliée aux données directement. Nous notons aussi $\lambda^{(i)} := \lambda_{\theta^*}^{(i)}$ l'intensité associée aux vrais paramètres θ^* et aux sauts $N^{(i)}$ observés.

• **Définition II.3 – 1 :**

2

Soit $\theta \in \Theta$. On définit la *martingale compensée* M , associée au processus de Hawkes multivarié N de paramètres θ , par l'égalité suivante pour tout $j \in \llbracket 1, M \rrbracket$:

$$\forall t \geq 0, M_{\theta,j}(t) := N_j([0, t]) - \int_0^t \lambda_{\theta,j}(s) ds.$$

Nous noterons avec la même logique que pour λ , la martingale $M := N - \int_0^t \lambda$ et $M^* := M_{\theta^*}$ qui est la martingale compensée associée à l'intensité λ^* dont uniquement la loi est pertinente dans notre contexte. Ce processus M_{θ} que l'on appelle martingale compensée est bien une martingale par définition de l'intensité d'un processus ponctuel. On notera aussi

$$dM_{\theta,j}(t) := dN_j(t) - \lambda_{\theta,j}(t)dt.$$

On remarquera que l'intégrale de tout processus prévisible avec $M_{\theta,j}$ restera une martingale. De plus, on notera $M_{\theta}^{(i)}$ la martingale compensée du processus $N^{(i)}$ pour tout $i \in \llbracket 1, n \rrbracket$ et pour paramètres $\theta \in \Theta$.

Dans la suite nous aurons besoin de simplifier de nombreuses fois les termes d'ordre de grandeur inférieur dans les inégalité, pour cela on introduit la notation suivante et nous introduisons différentes métriques qui seront utiles par la suite.

• **Définition II.3 – 2 :**

3

- On note $a_n \lesssim b_n$ si et seulement s'il existe une suite ε telle que pour tout $n \geq 0$, $a_n \leq b_n + \varepsilon_n$ où $\varepsilon_n \underset{n \rightarrow +\infty}{=} o(b_n)$.
- Pour tout ensemble I et toute norme $\|\cdot\|$ sur $\mathbb{R}^I$, la norme $\|\cdot\|_S := \|\cdot|_S\|$ sur $\mathbb{R}^I$ pour tout $S \subseteq I$.
- Pour toutes fonctions $f, g : [0, T] \rightarrow \mathbb{R}^M$, on définit le produit scalaire suivant :

$$\langle f | g \rangle_T := \frac{1}{T} \int_0^T f(t)^\top g(t) dt.$$

La norme associée sera dénotée par $\|\cdot\|_T$.

- On notera pour $p, q \in \mathbb{N}^*$, $A \in \mathbb{R}^{d \times r}$, I ensemble et $u \in \mathbb{R}^I$,

$$\|u\|_p := \left(\sum_{i \in I} |u_i|^p \right)^{1/p} \quad \text{et} \quad \|A\|_{p \rightarrow q} := \sup_{\|v\|_p=1} \|Av\|_q.$$

- On notera $\|\cdot\|_F$ la norme de Frobenius induite sur les matrices par le produit scalaire

$$\langle \cdot | \cdot \rangle_F : A, B \mapsto \text{Tr}(A^\top B).$$

Nos inégalités auront la mauvaise tendance à faire apparaître des termes d'ordre inférieurs et dépendant du nombre d'observations n , dont dépendra aussi la probabilité de l'évènement sur lequel cette dernière sera vérifiée. Pour cela nous devons préciser notre définition du sigle $\lesssim$ dans un contexte de variables aléatoires.

• **Définition II.3 – 3 :**

4

Soient $(X_n, Y_n)_{n \geq 0}$ des variables aléatoires réelles sur un même espace de probabilité et $p \in [0, 1]^{\mathbb{N}}$ une suite convergeant vers 0. On note $X_n \underset{n \rightarrow +\infty}{\lesssim} Y_n$ avec probabilité $1 - p_n$, s'il existe une suite de variables aléatoires sur le même espace de probabilité $(\varepsilon_n)_{n \geq 0}$ positive telle que :

$$\mathbb{P}(X_n > \varepsilon_n Y_n) \underset{n \rightarrow +\infty}{\lesssim} p_n \quad \text{et} \quad \varepsilon_n \xrightarrow{n \rightarrow +\infty} 1 \text{ presque sûrement.}$$

Nous omettrons, lorsque évident, le sous-titre $n \rightarrow +\infty$ de la notation $\underset{n \rightarrow +\infty}{\lesssim}$ qui deviendra donc simplement $\lesssim$. Cette définition va nous permettre de faciliter les conclusions des inégalités obtenues tout en gardant un socle rigoureux. De

cette notation on déduit un fait assez pratique dans son application lors d'inégalités enchaînées. Nous énonçons donc la propriété de transitivité suivante portant sur la relation $\lesssim$ entre variables aléatoires.

► **Proposition II.3-4 :**

5

Soient X_n, Y_n, Z_n des variables aléatoires sur un même espace de probabilité telles que $X_n \lesssim Y_n$ avec probabilité $1 - p_n$ et $Y_n \lesssim Z_n$ avec probabilité $1 - q_n$. Alors nous avons le résultat suivant :

$$X_n \lesssim Z_n \text{ avec probabilité } 1 - (p_n + q_n).$$

Preuve (Proposition II.3-4) : Considérons les notations de Proposition II.3-4. Alors, d'après la Définition II.3-3, on dispose de ε_n et δ_n deux suites de variables aléatoires positives sur le même espace que X_n, Y_n et Z_n qui convergent chacune vers 1 presque sûrement et vérifiant :

$$\mathbb{P}(X_n > \varepsilon_n Y_n) \lesssim p_n \text{ et } \mathbb{P}(Y_n > \delta_n Z_n) \lesssim q_n.$$

Alors on en déduit directement :

$$\mathbb{P}(X_n > \varepsilon_n \delta_n Z_n) \leq 1 - \mathbb{P}(X_n \leq \varepsilon_n Y_n \text{ et } Y_n \leq \delta_n Z_n) \leq \mathbb{P}(X_n > \varepsilon_n Y_n) + \mathbb{P}(Y_n > \delta_n Z_n) \lesssim p_n + q_n,$$

ce qui conclut. □

Avec ces définitions et notre modèle statistique clairement exposés, il faut maintenant trouver une méthode d'estimation adaptée au problème. Nous illustrons ci-dessous la méthode choisie qui est la procédure Lasso dans le cas d'un modèle linéaire simple en grande dimension.

4 - Rappel du modèle linéaire

Nous allons illustrer la procédure Lasso dans le cas d'un modèle linéaire. Nous nous baserons fortement sur le chapitre 11 de [17].

La régression linéaire standard consiste à considérer le modèle suivant

$$y = X\theta + w$$

où X est une matrice de taille $n \times p$ où n est le nombre d'observations et p leur dimension. Le vecteur $w \in \mathbb{R}^n$ est un vecteur de bruits tandis que le vecteur $\theta \in \mathbb{R}^p$ est le paramètre et on suppose l'existence de $\theta^* \in \mathbb{R}^p$ tel que $y = X\theta^* + w$. Nous nous plaçons dans le cadre de la grande dimension, c'est dire avec la possibilité pour la dimension p de diverger : $p \rightarrow +\infty$ avec la condition $p > n$. On remarque alors que la fonction $f(\theta) := \frac{1}{2} \|X\theta - y\|_2^2$ vérifie $\nabla^2 f(\theta) = X^\top X$. Ainsi, si l'on souhaite avoir une forte convexité de f , il faudra que $\text{Sp}(X^\top X)$ soit uniformément borné loin de 0. Or, ceci n'est pas le cas dès lors que $p > n$ car son rang est majoré par $\min(n, p)$. C'est là que l'on observe un phénomène de grande dimension, on a trop de paramètres par rapport au nombre d'observations. Une réponse très répandue à ce genre de problème est de considérer une pénalité L^1 en plus sur le risque empirique qui aura pour effet d'accentuer la sélection de variable dans le sens où de nombreuses variables vont être sélectionnées comme égales à 0 tandis que les variables importantes seront de l'ordre de grandeur d'une constante. Nous définissons alors ci-dessous l'estimateur Lasso associé à ce problème :

• **Définition II.4-1 :**

6

Pour $\lambda > 0$, on dit que $\hat{\theta}_\lambda$ est un estimateur Lasso s'il vérifie la condition suivante :

$$\hat{\theta}_\lambda \in \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \left(\frac{1}{2n} \|X\theta - y\|_2^2 + \lambda \|\theta\|_1 \right). \quad (2)$$

La définition d'un estimateur Lasso avec la pénalité $\lambda \|\cdot\|_1$ est en fait équivalente à un problème d'optimisation sous contrainte L^1 comme le stipule la proposition suivante. La Proposition II.4-2 peut être prouvé avec la dualité Lagrangienne comme discuté dans [17].

► **Proposition II.4-2 :**

7

Il y a une correspondance entre les solutions du problème pénalisé (2) et les solutions du problème contraint suivant

$$\underset{\|\theta\|_1 \leq R}{\operatorname{argmin}} \|X\theta - y\|_2^2.$$

Rappelons que pour toute matrice A de taille $p \times p$, si on note $\text{Sp}(A)$ l'ensemble des valeurs propres de A ,

$$\min \text{Sp}(A) = \inf_{v \neq 0} \frac{v^\top A v}{\|v\|_2^2}.$$

Nous pouvons alors définir une notion plus faible de valeur propre minimal en ne considérant qu'un nombre réduit de directions où la matrice $X^\top X$ pourrait y avoir des valeurs propres uniformément loin de 0.

• **Définition II.4 – 3 :**

8

Si on considère une partie $\mathcal{C}$ de $\mathbb{R}^p$ on dira que A vérifie la condition des *valeurs propres restreintes* à $\mathcal{C}$, de constante $\phi_0 > 0$ si pour tout $v \in \mathcal{C}$,

$$\forall v \in \mathcal{C} \setminus \{0\}, \left\langle v \mid \left(\frac{1}{n} A \right) v \right\rangle \geq \phi_0 \|v\|_2^2.$$

Nous pouvons alors énoncer le théorème suivant, en faisant apparaître la condition des valeurs propres restreintes, que l'on peut retrouver sous le nom de Théorème 11.1 dans [17]. On notera $\mathcal{C}(S, k)$ l'ensemble des v tels que $\alpha \|v\|_{1,S} \geq \|v\|_{1,S^c}$.

★ **Théorème II.4-4 :**

9

Supposons que $X^\top X$ vérifie la condition des valeurs propres restreintes à $\mathcal{C}(S, 3)$ avec pour constante $\phi_0 > 0$ et que θ^* est s^* -sparse. Alors pour $\lambda \geq 2 \|X^\top w\|_\infty / n > 0$, chaque estimateur Lasso $\hat{\theta}_\lambda$ vérifie la relation suivante :

$$\|\hat{\theta}_\lambda - \theta^*\|_2 \leq \frac{3}{\phi_0} \sqrt{s^* \lambda}.$$

Preuve (Théorème II.4 – 4) : On commence d'abord à remarquer par définition de $\hat{\theta} := \hat{\theta}_\lambda$,

$$\frac{1}{2n} \|X\hat{\theta} - y\|_2^2 \leq \frac{1}{2n} \|X\theta^* - y\|_2^2 + \lambda \left(\|\theta^*\|_1 - \|\hat{\theta}\|_1 \right)$$

ce qui donne après ré-arrangement l'inégalité de type oracle suivante :

$$\frac{1}{2n} \|X(\hat{\theta} - \theta)\|_2^2 \leq \frac{1}{n} \left\langle X(\hat{\theta} - \theta^*) \mid w \right\rangle + \lambda \left(\|\theta^*\|_1 - \|\hat{\theta}\|_1 \right). \quad (3)$$

Or, en notant $S = \text{supp}(\theta^*)$ on peut écrire $\|\hat{\theta}\|_1 = \|\theta^* + (\hat{\theta} - \theta^*)\|_{1,S} + \|\hat{\theta} - \theta^*\|_{1,S^c}$, de quoi on déduit

$$\|\hat{\theta}\|_1 \geq \|\theta^*\|_{1,S} - \|\hat{\theta} - \theta^*\|_{1,S} + \|\hat{\theta} - \theta^*\|_{1,S^c},$$

qui, en combinant avec (3) et l'hypothèse $\lambda \geq \frac{2\|X^\top w\|_\infty}{n}$, on obtient la majoration suivante :

$$\begin{aligned} \frac{1}{2n} \|X(\hat{\theta} - \theta^*)\|_2^2 &\leq \frac{1}{n} \left\langle w \mid X(\hat{\theta} - \theta^*) \right\rangle + \lambda \left(\|\hat{\theta} - \theta^*\|_{1,S} - \|\hat{\theta} - \theta^*\|_{1,S^c} \right) \\ &\leq \frac{1}{n} \|X^\top w\|_\infty \|\hat{\theta} - \theta^*\|_1 + \lambda \left(\|\hat{\theta} - \theta^*\|_{1,S} - \|\hat{\theta} - \theta^*\|_{1,S^c} \right) \\ &\leq \lambda \left(\frac{3}{2} \|\hat{\theta} - \theta^*\|_{1,S} - \frac{1}{2} \|\hat{\theta} - \theta^*\|_{1,S^c} \right) \leq \frac{3}{2} \sqrt{s^* \lambda} \|\hat{\theta} - \theta^*\|_2. \end{aligned}$$

Cette inégalité montre d'abord que $\hat{\theta} - \theta^* \in \mathcal{C}(S, 3)$. Cette observation nous permet d'appliquer l'hypothèse des valeurs propres restreintes dans l'énoncé pour passer le l'inégalité oracle à une inégalité sur la norme 2 de $\hat{\theta} - \theta$:

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{3\lambda}{\phi_0} \sqrt{s^*}$$

□

Pour mieux comprendre la majoration de l'erreur quadratique par la quantité $\frac{3\lambda}{\gamma}\sqrt{k}$ pour tout $\lambda \geq 2\|X^\top w\|_\infty/n$, il faut étudier le terme $\|X^\top w\|_\infty$. Dans le cadre d'une matrice de design X fixée et d'un bruit $w \sim \mathcal{N}(0, \sigma^2 I_n)$, on peut trouver dans le point 11.1 de [17] la majoration

$$\frac{\|X^\top w\|_\infty}{n} \leq \sigma \sqrt{\frac{\tau \log(p)}{n}} \text{ avec probabilité d'au moins } 1 - 2 \exp\left(-\frac{1}{2}(\tau - 2) \log(M)\right)$$

pour un $\tau > 2$. Ainsi, le choix de $\lambda = 2\sigma \sqrt{\frac{\tau \log(p)}{n}}$ donne la majoration suivante du risque quadratique associé à l'estimateur $\hat{\theta}_\lambda$:

$$\|\hat{\theta}_\lambda - \theta^*\|_2 \leq \frac{3\sigma}{\phi_0} \sqrt{\frac{s^* \tau \log(p)}{n}}. \quad (4)$$

Dans la partie suivante, nous allons présenter plus précisément ce qui a été fait dans la littérature des estimations de processus de Hawkes en grande dimension, que ce soit en temps long ou en observations répétées.

5 - Revue de la littérature

L'article [2] se positionne dans un contexte d'étude de processus de Hawkes à temps long. c'est-à-dire que les observations considérées sont les réalisations d'événements aléatoires dans $[0, T]$ d'un processus de Hawkes multidimensionnel avec un horizon T qui diverge : $T \rightarrow +\infty$. Nous remarquons que ce point de vue diffère de celui présenté précédemment qui sera étudié dans la suite du manuscrit. Le risque quadratique considéré dans [2] (dans le cas particulier $K = 1$) s'écrit pour tout $\theta \in \Theta$:

$$R_T(\theta) := \|\lambda_\theta\|_T^2 - 2 \sum_{j=1}^M \frac{1}{T} \int_0^T \lambda_{\theta,j}(t) dN_j(t), \quad (5)$$

où l'espace d'état Θ et l'intensité λ_θ sont définis comme précédemment. Ce risque est effectivement intéressant à étudier car c'est une approximation empirique du risque quadratique à une constante indépendante des paramètres θ près. En effet,

$$\|\lambda_\theta - \lambda^*\|_T^2 = \|\lambda_\theta\|_T^2 + \|\lambda^*\|_T^2 - 2 \sum_{j=1}^M \frac{1}{T} \int_0^T \lambda_{\theta,j}(t) \lambda_j^*(t) dt,$$

or, en utilisant la martingale compensée associée au processus, on peut écrire $\lambda_j^*(t) dt = dN_j(t) - dM_j^*(t)$ donc en espérance,

$$\mathbb{E} \left(\int_0^T \lambda_{\theta,j}(t) \lambda_j^*(t) dt \right) = \mathbb{E} \left(\int_0^T \lambda_{\theta,j}(t) dN_j(t) \right),$$

ce qui montre la relation suivante :

$$\mathbb{E} \left(\|\lambda_\theta - \lambda^*\|_T^2 \right) = \mathbb{E} (R_T(\theta)) + \|\lambda^*\|_T^2.$$

L'estimateur considéré dans [2] est associé à une fonction de pénalité pen et vérifie la condition suivante :

$$\hat{\theta} \in \operatorname{argmin}_{\theta \in \Theta} (R_T(\theta) + \operatorname{pen}(\theta)).$$

Nous allons introduire quelques notations pour définir la pénalité introduite dans ce cadre.

• Définition II.5 – 1 :

10

— Pour toute matrice $A \in \mathbb{R}^{p \times q}$, on définit la *norme nucléaire* de A comme étant la somme de ses valeurs singulières :

$$\|A\|_* := \sum_{k=1}^{\min(p,q)} \sigma_k(A).$$

— Pour tout ensemble I , tout $v \in \mathbb{R}^I$ et $p \in (\mathbb{R}_+)^I$, on note

$$\|v\|_{1,p} := \sum_{i \in I} p_i |v_i|$$

la norme 1 de v , pondérée par p .

Alors la pénalisation introduite par [2] sera pilotée par les données, c'est-à-dire que l'on va considérer $(\hat{p}, \hat{P}) = \hat{\pi} \in (\mathbb{R}_+)^{M \times (M+1)}$ et $\hat{\tau} \in \mathbb{R}_+$, à préciser, ne dépendant que des données $(N(t))_{0 \leq t \leq T}$, avec $\hat{p} \in \mathbb{R}^M$ et $\hat{P} \in \mathbb{R}^{M \times M}$. On définit alors la pénalité suivante

$$\text{pen}(\theta) := \|\theta\|_{1, \hat{\pi}} + \hat{\tau} \|A\|_*.$$

La norme L^1 favorise des estimateurs parcimonieux en annulant de nombreux coefficients, ce qui permet de sélectionner les interactions pertinentes entre composantes. La norme nucléaire favorise quant à elle des matrices de faible rang, traduisant l'existence d'une structure latente de faible dimension gouvernant les interactions. Ces deux normes restent convexe et permettent d'ajouter des pondérations par des paramètres à estimer pour améliorer les résultats de convergence.

On utilise la décomposition SVD de la matrice $A^* = U \Sigma V^\top$, en notant u_j et v_k les colonnes de U et V , qui sont respectivement les vecteurs singuliers orthogonaux gauche et droit de A . On note pour toute matrice B ,

$$\mathcal{P}_A(B) = UU^\top B + BVV^\top - UU^\top BVV^\top, \text{ et } \mathcal{P}_A^\perp(B) = (I - UU^\top)B(I - VV^\top).$$

Alors on observe dans [2] une hypothèse semblable à notre hypothèse de valeurs propres restreintes **(H3)**. En effet, cette hypothèse qui se trouve dans la définition 5 de [2] s'énonce comme l'existence d'une constante $+\infty > \kappa(\theta^*) > 0$ telle que pour tout $(\mu, A) = \theta \in \Theta$ vérifiant

$$\frac{1}{3} \|\mu\|_{1, \hat{p}, \text{supp}(\mu^*)^c} + \frac{1}{2} \|A\|_{1, \hat{P}, A^{*c}} + \frac{1}{2} \hat{\tau} \|\mathcal{P}_{A^*}^\perp(A)\|_* \leq \frac{5}{3} \|\mu\|_{1, \hat{p}, \text{supp}(\mu^*)} + \frac{3}{2} \|A\|_{1, \hat{P}, \text{supp}(A^*)} + \frac{3}{2} \hat{\tau} \|\mathcal{P}_{A^*}(A)\|_*$$

alors,

$$\max \left(\|\mu\|_{2, \text{supp}(\mu^*)}, \|A\|_{F, \text{supp}(A^*)}, \|\mathcal{P}_{A^*}(A)\|_* \right) \leq \kappa(\theta^*) \|\lambda_\theta\|_T.$$

On remarquera que la restriction sur laquelle [2] se base pour supposer une inégalité de valeurs propres sur cette restriction est clairement la paramétrisation d'un cône, très proche de notre hypothèse **(H3)**. Cet article ne stipule pas précisément que c'est une hypothèse et laisse entendre que $\kappa(\theta^*)$ peut être égal à $+\infty$. Or, dans ce cas le résultat principal d'inégalité oracle présenté dans l'article est vide puisqu'il peut s'écrire de la sorte :

$$\|\lambda_{\hat{\theta}} - \lambda_{\theta^*}\|_T \leq \frac{5}{4} \kappa(\theta^*)^2 \left(\|\hat{\pi}\|_{F, \text{supp}(\theta^*)}^2 + \hat{\tau}^2 \text{rg}(A^*) \right)$$

avec une probabilité à préciser selon un paramètre apparaissant dans $\hat{\pi}$ et $\hat{\tau}$. L'article [2] commente que l'on peut s'attendre à vitesse de convergence de $\lambda_{\hat{\theta}}$ vers la vraie intensité de l'ordre de grandeur de $\sqrt{\frac{\log(M)}{T}}$ en norme $\|\cdot\|_T$:

$$\|\lambda_{\hat{\theta}} - \lambda\|_T \lesssim \sqrt{\frac{\log(M)}{T}}.$$

Nous pouvons aussi mentionner que [8] obtient un résultat similaire avec une intensité supposée bornée et des noyaux à supports compacts dans un contexte plus large de processus de Hawkes non linéaire qui permet d'accepter les effets de noyaux négatifs (inhibition).

L'article [13] se place dans un contexte plus proche du nôtre. En effet, les auteurs y considèrent le même modèle statistique, présenté en 3. Des observations $N^{(i)}$ pour $i \in \llbracket 1, n \rrbracket$, ils définissent le risque quadratique suivant

$$R_n(\theta) := \frac{1}{n} \sum_{i=1}^n \left(\|\lambda_\theta^{(i)}\|_T^2 - \frac{2}{T} \sum_{j=1}^M \int_0^T \lambda_{\theta, j}^{(i)}(t) dN_j^{(i)}(t) \right).$$

Ce contraste se rapproche de celui introduit dans [2] que l'on a explicité en (5) mais en y incorporant la dimension $i \in \llbracket 1, n \rrbracket$. Les auteurs y étudient notamment les questions d'approximation en norme uniforme et d'obtention du support des paramètres du processus sous-jacent. Pour cela ils considèrent un estimateur de type Lasso :

$$\hat{\theta} \in \underset{\theta=(\mu, A) \in \Theta}{\text{argmin}} (R_n(\theta) + \kappa \|A\|_1), \quad (6)$$

où l'on remarque que l'estimateur est pénalisé en norme L^1 seulement selon ses coordonnées faisant référence à la matrice d'adjacence. Leur hypothèse de parcimonie ne porte aussi que sur les coordonnées des paramètres portant sur la matrice d'adjacence, c'est-à-dire qu'ils supposent que $|\text{supp}(A^*)| \leq s^* \ll M^2$ en plus d'une hypothèse sur la taille des coefficients de A^* qui stipule que $\|A^*\|_F/M$ est borné. Par rapport au vecteur de baseline, ils le supposent borné et positif, au sens où il existe des constantes μ_0 et μ_1 strictement positives telles que $\mu_0 < \mu^* < \mu_1$. Ils utilisent une hypothèse de stabilité classique en supposant un noyau h positif d'intégrale 1 sur son domaine et suppose donc le rayon

spectral de A^* strictement inférieur à 1. Pour contrôler le nombre d'évènements de leurs processus ils introduisent une hypothèse de moments exponentiels sur les nombres de sauts du processus observé :

$$\exists a > 0, \exists C > 0, \max_{1 \leq j \leq M} \mathbb{E}(\exp(aN_j(T))) < C,$$

où a et C sont constantes par rapport à n . En plus de ces hypothèses, les auteurs introduisent les conditions suivantes qui sont respectivement nommées condition de valeurs propres minimales et condition d'incohérence mutuelle, chacune vraie presque sûrement :

$$\min \text{Sp} \left(\frac{1}{n} \mathbb{H}_{S_{\theta_j^*}^*, S_{\theta_j^*}^*} \right) \geq \Lambda_0 > 0 \quad \text{et} \quad \max_{1 \leq j \leq M} \left\| \mathbb{H}_{S_{\theta_j^*}^*, S_{\theta_j^*}^*}^{-1} \mathbb{H}_{S_{\theta_j^*}^*, S_{\theta_j^*}^*} \right\|_{\infty} \leq 1 - \gamma,$$

où l'on a noté Sp l'ensemble des valeurs propres d'une matrice, Λ_0 et $\gamma \in]0, 1[$ des constantes et pour des sous-ensembles d'indices U, V et une matrice A , on note $A_{U,V}$ la matrice $(A_{u,v})_{u \in U, v \in V}$. Sous ces hypothèses, en notant $\kappa = \frac{\log^4(nM^2)}{\sqrt{n}}$, ils obtiennent sous la condition de signal minimal suivante :

$$\min_{j, j' \in \text{supp}(\theta^*)} |\theta_{j,j'}^*| > \Lambda_0^{-1} \frac{\log^4(nM^2) \sqrt{s^*}}{\sqrt{n}},$$

l'unicité de la solution $\hat{\theta}$ au problème de minimisation associé à l'équation (6) avec le coefficient de pénalité Lasso κ et avec probabilité supérieure à $1 - \frac{C}{n}$, le support est retrouvé $\text{supp}(\hat{\theta}) = \text{supp}(\theta^*)$, et,

$$\left\| \hat{\theta} - \theta^* \right\|_{\infty} \leq \frac{\sqrt{s^*} \log^4(nM^2)}{\Lambda_0 \sqrt{n}}.$$

Nous allons maintenant passer à la présentation de nos résultats.

III – Résultats

1 - Hypothèses principales

Remarquons que dans le cas linéaire, si nous essayons simplement de minimiser le risque quadratique, on se retrouve à étudier les valeurs propres de la matrice $X^\top X$ qui est de rang au plus $\min(n, p)$. Ainsi, lorsque la dimension est plus grande que le nombre d'observations, nous aurons forcément des valeurs propres nulles associées à $X^\top X$ et donc une fonction qui ne sera pas fortement convexe à minimiser. Pour remédier à cela, nous avons introduit la notion de parcimonie des paramètres, c'est-à-dire que l'on a imposé au support des vrais paramètres d'être de taille au plus constante par rapport à la dimension p ou au nombre d'observation n .

Dans notre contexte, nous avons un nombre total de $M(M + 1)$ paramètres à estimer et nous allons alors nous positionner dans un régime de grande dimension dans le sens précis suivant :

$$M, n \longrightarrow +\infty \text{ et } \frac{M}{n} \longrightarrow 0 \text{ avec } M^2 > n \text{ autorisé.}$$

Sous ce régime nous allons imposer une hypothèse de parcimonie sur les paramètres θ^* comme l'indique précisément l'hypothèse suivante :

⊙ Hypothèse (H2) :

11

Si on note $\theta^* = (\mu^*, A^*)$ les paramètres du modèle, alors on suppose que

$$|\text{supp}(\mu^*)| \leq s^* \text{ et } |\text{supp}(A^*)| \leq M s^*, \quad (7)$$

où s^* est une constante par rapport à n .

Nous allons maintenant introduire ce qui va être l'équivalent de la matrice $X^\top X$ du modèle linéaire dans notre modèle de Hawkes. Pour cela on introduit les quantités suivantes pour tout $i \in \llbracket 1, n \rrbracket$ et tout $j \in \llbracket 0, M \rrbracket$:

$$H_j^{(i)}(t) = \begin{cases} \int_0^{t^-} h(t-s) dN_j^{(i)}(s) & \text{si } j \neq 0 \\ 1 & \text{sinon} \end{cases} \quad (8)$$

et $\mathbb{H}_t$ la matrice de taille $n \times (M + 1)$ composée de ces coefficients. Un fait remarquable sur cet opérateur est l'égalité suivante pour tout $j \in \llbracket 1, M \rrbracket$,

$$\mathbb{H}_t \theta_j^\top = \left(\lambda_{\theta, j}^{(i)}(t) \right)_{1 \leq i \leq n} \text{ et donc } \mathbb{H}_t \theta^\top = \left(\lambda_{\theta, j}^{(i)}(t) \right)_{1 \leq i \leq n, 1 \leq j \leq M}. \quad (9)$$

On introduit alors l'opérateur $\mathbb{H}$ définit comme suit

$$\mathbb{H} := \frac{1}{T} \int_0^T \mathbb{H}_t^\top \mathbb{H}_t. \quad (10)$$

Comme dans le cas linéaire présenté précédemment, nous allons obtenir une inégalité oracle sur notre estimateur Lasso, pour pouvoir en déduire une inégalité sur la distance en norme 2 de notre estimateur et des paramètres réels, nous avons besoin d'une hypothèse de valeurs propres restreintes sur l'opérateur $\mathbb{H}$. Cette hypothèse permettra, en outre, d'assurer l'identifiabilité du modèle statistique considéré au sens de la [Proposition III.1 – 2](#).

⊙ Hypothèse (H3) :

12

On note $\mathcal{C}(S, \alpha)$ le cône de support S et de paramètre α comme étant la partie de $\mathbb{R}^{M \times (M+1)}$ définie comme suit :

$$\mathcal{C}(S, \alpha) := \{v \in \Theta \mid \|v\|_{1, S^c} \leq \alpha \|v\|_{1, S}\}.$$

On suppose que $\mathbb{H}$ satisfait la *condition en moyenne des valeurs propres restreintes* de constantes $\phi_0 > 0$ et $\alpha > 1$ avec $S = \text{supp}(\theta^*)$. c'est-à-dire, pour tout $v \in \mathcal{C}(S, \alpha)$,

$$\left\langle v^\top \mid \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right\rangle_F \geq \phi_0 \|v\|_F^2. \quad (11)$$

On remarquera que la condition $\alpha > 1$ inclue la condition classique $\alpha = 3$ que l'on peut retrouver dans le cas linéaire, vu ici dans le **Théorème II.4 – 4** issu de [17].

Il est important de noter que cette hypothèse n'est pas une relation presque sûre comme on peut retrouver dans les hypothèses 6 ou 7 de [13] mais une hypothèse ne dépendant que de θ^* . On remarquera que la matrice symétrique $\frac{1}{n}\mathbb{E}(\mathbb{H})$ ne dépend pas des données mais seulement de θ^* puisque l'on peut écrire ses coefficients ainsi pour tout $0 \leq k \leq l \leq M$:

$$\left(\frac{1}{n}\mathbb{E}(\mathbb{H})\right)_{k,l} = \begin{cases} \frac{1}{T} \int_0^T \mathbb{E} \left(\int_0^{t^-} \int_0^{t^-} h(t-s)h(t-u) dN_k(s) dN_l(u) \right) dt & \text{si } k \neq l \\ \frac{1}{T} \int_0^T \mathbb{E} \left(\int_0^{t^-} h(t-s) dN_l(s) \right) dt & \text{si } 0 = k < l \\ 1 & \text{sinon} \end{cases}$$

où N est un processus de Hawkes induit par θ^* .

► **Proposition III.1-1 : Valable sous les hypothèses (H1) et (H2).**

13

Les deux exemples de paramètres θ^* suivants vérifient la propriété (H3) pour tout $\alpha > 1$ avec pour constante ϕ_0 précisée ci-dessous :

- Lorsque $A^* = 0$ avec $\phi_0 \geq \frac{1}{1+\alpha^2 s^*}$.
- Dans le cadre d'un noyau exponentiel $h : t \mapsto \beta e^{-\beta t}$ avec $\beta > 0$, lorsque A^* est diagonale telle que son j -ième coefficient diagonal est nul en dehors de $S = \text{supp}(\mu^*)$ et que $\underline{\mu} := \min_{j \in S} \mu_j^*$ est minorée par une constante. On obtient $\phi_0 \geq \frac{1 \wedge K \mu^2}{1+\alpha^2 s^*}$ avec $K = 1 - \frac{3-4e^{-\beta T} + e^{-2\beta T}}{2\beta T} > 0$.

Preuve (Proposition III.1 – 1) : Dans le cas poissonien où $A^* = 0$, les $(N_j)_{1 \leq j \leq M}$ sont tous indépendants les uns des autres et suivent chacun la même loi qu'un processus de Poisson de paramètre μ_j^* . Ainsi, pour tout $k, l \in \llbracket 1, M \rrbracket$ distincts,

$$\mathbb{E} \left(\int_0^{t^-} h(t-s) dN_k(s) \right) = \mu_k^* \int_0^t h(s) ds \quad \text{et} \quad \mathbb{E} \left(\int_0^{t^-} \int_0^{t^-} h(t-s)h(t-u) dN_k(s) dN_l(u) \right) = \mu_k^* \mu_l^* \left(\int_0^t h(s) ds \right)^2.$$

Dès lors, si $v = (\mu, A) \in \mathcal{C}(\text{supp}(\theta^*), \alpha)$, on trouve la relation suivante pour tout $j, j' \in \llbracket 1, M \rrbracket$,

$$\left(v \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right)_{j,j'} = \sum_{k=1}^M A_{j,k} \mu_k^* \left(C \sum_{l=1}^M \mu_k^* A_{k,l} + c \mu_{j'} \right) + \mu_j \left(\mu_{j'} + c \sum_{l=1}^M A_{j',l} \mu_l^* \right),$$

où $C = \frac{1}{T} \int_0^T \left(\int_0^t h(s) ds \right)^2 dt$ et $c = \frac{1}{T} \int_0^T \int_0^t h(s) ds dt$. On en déduit donc la relation suivante :

$$\left\langle v^\top \mid \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right\rangle_F = \|\mu\|_2^2 + 2c \left\langle \mu^* \mid \sum_{j=1}^M \mu_j A_j \right\rangle + C \sum_{k,l,j=1}^M \mu_k^* \mu_l^* A_{k,l} A_{j,k} \geq \|\mu\|_2^2.$$

Or, nous savons que $v \in \mathcal{C}(\text{supp}(\theta^*), \alpha)$ donc si $S = \text{supp}(\mu^*)$, $\alpha \|\mu\|_{1,S} \geq \|A\|_1 + \|\mu\|_{1,S^c}$. On peut en déduire les relations suivantes par Cauchy-Schwarz :

$$\sqrt{s^*} \alpha \|\mu\|_2 \geq \sqrt{s^*} \alpha \|\mu\|_{2,S} \geq \alpha \|\mu\|_{1,S} \geq \|A\|_1 + \|\mu\|_{1,S^c} \geq \|A\|_F,$$

de quoi on peut déduire, en remarquant que $(1 + \alpha^2 s^*)^{-1}$ est constant par rapport à n ,

$$\|\mu\|_2^2 \geq \frac{1}{1 + \alpha^2 s^*} \left(\|\mu\|_2^2 + \|A\|_F^2 \right) = \phi_0 \|v\|_F^2 \quad \text{donc} \quad \left\langle v^\top \mid \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right\rangle_F \geq \frac{1}{1 + \alpha^2 s^*} \|v\|_F^2.$$

Plaçons nous dans le second cas de la proposition. L'hypothèse (H1) fourni la relation $A_{j,j}^* \in [0, 1[$ pour tout $j \in \llbracket 1, M \rrbracket$ et on remarque que les dimensions du processus de Hawkes sont indépendantes dans leur ensemble. Remarquons que pour tout $t > 0$, la fonction $m_j(t) = \mathbb{E}(\lambda_j^*(t))$ vérifie la relation suivante :

$$\begin{aligned} m_j(t) &= \mathbb{E} \left(\mu_j^* + A_{j,j}^* \int_0^{t^-} h(t-s) dN_j(s) \right) = \mu_j + A_{j,j}^* \int_0^t h(t-s) m(s) ds \\ &= \mu_j + (A_{j,j}^* h * m)(t) = \dots = \mu_j \left(1 + \sum_{k \geq 1} (A_{j,j}^*)^k (h^{*k} * 1)(t) \right), \end{aligned}$$

où la série est bien définie puisque $\|A_{j,j}^* h\|_1 = A_{j,j}^* < 1$. Il faut donc calculer les itérées de la convolution pour le noyau exponentiel. Pour ce faire on va montrer par récurrence que pour tout $n \geq 1$ et $t \geq 0$,

$$h^{*n}(t) = \frac{\beta^n t^{n-1}}{(n-1)!} \exp(-\beta t).$$

En effet, le résultat est évident pour $n = 1$ et s'il est vrai au rang $n \geq 1$, alors pour tout $t \geq 0$,

$$h^{*(n+1)}(t) = h^{*n} * h = \frac{\beta^{n+1}}{(n-1)!} \int_0^t s^{n-1} \exp(-\beta s) ds = \frac{\beta^{n+1} t^n}{n!} \exp(-\beta t).$$

Ainsi, pour tout $t > 0$, on peut calculer pour tout $k \geq 1$:

$$(h^{*k} * 1)(t) = \int_0^t h^{*k}(s) ds = 1 - \exp(-\beta t) \sum_{l=0}^{k-1} \frac{(\beta t)^l}{l!},$$

et donc obtenir :

$$\begin{aligned} \sum_{k \geq 1} (A_{j,j}^*)^k (h^{*k} * 1)(t) &= \sum_{k \geq 1} (A_{j,j}^*)^k \left(1 - \exp(-\beta t) \sum_{l=0}^{k-1} \frac{(\beta t)^l}{l!} \right) \\ &= \frac{A_{j,j}^*}{1 - A_{j,j}^*} - A_{j,j}^* \exp(-\beta t) \sum_{k \geq 0} \sum_{l=0}^k \frac{(\beta t)^l}{l!} (A_{j,j}^*)^k \\ &= \frac{A_{j,j}^*}{1 - A_{j,j}^*} - A_{j,j}^* \exp(-\beta t) \sum_{l=0}^{\infty} \frac{(\beta t)^l}{l!} \sum_{k \geq l} (A_{j,j}^*)^k \\ &= \frac{A_{j,j}^*}{1 - A_{j,j}^*} (1 - \exp(-\beta t(1 - A_{j,j}^*))). \end{aligned}$$

On en déduit finalement la formule suivante pour tout $j \in \llbracket 1, M \rrbracket$:

$$\mathbb{E} \left(\int_0^{t-} h(t-s) dN_j(s) \right) = \frac{\mu_j^*}{1 - A_{j,j}^*} (1 - \exp(-\beta t(1 - A_{j,j}^*))).$$

On trouve donc pour $k \in \llbracket 1, M \rrbracket$, les coefficients $(\frac{1}{n} \mathbb{E}(\mathbb{H}))_{k,0} = (\frac{1}{n} \mathbb{E}(\mathbb{H}))_{0,k}$ d'après la formule précédente :

$$\left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{k,0} = \frac{\mu_k^*}{1 - A_{k,k}^*} \left(1 - \frac{1}{T} \int_0^T \exp(-\beta(1 - A_{k,k}^*)t) dt \right) = \frac{\mu_k^*}{1 - A_{k,k}^*} \left(1 - \frac{1}{\beta T(1 - A_{k,k}^*)} (1 - \exp(-\beta(1 - A_{k,k}^*)T)) \right).$$

On peut en déduire par indépendance les autres coefficients de la matrice $\frac{1}{n} \mathbb{E}(\mathbb{H})$, pour tout $k, l \in \llbracket 1, M \rrbracket$,

$$\left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{k,l} = \frac{\mu_k^* \mu_l^*}{(1 - A_{k,k}^*)(1 - A_{l,l}^*)} \frac{1}{T} \int_0^T (1 - \exp(-\beta(1 - A_{k,k}^*)t)) (1 - \exp(-\beta(1 - A_{l,l}^*)t)) dt.$$

Remarquons que si l'on pose les fonctions $g_t(x) := \frac{1 - \exp(-\beta t x)}{x}$ alors

$$\left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{k,l} = \mu_k^* \mu_l^* \frac{1}{T} \int_0^T g_t(1 - A_{k,k}^*) g_t(1 - A_{l,l}^*) dt,$$

d'après le Théorème de dérivation sous le signe intégrale et l'étude des fonctions g_t , on peut montrer que par rapport aux paramètres $A_{k,k}^*, A_{l,l}^* \in [0, 1[$, la quantité $(\frac{1}{n} \mathbb{E}(\mathbb{H}))_{k,l}$ admet un minimum en $(A_{k,k}^*, A_{l,l}^*) = (0, 0)$ qui vaut $\mu_k^* \mu_l^* \left(1 - \frac{3 - 4e^{-\beta T} + e^{-2\beta T}}{2\beta T} \right)$ et une étude immédiate de l'application $x \mapsto 1 - \frac{3 - 4e^{-x} + e^{-2x}}{2x}$ montre que cette dernière est strictement croissante et continue sur $[0, +\infty[$ après prolongement. Ainsi, puisque $\beta T > 0$ on sait qu'il existe une constante $K > 0$ telle que

$$1 - \frac{3 - 4e^{-\beta T} + e^{-2\beta T}}{2\beta T} \geq K.$$

Dès lors, puisque $A_{j,j}^* \in [0, 1[$ et $\mu_j \geq 0$, nous avons l'inégalité suivante pour tout $j \in \llbracket 1, M \rrbracket$ et $(\mu, A) = v \in \mathcal{C}(\text{supp}(\theta^*), \alpha)$:

$$\begin{aligned} \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{j,j} &\geq \mu_j^2 + 2\mu_j \sum_{k=1}^M \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{k,0} A_{j,k} \\ &\geq \mu_j^2 + K \sum_{k,l=1}^M \mu_k^* \mu_l^* A_{j,k} A_{j,l}. \end{aligned}$$

On trouve alors, en notant $\underline{\mu} = \min_S \mu^*$ où S est le support de μ^* ,

$$\left\langle v^\top \mid \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right\rangle_F \geq \|\mu\|_2^2 + \underline{\mu}^2 K \sum_{j \in S} A_{j,j}^2. \quad (12)$$

Or, nous savons que v est dans le cône $\mathcal{C}(\text{supp}(\theta^*), \alpha)$ donc l'inégalité suivante est vérifiée :

$$\alpha \left(\sum_{j \in S} A_{j,j} + \sum_{j \in S} \mu_j \right) \geq \sum_{j \neq j'} A_{j,j'} + \sum_{j \notin S} A_{j,j} + \sum_{j \notin S} \mu_j,$$

où l'on suppose que l'ensemble des $j \in \llbracket 1, M \rrbracket$ tels que $A_{j,j}^* > 0$ est exactement égal à S . Ainsi, nous pouvons faire les majorations suivantes :

$$\sum_{j \neq j'=1}^M A_{j,j'}^2 + \sum_{j \notin S} A_{j,j}^2 + \sum_{j \notin S} \mu_j^2 \leq \left(\sum_{j \neq j'=1}^M A_{j,j'} + \sum_{j \notin S} A_{j,j} + \sum_{j \notin S} \mu_j \right)^2 \leq \alpha^2 \left(\sum_{j \in S} (A_{j,j} + \mu_j) \right)^2. \quad (13)$$

Or, par l'inégalité de Cauchy-Schwarz,

$$\sum_{j \in S} (A_{j,j} + \mu_j) \leq s^* \left(\sqrt{\sum_{j \in S} A_{j,j}^2} + \sqrt{\sum_{j \in S} \mu_j^2} \right),$$

de quoi on déduit en combinant en plus avec l'équation (13), l'inégalité suivante :

$$\|\mu\|_2^2 + \|A\|_F^2 \leq \sum_{j \in S} \mu_j^2 + \sum_{j \in S} A_{j,j}^2 + 2\alpha^2 s^* \left(\sum_{j \in S} \mu_j^2 + \sum_{j \in S} A_{j,j}^2 \right).$$

Mais nous avons la majoration suivante :

$$\sum_{j \in S} \mu_j^2 + \sum_{j \in S} A_{j,j}^2 \leq \left(1 \vee \frac{1}{K \underline{\mu}^2} \right) \left(\sum_{j \in S} \mu_j^2 + \underline{\mu}^2 K \sum_{j \in S} A_{j,j}^2 \right) \leq \left(1 \vee \frac{1}{K \underline{\mu}^2} \right) \left(\|\mu\|_2^2 + \underline{\mu}^2 K \sum_{j \in S} A_{j,j}^2 \right).$$

On peut donc conclure en utilisant l'inégalité (12) :

$$\begin{aligned} \left\langle v^\top \mid \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right) v^\top \right\rangle_F &\geq \|\mu\|_2^2 + \underline{\mu}^2 K \sum_{j \in S} A_{j,j}^2 \\ &\geq (1 \wedge \underline{\mu}^2 K) \left(\sum_{j \in S} \mu_j^2 + \sum_{j \in S} A_{j,j}^2 \right) \\ &\geq \frac{1 \wedge \underline{\mu}^2 K}{1 + 2\alpha^2 s^*} \|v\|_F^2. \end{aligned}$$

□

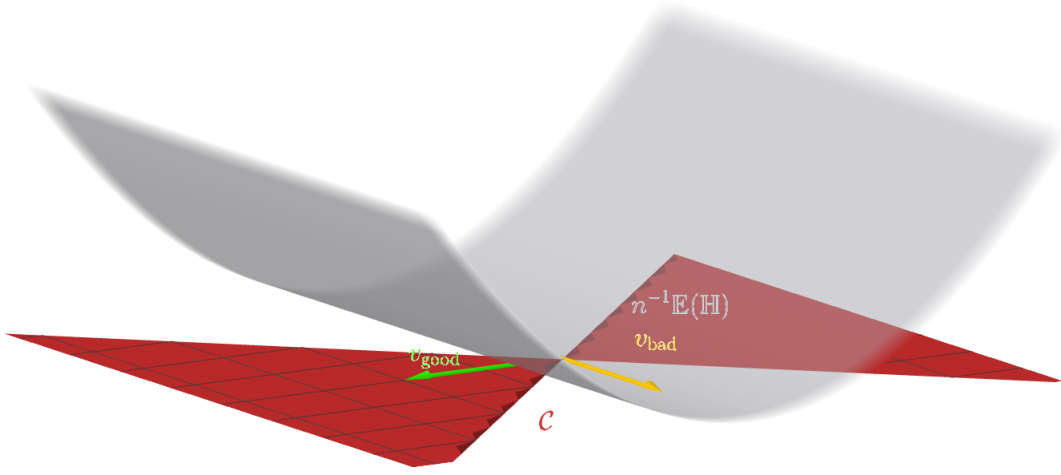


Illustration des valeurs propres restreintes à un cône $\mathcal{C}$.

L'hypothèse **(H3)** permet d'assurer que, selon un certain sous-ensemble de directions, ici le cône $\mathcal{C}(S, \alpha)$, l'opérateur $\mathbb{H}$ possède des valeurs propres toutes plus grande qu'un certain seuil, de l'ordre de grandeur de n . Nous remarquerons que cette hypothèse permet en particulier d'assurer l'identifiabilité de notre modèle, c'est ce que montre la proposition suivante.

► **Proposition III.1-2 : Valable sous l'hypothèse (H3).**

14

Si les paramètres θ^* sont tels qu'il existe une dimension $j \in \llbracket 1, M \rrbracket$ pour laquelle $N_j(T) = 0$ presque sûrement, alors $\theta_j^* = 0$.

Preuve (Proposition III.1 – 2) : Rappelons que pour tout $k, l \in \llbracket 0, M \rrbracket$,

$$\frac{1}{n} \mathbb{E}(\mathbb{H})_{k,l} = \frac{1}{T} \int_0^T \mathbb{E}(H_k(t)H_l(t)) dt,$$

avec $H_j(t) = \int_0^{t-} h(t-s)dN_j(s)$ si $j \neq 0$ et 1 sinon. Ainsi, cette quantité ne dépend que de θ^* qui caractérise la loi de N et donc de H . Considérons j tel que $N_j(T) = 0$ presque sûrement. Dans ce cadre nous avons forcément $\mu_j^* = 0$ car dans le cas contraire, pour tout $t > 0$, $\lambda_j^*(t) \geq \mu_j^*$ et donc $N_j(T) > 0$ avec probabilité d'au moins $1 - e^{-\mu_j^* T} > 0$. Ainsi, si l'on suppose par l'absurde que $\theta_j^* \neq 0$, alors on dispose de $j' \in \llbracket 1, M \rrbracket$ pour lequel $a_{j,j'}^* \neq 0$. Alors, la matrice $v := E_{j,j'}$ appartient au cône $\mathcal{C}(\text{supp}(\theta^*), \alpha)$ et,

$$\left\langle v^\top \mid \frac{1}{n} \mathbb{E}(\mathbb{H}) v^\top \right\rangle_F = \text{Tr} \left(E_{j,j'} \frac{1}{n} \mathbb{E}(\mathbb{H}) E_{j',j} \right) = \left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{j',j'} \geq \phi_0 > 0. \quad (14)$$

On en déduit que $N_{j'}(T)$ ne peut pas être nul presque sûrement car sinon $\int_0^{t-} h(t-s)dN_{j'}(s) = 0$ presque sûrement et donc on aurait $\left(\frac{1}{n} \mathbb{E}(\mathbb{H}) \right)_{j',j'} = 0$. On pourra alors supposer sans perdre de généralité que $\mu_{j'}^* > 0$ car dans le cas contraire, nous obtiendrions $j'' \in \llbracket 1, M \rrbracket$ pour lequel $a_{j',j''}^* > 0$ et donc nous pourrions itérer ce principe jusqu'à obtenir une dimension dont la baseline soit non nulle, et elle existe bien car dans le cas contraire nous aurions $N = 0$ presque sûrement, ce qui entraînerait $\frac{1}{n} \mathbb{E}(\mathbb{H}) = 0$ et donc une contradiction avec l'équation (14).

Considérons $T/2 > \delta > 0$ fourni par le [Lemme VII.1 – 1](#). Remarquons que pour tout $t \in [0, T]$,

$$\mathbb{P}(N_{j'}(t) > 0) \geq 1 - \exp(-t\mu_{j'}^*),$$

on peut alors se placer sur l'évènement $\tau_{j'} \in [0, \delta]$ où $\tau_{j'}$ est le temps de première activation de la dimension j' . On remarquera que $\tau_{j'}$ admet une densité proportionnelle à $t \mapsto \mathbb{E} \left(\lambda_{j'}^*(t) \exp \left(- \int_0^t \lambda_{j'}^*(s) ds \right) \right) \mathbb{1}_{[0, \delta]}$ par rapport à la mesure de Lebesgue avec comme coefficient de proportionnalité $L > 0$. Ainsi, pour tout $s \in [\delta, 2\delta]$,

$$\lambda_j^*(s) = \mu_j^* + \sum_{k=1}^M a_{j,k}^* \int_0^{s-} h(s-u)dN_k(u) \geq a_{j,j'}^* h(s - \tau_{j'}).$$

Or, $N_j(T)$ est nulle avec la probabilité suivante :

$$\begin{aligned}\mathbb{P}(N_j(T) = 0) &= \mathbb{E} \left(\exp \left(- \int_0^T \lambda_j^*(s) ds \right) \right) \leq \mathbb{E} \left(\exp \left(- a_{j,j'}^* \int_\delta^T h(s - \tau_{j'}) ds \right) \right) \\ &= L \int_0^\delta \exp \left(- a_{j,j'}^* \int_\delta^T h(s - \tau) ds \right) \mathbb{E} \left(\lambda_{j'}^*(\tau) \exp \left(- \int_0^\tau \lambda_{j'}^*(s) ds \right) \right) d\tau.\end{aligned}$$

Or, pour τ entre 0 et δ , nous avons l'inégalité suivante :

$$\exp \left(- \int_\delta^T h(s - \tau) ds \right) \leq \exp \left(- \int_\delta^{2\delta} h(s - \tau) ds \right) = \exp \left(- \int_{\delta-\tau}^{2\delta-\tau} h(s) ds \right).$$

Or, d'après le [Lemme VII.1 – 1](#), on sait que pour un nombre non Lebesgue-négligeable de $\tau \in [0, \delta]$, on a $\int_{\delta-\tau}^{2\delta-\tau} h(s) ds > 0$. Cette propriété montre alors l'inégalité suivante, puisque nous savons que $\lambda_{j'}^*(\tau) \geq \mu_{j'}^* > 0$,

$$\mathbb{P}(N_j(T) > 0) \geq L \int_0^\delta \left(1 - \exp \left(- a_{j,j'}^* \int_\delta^{2\delta} h(s - \tau) ds \right) \right) \mathbb{E} \left(\lambda_{j'}^*(\tau) \exp \left(- \int_0^\tau \lambda_{j'}^*(s) ds \right) \right) d\tau > 0,$$

puisque si f est Lebesgue mesurable, positive et strictement positive sur un borélien I de mesure non nulle, alors $\int_I f > 0$. Cette inégalité entre en contradiction avec notre hypothèse précédente et conclut la preuve. $\square$

2 - Estimateurs

Contrairement à ce que nous avons pu voir dans [\[2\]](#), nous ne considérons pas des processus de Hawkes en temps long mais une multitude d'observations i.i.d. de processus de Hawkes à horizon fini $T > 0$. C'est justement dans ce contexte que s'inscrit l'article [\[13\]](#) dans laquelle les auteurs étudient la récupération du support des paramètres sous des hypothèses de sparsité en grande dimension. Notre objectif sera différent de ce dernier puisque nous nous intéressons à l'erreur L^2 associé aux paramètres et non à l'erreur L^∞ étudié dans [\[13\]](#). Pour définir notre estimateur Lasso nous allons donc considérer le risque quadratique moyenné sur les données :

$$R_n(\theta) := \frac{1}{n} \sum_{i=1}^n \left(\left\| \lambda_\theta^{(i)} \right\|_T^2 - \frac{2}{T} \sum_{j=1}^M \int_0^T \lambda_{\theta,j}^{(i)}(t) dN_j^{(i)}(t) \right), \quad (15)$$

associé aux paramètres θ de notre processus de Hawkes, où l'on rappelle que les intensités sont calculées à l'aide des activations reçues $(N_j^{(i)}(t))_{0 \leq t \leq T}$. c'est-à-dire que pour tout $\theta \in \Theta$ paramètre, on a

$$\lambda_{\theta,j}^{(i)}(t) = \mu_j + \sum_{j'=1}^M a_{jj'} \int_0^{t^-} h(t-s) dN_{j'}^{(i)}(s).$$

Alors on choisit $\kappa > 0$ et on considère un estimateur $\hat{\theta}$ qui vérifie la condition suivante :

$$\hat{\theta} \in \operatorname{argmin}_{\theta \in \Theta} (R_n(\theta) + \kappa \|\theta\|_1). \quad (16)$$

Cet estimateur ne coïncide pas exactement avec l'estimateur Lasso introduit dans [\[13\]](#). En effet, les auteurs considèrent le même risque quadratique R_n que nous mais ne pénalisent pas le vecteur des baselines μ . Cette différence vient des hypothèses de parcimonie qui diffèrent entre nos deux études : ils ne supposent pas de parcimonie sur μ^* tandis que nous l'imposons avec l'hypothèse **(H2)**.

Nous allons chercher à trouver une majoration de l'écart $\left\| \hat{\theta} - \theta^* \right\|_F$ en fonction de n et pour cela nous aurons besoin de contrôler le nombre d'évènements des processus de Hawkes $N^{(i)}$ et des opérateurs en les martingales compensées introduites en [Définition II.3 – 1](#). par des inégalités de concentration

3 - Inégalités de concentration

Nous n'avons pour l'instant pas de moyen de contrôler le nombre d'activations durant notre intervalle de temps d'observation. Nous allons donc étudier des inégalités de concentration et de dispersion sur le nombre d'activations du processus et sur des opérateurs pertinents à notre étude qui dépendent notamment des martingales compensées associées à nos processus de Hawkes. Pour pouvoir obtenir des inégalités de concentrations intéressantes sur nos processus nous avons besoin d'une petite hypothèse en plus portant sur la taille en norme uniforme des paramètres et du noyau.

⊙ **Hypothèse (H4) :**

15

Le noyau h du processus de Hawkes est borné sur $\mathbb{R}_+$ et il existe une constante $m > 0$ telle que

$$\|\mu\|_\infty \leq m. \quad (17)$$

Nous verrons en 5 que nous pourrions supposer moins que le caractère constant de m par rapport à n , au coût de réduire la probabilité pour laquelle la majoration de la distance L^2 entre $\hat{\theta}$ et θ^* est vérifiée. Nos hypothèses actuelles nous permettent donc d'appliquer un cas particulier du Théorème 1 de [21] que l'on explicite ci-dessous :

► **Proposition III.3-1 :**

16

Si $\|h\|_1 = 1$ avec N_j associé aux paramètres (μ, A) , alors si $\rho(A) < 1$, il existe deux constantes $a, K > 0$ telles que

$$\mathbb{E} \left(\prod_{j=1}^M e^{a N_j(T)} \right) \leq \exp \left(KT \sum_{j=1}^M \mu_j \right).$$

Nous disposons de formules explicites suivantes, pour $r \in]0, 1[$ et $W > 0$ tels que $\sup_{n \geq 0} \|A^n / r^n\|_{\infty \rightarrow \infty}$ soit majoré par W ,

$$a = \frac{\log \left(\frac{1+r}{2r} \right)}{1 + \frac{2W}{1-r}} \text{ et } K = \frac{1-r}{2r} \text{ conviennent.}$$

En effet, nous avons supposé que nos observations $(N_j^{(i)})_{\substack{1 \leq j \leq M \\ 1 \leq i \leq n}}$ sont des processus de Hawkes i.i.d. comme défini auparavant avec les paramètres $\theta^* = (\mu^*, A^*)$ et comme noyau h . Or, on sait que h est d'intégrale 1 et positive donc $\|h\|_1 = 1$ et d'après l'hypothèse (H1), $\rho(A^*) \leq \|A^*\|_\infty < 1$. Ainsi nos $(N_j^{(i)})_{\substack{1 \leq j \leq M \\ 1 \leq i \leq n}}$ rentrent bien dans le cadre de la

Proposition III.3 – 1 ce qui nous permet d'en dériver les inégalités de concentration sur $\sum_{j=1}^M N_j^{(i)}$ qui vont suivre.

— **Corollaire III.3-2 : Valable sous les hypothèses (H1), (H2) et (H4).**

17

Nos observations $(N_j^{(i)})_{\substack{1 \leq j \leq M \\ 1 \leq i \leq n}}$ vérifient pour tout $x > 0$ les inégalités suivantes :

$$\forall i \in \llbracket 1, n \rrbracket, \mathbb{P} \left(\sum_{j=1}^M N_j^{(i)}(T) > x \right) \leq C(\gamma, T, m, s^*) e^{-a(\gamma)x}$$

$$\mathbb{P} \left(\forall i \in \llbracket 1, n \rrbracket, \sum_{j=1}^M N_j^{(i)}(T) > x \right) \leq C(\gamma, T, m, s^*) n e^{-a(\gamma)x},$$

avec des constantes a et C explicites ne dépendants que des paramètres γ, T, m , et s^* :

$$a(\gamma) = \frac{\log \left(\frac{3+\gamma}{2(1+\gamma)} \right)}{1 + \frac{4(1+\gamma)}{(1-\gamma)^2}} > 0 \text{ et } C(\gamma, T, m, s^*) = \exp \left(\frac{1-\gamma}{2(1+\gamma)} T m s^* \right) < +\infty.$$

Preuve (Corollaire III.3 – 2) : En utilisant les hypothèses (H2) et (H4), nous obtenons la borne $\sum_{j=1}^M \mu_j \leq s^* m$. Nous pouvons alors déduire de la Proposition III.3 – 1 l'inégalité suivante sur l'existence d'un moment exponentiel sur la

somme $\sum_{j=1}^M N_j(T)$:

$$\mathbb{E} \left(\exp \left(a \sum_{j=1}^M N_j(T) \right) \right) = \mathbb{E} \left(\prod_{j=1}^M e^{aN_j(T)} \right) \leq \exp \left(KT \sum_{j=1}^M \mu_j \right) \leq \exp(KTms^*) \leq C < +\infty.$$

Ainsi, pour tout $x > 0$, on déduit de l'inégalité de Markov le résultat suivant

$$\mathbb{P} \left(\sum_{j=1}^M N_j(T) > x \right) \leq e^{-ax} \mathbb{E} \left(\exp \left(a \sum_{j=1}^M N_j(T) \right) \right) \leq Ce^{-ax}.$$

De plus, puisque les $(N^{(i)})_{1 \leq i \leq n}$ sont i.i.d., on peut en déduire les inégalités suivantes :

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq i \leq n} \sum_{j=1}^M N_j^{(i)}(T) > x \right) &= 1 - \left(1 - \mathbb{P} \left(\sum_{j=1}^M N_j(T) > x \right) \right)^n \\ &\leq n \mathbb{P} \left(\sum_{j=1}^M N_j(T) > x \right) \leq Cne^{-ax}. \end{aligned}$$

Il suffit maintenant de borner le $a = \frac{\log(\frac{1+r}{1+\frac{2W}{1-r}})}$ apparaissant dans **Proposition III.3 – 1** par une quantité indépendante de n . Pour cela on considère la formule de Cauchy suivante pour $r > \rho(A)$:

$$A^n = \frac{1}{2i\pi} \oint_{|z|=r} (zI - A)^{-1} z^n dz \text{ ce qui donne } \|A^n/r^n\|_{\infty \rightarrow \infty} \leq r \sup_{|z|=r} \|(zI - A)^{-1}\|_{\infty \rightarrow \infty}.$$

Ainsi, puisque nous sommes sous l'hypothèse **(H1)**, nous en déduisons que pour $|z| = r$ la série de terme général A^n/r^n converge normalement car $\|A/r\|_{\infty \rightarrow \infty} < 1$. Nous obtenons alors les relations suivantes :

$$\|(zI - A)^{-1}\|_{\infty \rightarrow \infty} = \left\| \frac{1}{z} \sum_{k \geq 0} \left(\frac{A}{z} \right)^k \right\|_{\infty \rightarrow \infty} \leq \frac{1}{|z| - \|A\|_{\infty \rightarrow \infty}},$$

ce qui donne la majoration suivante : $\|A^n/r^n\|_{\infty \rightarrow \infty} \leq \frac{r}{r - \|A\|_{\infty \rightarrow \infty}}$. Pour conclure, le choix de $r = \frac{1+\gamma}{2}$ permet de choisir a et C comme suit :

$$a := a(\gamma) = \frac{\log\left(\frac{1+r}{1+\frac{2W}{1-r}}\right)}{1 + \frac{2W}{1-r}} = \frac{\log\left(\frac{3+\gamma}{2(1+\gamma)}\right)}{1 + \frac{4(1+\gamma)}{(1-\gamma)^2}} \text{ et } C := C(\gamma, T, m, s^*) = \exp\left(\frac{1-\gamma}{2(1+\gamma)} Tms^*\right).$$

□

Le **Corollaire III.3 – 2** va nous permettre de majorer le nombre d'activations de nos processus de Hawkes avec grande probabilité et de pouvoir appliquer le théorème suivant, un cas particulier du Theorème 4 de [2], portant sur la concentration d'opérateurs linéaires en les martingales associées à nos processus de Hawkes.

Pour introduire l'inégalité de concentration au centre de [2] nous devons introduire des quantités apparaissant respectivement comme des variances et des biais d'une inégalité de Bernstein classique.

• **Définition III.3 – 3 :**

18

Soient $(T_j^{(i)})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}}$ des fonctions prévisibles et localement bornées de $[0, T]$ dans $\mathbb{R}$. On note $V(T)$ la variance expérimentale et $b(T)$ le biais comme étant les quantités suivantes :

$$V(T) := \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \int_0^T (T_j^{(i)}(s))^2 dN_j^{(i)}(s) \text{ et } b(T) := \sup_{0 \leq s \leq T} \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} |T_j^{(i)}(s)|.$$

On introduit aussi un terme de correction logarithmique :

$$\ell_x(T) := 2 \log \log \left(2 \vee \frac{4V(T)}{x} \right) + 2 \log \log (2 \vee 4b(T)).$$

Nous pouvons enfin énoncer le Théorème suivant, démontré dans [2] où il est connu sous le nom de Théorème 4. Ce théorème nécessite de vérifier une condition sur la taille moyenne des intensités, pour cela nous allons l'introduire dans un contexte de conditionnement. En effet, les inégalités de concentration montrées dans le Corollaire III.3 – 2 vont permettre de se placer sur des événements de grande probabilités où nous pouvons contrôler les valeurs des $(N_j^{(i)})_{i,j}$ et donc des intensités.

Cette inégalité est une inégalité de type Bernstein pilotée par les données, c'est-à-dire qu'il n'y a pas de constante dépendant du modèle qui y apparaît. En effet, on observe que chacune des quantités introduites dans la Définition III.3 – 3 ne dépend que des données $(N_j^{(i)})_{\substack{1 \leq j \leq M \\ 1 \leq i \leq n}}$, y compris les martingales $M^{(i)}$.

★ **Théorème III.3-4 :**

19

Soient $\theta \in \Theta$ et Ω un événement de probabilité non nulle, en notant $\mathbb{P}_\Omega$ et $\mathbb{E}_\Omega$ les probabilités et espérances conditionnées par Ω , alors, avec les notations de la Définition III.3 – 3, il existe une constante $15 \geq c > 0$ telle que pour tout $x > 0$:

$$\mathbb{P}_\Omega \left(\left| \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \int_0^T T_j^{(i)}(s) dM_{\theta,j}^{(i)}(s) \right| \geq 2\sqrt{V(T)(x + \ell_x(T))} + c(x + \ell_x(T))(1 + b(T)) \right) \leq \frac{\pi^4}{18 \log(2)^4} e^{-x},$$

si la condition de réalisation suivante est vérifiée :

$$\mathbb{E}_\Omega \left(\int_0^T \varphi(\|T_s\|_\infty) W_{\theta,T}(s) ds \right) < +\infty \quad \text{où} \quad W_{\theta,T}(s) = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} T_j^{(i)}(s) \lambda_{\theta,j}^{(i)}(s),$$

où φ est le prolongement par continuité de la fonction $x \mapsto x^{-2}(e^{3x} - 1 - 3x)$ sur $\mathbb{R}$.

Nous allons tout d'abord utiliser le Théorème III.3 – 4 pour améliorer d'un facteur $\log$ la borne annoncée dans le Lemme C.1 de [13] grâce au corollaire suivant :

— **Corollaire III.3-5 : Valable sous les hypothèses (H1), (H2) et (H4).**

20

Notons pour tout $j, j' \in \llbracket 0, M \rrbracket \times \llbracket 1, M \rrbracket$,

$$Z_{j,j'} = \sum_{i=1}^n \int_0^T H_j^{(i)}(t) dM_{\theta^*,j'}^{(i)}(t) \quad \text{où} \quad H_j^{(i)}(t) := \int_0^{t^-} h(t-s) dN_j^{(i)}(s) \text{ si } j \neq 0 \text{ et } 1 \text{ sinon.} \quad (18)$$

Alors la majoration suivante est vraie avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n} + o\left(\frac{1}{n}\right)$ pour n plus grand qu'un rang ne dépendant que de h et γ :

$$\frac{1}{n} \max_{\substack{0 \leq j \leq M \\ 1 \leq j' \leq M}} |Z_{j,j'}| \leq \frac{4\sqrt{2}\|h\|_\infty}{a(\gamma)^{3/2}} \frac{\log(nM(M+1))^2}{\sqrt{n}} + 4c \frac{\log(nM(M+1))}{n}.$$

Preuve (Corollaire III.3 – 5) : Fixons $j, j' \in \llbracket 1, M \rrbracket \times \llbracket 0, M \rrbracket$. On remarque que si l'on choisit pour tout k, i , $T_k^{(i)}(t) := \delta_{k,j} H_{j'}^{(i)}(t)$, alors on retrouve les égalités suivantes :

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq k \leq M}} \int_0^T T_k^{(i)}(s) dM_{\theta^*,k}^{(i)}(s) = \sum_{i=1}^n \int_0^T H_{j'}^{(i)}(s) dM_{\theta^*,j}^{(i)}(s) =: Z_{j,j'}. \quad (19)$$

On calcule ensuite les quantités associées de variance et de biais $V(T)$ et $b(T)$ et on trouve les relations suivantes :

$$V(T) = \sum_{i=1}^n \int_0^T \left(H_{j'}^{(i)}(t) \right)^2 dN_j^{(i)}(t) \leq n \left(\max_{1 \leq i \leq n} N_j^{(i)}(T) \right) \sup_{0 \leq t \leq T} \max_{1 \leq i \leq n} \left(H_{j'}^{(i)}(t) \right)^2.$$

Or, pour tout t , en vertu de l'hypothèse **(H4)**, en utilisant la convention $N_0 = 0$ p.s. :

$$\left(H_{j'}^{(i)}(t)\right)^2 \leq 1 \vee \int_0^t \int_0^{t^-} h(t-s)h(t-u) dN_{j'}^{(i)}(s) dN_{j'}^{(i)}(u) \leq 1 \vee \|h\|_\infty^2 N_{j'}^{(i)}(T)^2.$$

Sur l'évènement

$$\Omega := \left(\max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} N_j^{(i)}(T) \leq \frac{2}{a(\gamma)} \log(n) \right), \quad (20)$$

on trouve pour n assez grand,

$$V(T) \leq \frac{8\|h\|_\infty^2}{a(\gamma)^3} n \log(n)^3.$$

De même, sur Ω , on obtient une borne sur le biais : $b(T) \leq \frac{2\|h\|_\infty}{a(\gamma)} \log(n)$. Dès lors, posons $x_n = \log(nM(M+1))$, et restons toujours sur l'évènement Ω . On remarque que pour n assez grand,

$$\ell_{x_n}(T) \leq \log \log \left(2 \vee \frac{32\|h\|_\infty n \log(n)^3}{a(\gamma)^3 \log(nM(M+1))} \right) + 2 \log \log (2 \vee 4\|h\|_\infty \log(n)) \leq \log(nM(M+1)).$$

Ainsi, en remarquant que pour n assez grand,

$$c(x_n + \ell_{x_n}(T))(1 + b(T)) \leq 4c \log(nM(M+1)),$$

on obtient du **Théorème III.3 – 4** la borne suivante :

$$\frac{1}{n} |Z_{j,j'}| \leq \frac{4\sqrt{2}\|h\|_\infty}{a(\gamma)^{3/2}} \frac{\log(nM(M+1))^2}{\sqrt{n}} + 4c \frac{\log(nM(M+1))}{n}$$

avec probabilité d'au moins $1 - \frac{\pi^4}{18 \log(2)^4 n M(M+1)}$ puisque sur Ω , la condition de réalisation s'écrit

$$\mathbb{E}_\Omega \left(\int_0^T \varphi \left(\max_{1 \leq i \leq n} H_{j'}^{(i)}(s) \right) \sum_{i=1}^n H_j^{(i)}(s) \lambda_{\theta^*, j'}^{(i)}(s) ds \right) < +\infty.$$

Or, sur Ω , pour tout $s \in [0, T]$, on a les bornes suivantes pour n assez grand :

$$\begin{aligned} \max_{\substack{1 \leq i \leq n \\ 0 \leq j' \leq M}} H_j^{(i)}(s) &\leq \frac{2\|h\|_\infty}{a(\gamma)} \log(n) \quad \text{et} \quad \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \lambda_{\theta^*, j}^{(i)}(s) \leq m + \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \sum_{l=1}^M a_{j,l} H_l^{(i)}(s) \\ &\leq m + \|A\|_{\infty \rightarrow \infty} \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} H_j^{(i)}(s) \leq m + \frac{2\|h\|_\infty \|A\|_{\infty \rightarrow \infty}}{a(\gamma)} \log(n), \end{aligned}$$

ce qui vérifie bien la condition de réalisation. Ainsi pour obtenir l'inégalité finale, il suffit de prendre le maximum sur les $M(M+1)$ indices j, j' pour obtenir

$$\frac{1}{n} \max_{\substack{0 \leq j \leq M \\ 1 \leq j' \leq M}} |Z_{j,j'}| \leq \frac{4\sqrt{2}\|h\|_\infty}{a(\gamma)^{3/2}} \frac{\log(nM(M+1))^2}{\sqrt{n}} + 4c \frac{\log(nM(M+1))}{n}. \quad (21)$$

Cette inégalité est vraie avec probabilité supérieure à $1 - \frac{\pi^4}{18 \log(2)^4 n}$ selon $\mathbb{P}_\Omega$. Pour conclure, il suffit de remarquer que

le **Corollaire III.3 – 2** montre que Ω est un évènement de probabilité au moins $1 - \frac{C(\gamma, T, m, s^*)}{n}$, ce qui permet de dire que l'inégalité de concentration montrée est vraie avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n} + o\left(\frac{1}{n}\right)$. $\square$

Nous allons faire usage du **Théorème III.3 – 4** une seconde fois pour prouver une inégalité de concentration sur

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} M_{\theta, j}^{(i)}(T).$$

— **Corollaire III.3-6 : Valable sous les hypothèses (H1), (H2) et (H4).**

21

Alors pour tout $\theta \in \Theta$, la majoration suivante est vraie avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4}\right) \frac{1}{n} + o\left(\frac{1}{n}\right)$ pour n plus grand qu'un rang ne dépendant que de h et γ :

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} M_{\theta,j}^{(i)}(T) \leq \frac{4}{\sqrt{a(\gamma)}} \log(nM) \sqrt{n} + 4c \log(nM).$$

Preuve (Corollaire III.3 – 6) : Considérons pour tout $j \in \llbracket 1, M \rrbracket$ et $i \in \llbracket 1, n \rrbracket$ la quantité $T_j^{(i)}(t) := 1$, ce qui nous donne d'abord $b(T) = 1$ puis

$$V(T) = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} N_j^{(i)}(T) \leq n \max_{1 \leq i \leq n} \sum_{j=1}^M N_j^{(i)}(T) \leq \frac{2}{a(\gamma)} n \log(n),$$

sur l'évènement $\Omega := \left(\max_{1 \leq i \leq n} \sum_{j=1}^n N_j^{(i)}(T) \leq \frac{2}{a(\gamma)} \log(n) \right)$. En notant $x_n = \log(nM)$, on déduit l'inégalité suivante : $\ell_{x_n}(T) \leq 2 \log \log \left(2 \vee \frac{8n \log(n)}{a(\gamma) \log(nM)} \right) + \log \log(2) \leq \log(nM)$ pour n assez grand et enfin d'après le **Théorème III.3 – 4**, nous avons l'inégalité suivante avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4}\right) \frac{1}{n} + o\left(\frac{1}{n}\right)$ sur l'espace de probabilité entier :

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} M_{\theta,j}^{(i)}(T) \leq \frac{4}{\sqrt{a(\gamma)}} \log(nM) \sqrt{n} + 4c \log(nM). \quad (22)$$

En effet, la condition de réalisation du **Théorème III.3 – 4** s'écrit comme suit dans notre cadre :

$$\mathbb{E} \left(\int_0^T \varphi(1) \sum_{i=1}^n \lambda_{\theta,j}^{(i)}(s) ds \mid \Omega \right) < +\infty.$$

Or, comme on l'a vu dans la preuve du **Corollaire III.3 – 5**, on a la majoration suivante sur Ω :

$$\max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \lambda_{\theta,j}^{(i)}(s) \leq m + \frac{2\|h\|_\infty \|A^*\|_{\infty \rightarrow \infty}}{a(\gamma)} \log(n),$$

ce qui permet de vérifier la condition d'application du **Théorème III.3 – 4** et conclut. $\square$

• **Lemme III.3-7 : Valable sous les hypothèses (H1), (H2) et (H4).**

22

Nous avons l'inégalité de concentration suivante sur la norme d'opérateur de $\mathbb{H} - \mathbb{E}(\mathbb{H})$ associée aux normes 2 à l'arrivée et au départ :

$$\mathbb{P} \left(\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{2 \rightarrow 2} \leq \frac{16\sqrt{2}\|h\|_\infty^2}{a(\gamma)^2} \sqrt{n \log(nM)^{5/2}} \right) \geq 1 - \frac{2 + C(\gamma, T, m, s^*)}{n},$$

pour tout n assez grand.

Preuve (Lemme III.3 – 7) : Nous allons utiliser l'inégalité $\|A\|_{2 \rightarrow 2} \leq \sqrt{\|A\|_{1 \rightarrow 1} \|A\|_{\infty \rightarrow \infty}}$ valable pour toute matrice A , où l'on rappelle que

$$\|A\|_{1 \rightarrow 1} = \max_j \sum_i |a_{ij}| \quad \text{et} \quad \|A\|_{\infty \rightarrow \infty} = \max_i \sum_j |a_{ij}|.$$

Dans le cas où A est symétrique, on trouve tout de suite $\|A\|_{1 \rightarrow 1} = \|A\|_{\infty \rightarrow \infty}$. Nous allons utiliser cela car nous savons que $\sum_{j=1}^M N_j(T)$ possède un moment exponentiel d'après le **Corollaire III.3 – 2**. Nous allons utiliser l'inégalité de

Bernstein dans le sens précisé dans le **Théorème VII.2 – 1** pour en déduire une inégalité de concentration. En effet,

$$\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{1 \rightarrow 1} = \max_{0 \leq l \leq M} \sum_{k=0}^M |\mathbb{H}_{k,l} - \mathbb{E}(\mathbb{H}_{k,l})| = \max_{0 \leq l \leq M} \left| \sum_{i=1}^n X_l^{(i)} \right|,$$

avec

$$X_l^{(i)} = \sum_{k=0}^M \frac{1}{T} \int_0^T H_k^{(i)}(t) H_l^{(i)}(t) dt - \mathbb{E} \left(\frac{1}{T} \int_0^T H_k^{(i)}(t) H_l^{(i)}(t) dt \right). \quad (23)$$

Or, pour tous $i \in \llbracket 1, n \rrbracket$ et tous $l \in \llbracket 0, M \rrbracket$, on trouve la majoration suivante lorsque $l \neq 0$:

$$\begin{aligned} \left| \sum_{k=0}^M \frac{1}{T} \int_0^T H_k^{(i)}(t) H_l^{(i)}(t) dt \right| &\leq \frac{1}{T} \int_0^T \int_0^{t^-} h(t-s) dN_l^{(i)}(s) + \sum_{k=1}^M \frac{1}{T} \int_0^T \int_0^{t^-} \int_0^{t^-} h(t-s) h(t-u) dN_k^{(i)}(s) dN_l^{(i)}(u) \\ &\leq \|h\|_\infty \left(\max_{1 \leq k \leq M} N_k^{(i)}(T) \right) \left(1 + \|h\|_\infty \sum_{k=1}^M N_k^{(i)}(T) \right). \end{aligned}$$

et lorsque $l = 0$, nous nous retrouvons avec la majoration suivante :

$$\left| \sum_{k=0}^M \frac{1}{T} \int_0^T H_k^{(i)}(t) H_0^{(i)}(t) dt \right| = 1 + \sum_{k=1}^M \frac{1}{T} \int_0^T h(t-s) dN_k^{(i)}(s) \leq 1 + \|h\|_\infty \sum_{k=1}^M N_k^{(i)}(T).$$

Et, nous savons que $\mathbb{P} \left(\max_{1 \leq i \leq n} \sum_{k=1}^M N_k^{(i)}(T) > x \right) \leq nC(\gamma, T, m, s^*) e^{-a(\gamma)x}$, ainsi si on note

$$\tilde{\Omega} := \left(\max_{1 \leq i \leq n} \sum_{k=1}^M N_k^{(i)}(T) \leq \frac{2}{a(\gamma)} \log(n) \right), \quad \tilde{\Omega} \text{ est de probabilité d'au moins } 1 - \frac{C(\gamma, T, m, s^*)}{n} \quad (24)$$

et en se plaçant sur $\tilde{\Omega}$, nous avons la majoration suivante pour tout $i \in \llbracket 1, n \rrbracket$ et $l \in \llbracket 0, M \rrbracket$, puisque $\sum_{k=1}^M N_k^{(i)}(T) \geq \max_{1 \leq k \leq M} N_k^{(i)}(T)$, où $X_l^{(i)}$ a été défini en (23),

$$\begin{aligned} |X_l^{(i)}| &= \left| \sum_{k=0}^M \left(\frac{1}{T} \int_0^T H_k^{(i)}(t) H_l^{(i)}(t) dt - \mathbb{E} \left(\frac{1}{T} \int_0^T H_k^{(i)}(t) H_l^{(i)}(t) dt \right) \right) \right| \\ &\leq 2 \left(1 + \|h\|_\infty \frac{2 \log(n)}{a(\gamma)} \right) \left(1 \vee \|h\|_\infty \frac{2 \log(n)}{a(\gamma)} \right) \leq \frac{16 \|h\|_\infty^2}{a(\gamma)^2} \log(n)^2, \end{aligned}$$

pour tout n plus grand qu'un certain rang ne dépendant que de h et de γ . On déduit de ces majorations et du **Théorème VII.2 – 1** la borne suivante pour tout $t > 0$ et $l \geq 0$:

$$\mathbb{P}_{\tilde{\Omega}} \left(\left| \sum_{i=1}^n X_l^{(i)} \right| > t \right) \leq 2 \exp \left(- \frac{t^2}{2n \left(\frac{16 \|h\|_\infty^2}{a(\gamma)^2} \log(n)^2 \right)^2 + \frac{2t}{3} \frac{16 \|h\|_\infty^2}{a(\gamma)^2} \log(n)^2} \right),$$

donc le choix de $t = \frac{16\sqrt{2}\|h\|_\infty^2}{a(\gamma)^2} \sqrt{n} \log(nM)^{5/2}$ permet de trouver qu'avec probabilité d'au moins $1 - \frac{2}{nM} + o\left(\frac{1}{nM}\right)$, sur l'évènement $\tilde{\Omega}$,

$$\left| \sum_{i=1}^n X_l^{(i)} \right| \leq \frac{16\sqrt{2}\|h\|_\infty^2}{a(\gamma)^2} \sqrt{n} \log(nM)^{5/2}.$$

Or, nous avons la borne suivante pour passer au maximum sur l :

$$\begin{aligned} \mathbb{P}_{\tilde{\Omega}} (\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{1 \rightarrow 1} > t) &= \mathbb{P}_{\tilde{\Omega}} \left(\max_{0 \leq l \leq M} \left| \sum_{i=1}^n X_l^{(i)} \right| > t \right) \\ &= \mathbb{P}_{\tilde{\Omega}} \left(\bigcup_{0 \leq l \leq M} \left(\left| \sum_{i=1}^n X_l^{(i)} \right| > t \right) \right) \leq \sum_{l=0}^M \mathbb{P}_{\tilde{\Omega}} \left(\left| \sum_{i=1}^n X_l^{(i)} \right| > t \right) \\ &\leq (M+1) \frac{2}{nM} + (M+1) o\left(\frac{1}{nM}\right) = \frac{2}{n} + o\left(\frac{1}{n}\right). \end{aligned}$$

Ainsi, puisque nous pouvons contrôler la norme d'opérateur $2 \rightarrow 2$ d'une matrice symétrique par la norme d'opérateur $1 \rightarrow 1$ comme suit :

$$\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{2 \rightarrow 2} \leq \sqrt{\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{1 \rightarrow 1} \|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{\infty \rightarrow \infty}} = \|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{1 \rightarrow 1} \leq \max_{1 \leq l \leq M} \left| \sum_{i=1}^n X_l^{(i)} \right|,$$

nous pouvons conclure d'après (24) qu'avec probabilité d'au moins $1 - \frac{2+C(\gamma, T, m, s^*)}{n} + o\left(\frac{1}{n}\right)$, sur Ξ tout entier,

$$\|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{2 \rightarrow 2} \leq \frac{16\sqrt{2}\|h\|_{\infty}^2}{a(\gamma)^2} \sqrt{n} \log(nM)^{5/2}.$$

□

— **Corollaire III.3-8 : Valable sous les hypothèses (H1), (H2), (H3) et (H4).**

23

L'hypothèse **(H3)** imposant une borne inférieure aux valeurs propres de $\mathbb{E}(\mathbb{H})$ restreintes à un cône se généralise sur $\mathbb{H}$ avec grande probabilité en conservant le même ordre de grandeur de ces valeurs propres. Plus précisément, nous avons la minoration suivante :

$$\mathbb{P} \left(\forall \theta \in \mathcal{C}(S, \alpha), \frac{1}{n} \|\mathbb{H}\theta^\top\|_F^2 \geq \left(\phi_0 - \frac{16\sqrt{2}\|h\|_{\infty}^2 \log(nM)^{5/2}}{a(\gamma)^2 \sqrt{n}} \right) \|\theta\|_F^2 \right) \geq 1 - \frac{2 + C(\gamma, T, m, s^*)}{n}.$$

Preuve (Corollaire III.3 – 8) : Rappelons maintenant que l'hypothèse **(H3)** stipule que $\mathbb{E}(\mathbb{H})$ vérifie la condition de valeurs propres restreintes suivante :

$$\forall \theta \in \mathcal{C}(S, \alpha), \frac{1}{n} \|\mathbb{E}(\mathbb{H})^{1/2} \theta^\top\|_F^2 \geq \phi_0 \|\theta\|_F^2.$$

D'abord, essayons de relier les normes de Frobenius de l'hypothèse avec la norme d'opérateur $\|\cdot\|_{2 \rightarrow 2}$. Pour cela on procède simplement au calcul suivant pour une matrice $A \in \mathbb{R}^{n \times (M+1)}$ quelconque et $\theta \in \mathbb{R}^{M \times (M+1)}$ de lignes les $(\theta_j)_{1 \leq j \leq M}$:

$$\begin{aligned} \|A\theta^\top\|_F^2 &= \|A \begin{pmatrix} \theta_1^\top & \cdots & \theta_M^\top \end{pmatrix}\|_F^2 = \|(A\theta_1^\top \mid \cdots \mid A\theta_M^\top)\|_F^2 \\ &= \sum_{j=1}^M \|A\theta_j^\top\|_2^2 \leq \|A\|_{2 \rightarrow 2}^2 \sum_{j=1}^M \|\theta_j^\top\|_2^2 = \|A\|_{2 \rightarrow 2}^2 \|\theta\|_F^2. \end{aligned}$$

On en déduit donc simplement que pour tout paramètre $\theta \in \Theta$,

$$\|(\mathbb{H} - \mathbb{E}(\mathbb{H}))\theta^\top\|_F \leq \|\mathbb{H} - \mathbb{E}(\mathbb{H})\|_{2 \rightarrow 2} \|\theta\|_F$$

Observons maintenant que pour tout $\theta \in \mathcal{C}(S, \alpha)$,

$$\begin{aligned} \langle \theta^\top \mid \mathbb{H}\theta^\top \rangle &= \langle \theta^\top \mid \mathbb{E}(\mathbb{H})\theta^\top \rangle + \langle \theta^\top \mid (\mathbb{H} - \mathbb{E}(\mathbb{H}))\theta^\top \rangle \\ &\geq n\phi_0 \|\theta\|_F^2 - \|\theta\|_F \|(\mathbb{H} - \mathbb{E}(\mathbb{H}))\theta^\top\|_F \\ &\geq \|\theta\|_F^2 \left(n\phi_0 - \frac{16\sqrt{2}\|h\|_{\infty}^2}{a(\gamma)^2} \sqrt{n} \log(nM)^{5/2} \right) \gtrsim n\phi_0 \|\theta\|_F^2 \end{aligned}$$

avec probabilité supérieure à $1 - \frac{2+C(\gamma, T, m, s^*)}{n}$ d'après le [Lemme III.3 – 7](#). Ainsi, le bruit engendré par $\mathbb{H} - \mathbb{E}(\mathbb{H})$ est négligeable devant la taille demandée aux valeurs propres restreintes de l'hypothèse **(H3)**. □

4 - Inégalité oracle

Dans cette section nous allons adapter la preuve du Théorème 6 apparaissant dans [2] pour obtenir une inégalité oracle dans notre contexte. Nous rappelons que $\kappa > 0$ est la pénalité associée à notre estimateur Lasso, ce dernier vérifiant

$$\hat{\theta} \in \operatorname{argmin}_{\theta \in \Theta} (R_n(\theta) + \kappa \|\theta\|_1),$$

avec $R_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left(\left\| \lambda_{\theta}^{(i)} \right\|_T^2 - \frac{2}{T} \sum_{j=1}^M \int_0^T \lambda_{\theta,j}^{(i)}(t) dN_j^{(i)}(t) \right)$. Nous omettons d'indicer l'estimateur $\hat{\theta}$ par la pénalisation κ car nous aurons par la suite un choix unique de κ et aucune autre conditions au préalable sur sa valeur.

• **Lemme III.4-1 : Valable sous les hypothèses (H1), (H2) et (H4).**

24

Avec $\hat{\theta}$ l'estimateur Lasso de pénalité $\kappa = \left(1 + \frac{2}{\alpha-1}\right) \frac{72\sqrt{2}\|h\|_{\infty}}{a(\gamma)^{3/2}T} \frac{\log(n)^2}{\sqrt{n}}$ où α provient de (H3) et $a(\gamma)$ de Corollaire III.3 – 2, l'inégalité oracle suivante est vérifiée :

$$\frac{2}{n} \sum_{i=1}^n \left\| \lambda_{\hat{\theta}}^{(i)} - \lambda_{\theta^*}^{(i)} \right\|_T^2 \lesssim \left(1 + \frac{2}{\alpha-1}\right) \frac{72\sqrt{2}\|h\|_{\infty}}{a(\gamma)^{3/2}T} \frac{\log(n)^2}{\sqrt{n}} \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)},$$

avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n}$.

Preuve (Lemme III.4 – 1) : On peut calculer le gradient de R_n , pour tout $j, j' \in \llbracket 1, M \rrbracket \times \llbracket 1, M \rrbracket$:

$$\frac{\partial R_n}{\partial \mu_j}(\theta) = \frac{2}{Tn} \sum_{i=1}^n \left(\int_0^T \lambda_{\theta,j}^{(i)}(t) dt - \int_0^T dN_j^{(i)}(t) \right)$$

et

$$\frac{\partial R_n}{\partial a_{jj'}}(\theta) = \frac{2}{Tn} \sum_{i=1}^n \left(\int_0^T H_{j'}^{(i)}(t) \lambda_{\theta,j}^{(i)}(t) dt - \int_0^T H_{j'}^{(i)}(t) dN_j^{(i)}(t) \right).$$

Ces expressions permettent alors de calculer

$$\begin{aligned} \left\langle \nabla R_n(\hat{\theta}) \mid \hat{\theta} - \theta^* \right\rangle_F &= \frac{2}{Tn} \sum_{i=1}^n \left(\sum_{j=1}^M \int_0^T \left(\lambda_{\hat{\theta},j}^{(i)}(t) dt - dN_j^{(i)}(t) \right) \left(\hat{\mu}_j - \mu_j^* + \sum_{j'=1}^M (\hat{a}_{jj'} - a_{jj'}^*) H_{j'}^{(i)}(t) \right) \right) \\ &= \frac{2}{Tn} \sum_{i=1}^n \left(\sum_{j=1}^M \int_0^T \left(\lambda_{\hat{\theta},j}^{(i)}(t) dt - dN_j^{(i)}(t) \right) (\lambda_{\hat{\theta},j}^{(i)}(t) - \lambda_{\theta^*,j}^{(i)}(t)) \right) \\ &= \frac{2}{n} \sum_{i=1}^n \left\| \lambda_{\hat{\theta}}^{(i)} - \lambda_{\theta^*}^{(i)} \right\|_T^2 - \frac{2}{nT} \sum_{i=1}^n \sum_{j=1}^M \int_0^T (\lambda_{\hat{\theta},j}^{(i)} - \lambda_{\theta^*,j}^{(i)}) dM_{\theta^*,j}^{(i)}(t). \end{aligned} \quad (25)$$

Puisque nous ne minimisons pas sur $\mathbb{R}^{M \times (M+1)}$ mais sur $(\mathbb{R}_+)^{M \times (M+1)}$, nous ne pouvons pas avoir les conditions classique sur le gradient et le sous gradient de la norme 1 comme dans la preuve du Théorème 6 de [2]. Nous nous réduisons aux conditions KKT qui nous octroient un multiplicateur $\Lambda \geq 0$ pour lequel

$$\Lambda \odot \hat{\theta} = 0 \text{ et } \nabla R_n(\hat{\theta}) + \kappa J - \Lambda = 0 \text{ où } J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(M+1) \times M} \text{ et } \odot \text{ est le produit terme à terme.}$$

Dès lors, on peut exprimer le produit scalaire de $\nabla R_n(\hat{\theta})$ avec $\hat{\theta} - \theta^*$ comme suit :

$$\left\langle -\nabla R_n(\hat{\theta}) \mid \hat{\theta} - \theta^* \right\rangle_F = \kappa \left\langle J \mid \hat{\theta} - \theta^* \right\rangle_F - \left\langle \Lambda \mid \hat{\theta} - \theta^* \right\rangle_F = \kappa \left\langle J \mid \hat{\theta} - \theta^* \right\rangle_F + \left\langle J \mid \theta^* \right\rangle_F \geq \kappa \left\langle J \mid \hat{\theta} - \theta^* \right\rangle_F = \kappa \left(\left\| \hat{\theta} \right\|_1 - \left\| \theta^* \right\|_1 \right),$$

puisque $\theta^* \geq 0$ d'après l'hypothèse (H1) et $\hat{\theta} \geq 0$ par définition. Or, on peut décomposer ces normes 1 selon les coordonnées qui sont sur le support de θ^* ou pas :

$$\begin{aligned} \left\| \hat{\theta} \right\|_1 - \left\| \theta^* \right\|_1 &= \sum_{(j,j') \in \text{supp}(\theta^*)^c} \hat{\theta}_{j,j'} + \sum_{(j,j') \in \text{supp}(\theta^*)} \left(\hat{\theta}_{j,j'} - \theta_{j,j'}^* \right) \\ &\geq \sum_{(j,j') \in \text{supp}(\theta^*)^c} \hat{\theta}_{j,j'} - \sum_{(j,j') \in \text{supp}(\theta^*)} \left| \hat{\theta}_{j,j'} - \theta_{j,j'}^* \right| = \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)^c} - \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)}. \end{aligned}$$

On déduit de ces inégalités la relation suivante :

$$\left\langle \nabla R_n(\hat{\theta}) \mid \hat{\theta} - \theta^* \right\rangle_F \leq \kappa \left(\left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)} - \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)^c} \right). \quad (26)$$

Or, on peut développer le terme suivant pour faire apparaître l'opérateur Z défini dans (19), où l'on rappelle que $\theta_j = (\mu_j, a_{j,1}, \dots, a_{j,M})$ et $H^{(i)}$ est défini en (8) :

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^M \int_0^T (\lambda_{\hat{\theta},j}^{(i)}(t) - \lambda_{\theta^*,j}^{(i)}(t)) dM_{\theta^*,j}^{(i)}(t) &= \sum_{i=1}^n \sum_{j=1}^M \int_0^T \lambda_{\hat{\theta}-\theta^*,j}^{(i)}(t) dM_{\theta^*,j}^{(i)}(t) \\ &= \sum_{\substack{1 \leq j \leq M \\ 0 \leq j' \leq M}} \sum_{i=1}^n \int_0^T (\hat{\theta} - \theta^*)_{j,j'} H_{j'}^{(i)}(t) dM_{\theta^*,j}^{(i)}(t) = \langle \hat{\theta} - \theta^* | Z \rangle_F. \end{aligned} \quad (27)$$

Nous avons déjà vu en (21) que nous pouvions contrôler la quantité $\frac{1}{n} \left(\max_{\substack{0 \leq j \leq M \\ 1 \leq j' \leq M}} |Z_{j,j'}| \right)$ avec grande probabilité. En effet, d'après le [Corollaire III.3 – 5](#), nous avons la majoration suivante :

$$\frac{1}{n} \max_{j,j'} |Z_{j,j'}| \underset{n \rightarrow +\infty}{\lesssim} \frac{4\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} \sqrt{n}}$$

avec probabilité $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n}$. Nous rapellons que la notation $\lesssim$ a été introduite dans la [Définition II.3 – 3](#). Ensuite, nous obtenons avec probabilité supérieure à $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n}$:

$$\left| \frac{1}{n} \langle \hat{\theta} - \theta^* | Z(T) \rangle \right| \leq \frac{1}{n} \left(\max_{\substack{0 \leq j \leq M \\ 1 \leq j' \leq M}} |Z_{j,j'}| \right) \|\hat{\theta} - \theta^*\|_1 \lesssim \frac{4\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} \sqrt{n}} \|\hat{\theta} - \theta^*\|_1.$$

On en déduit grâce à l'équation précédente et aux équations (25), (26) et (27), l'inégalité portant directement sur la norme $\|\cdot\|_T$ des intensités, moyennée sur les échantillons :

$$\frac{2}{n} \sum_{i=1}^n \left\| \lambda_{\hat{\theta}}^{(i)} - \lambda^{(i)} \right\|_T^2 \lesssim \left(\kappa + \frac{8\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} T \sqrt{n}} \right) \|\hat{\theta} - \theta^*\|_{1, \text{supp}(\theta^*)} + \left(\frac{8\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} T \sqrt{n}} - \kappa \right) \|\hat{\theta} - \theta^*\|_{1, \text{supp}(\theta^*)^c}.$$

Maintenant, considérons $\alpha > 1$ apparaissant dans l'hypothèse [\(H3\)](#). Le bon choix de coefficient de pénalisation κ qui suit permet d'obtenir une condition d'appartenance de $\hat{\theta} - \theta^*$ à un cône. En effet, notons :

$$\kappa := \left(1 + \frac{2}{\alpha - 1} \right) \frac{8\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} T \sqrt{n}},$$

ce choix permet de trouver la condition suivante sur $\hat{\theta} - \theta^*$ grâce à l'équation précédente :

$$\left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)^c} \leq \alpha \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)},$$

c'est-à-dire $\hat{\theta} - \theta^* \in \mathcal{C}(\text{supp}(\theta^*), \alpha)$ où les cônes $\mathcal{C}(S, \alpha)$ ont été définis au préalable dans l'hypothèse [\(H3\)](#). On peut aussi déduire des équations précédentes l'inégalité oracle suivante, vérifiée avec probabilité d'au moins $1 - \left(C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4} \right) \frac{1}{n}$:

$$\frac{2}{n} \sum_{i=1}^n \left\| \lambda_{\hat{\theta}}^{(i)} - \lambda_{\theta^*}^{(i)} \right\|_T^2 \lesssim \left(1 + \frac{2}{\alpha - 1} \right) \frac{8\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} T \sqrt{n}} \left\| \hat{\theta} - \theta^* \right\|_{1, \text{supp}(\theta^*)}. \quad (28)$$

□

On remarquera que rien n'empêche α de dépendre de n , le nombre d'observations, puisque α provient de l'hypothèse [\(H3\)](#). Or dans la littérature, ce coefficient est le plus souvent supposé comme étant constant par rapport à la taille des échantillons n .

5 - Vitesse de convergence

Cette section propose une preuve de la proposition suivante qui précise le taux de convergence de notre estimateur en norme L^2 . On remarquera que la borne est inversement proportionnelle à T mais ce n'est en réalité pas un résultat applicable en temps long puisque nous rappelons que $C(\gamma, T, m, s^*) \rightarrow \infty$ lorsque $T \rightarrow +\infty$. Pour rentrer dans les détails, nous pourrions tout de même faire tendre T vers l'infini de manière contrôlée. En effet, on a $C(\gamma, T, m, s^*) =$

$\exp\left(\frac{1-\gamma}{2(1+\gamma)}Tms^*\right)$ donc pour toujours avoir la propriété suivante avec grande probabilité, il suffirait d'avoir $mT \ll \log(n)$ et on pourrait obtenir une borne finale en grande probabilité de l'ordre de grandeur de $\log(n)^{1+\varepsilon}/\sqrt{n}$ pour tout $\varepsilon > 0$.

► **Proposition III.5-1 : Valable sous les hypothèses (H1), (H2), (H3) et (H4).**

25

L'inégalité asymptotique suivante est vérifiée avec probabilité d'au moins $1 - \left(2 + 2C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4}\right) \frac{1}{n}$:

$$\frac{1}{\sqrt{M}} \|\hat{\theta} - \theta^*\|_F = \sqrt{\frac{1}{M} \sum_{j=1}^M \|\hat{\theta}_j - \theta_j^*\|_2^2} \lesssim \left(1 + \frac{2}{\alpha - 1}\right) \frac{36\sqrt{2}\|h\|_\infty \sqrt{s^*} \log(n)^2}{a(\gamma)^{3/2} T \phi_0} \frac{1}{\sqrt{n}}.$$

Avant de passer à la preuve du résultat, remarquons que la vitesse de convergence obtenue est du même ordre de grandeur que celle obtenue dans le Lemme 3.3 de [19].

On rappelle que l'on note $\mathbb{H}_t$ la matrice de taille $n \times (M+1)$ dont les coefficients sont les $H_j^{(i)}(t)$, définis en (8). Cet opérateur aléatoire permet de relier les paramètres θ du processus de Hawkes aux intensités induites par θ , voir les équations (9). On rappelle que l'on a défini l'opérateur $\mathbb{H}$ en (10). Ces égalités vont permettre de relier directement les normes $\|\cdot\|_T$ des intensités avec les normes des paramètres inscrits dans θ . En effet,

$$\begin{aligned} \sum_{i=1}^n \|\lambda_\theta^{(i)}\|_T^2 &= \sum_{i=1}^n \frac{1}{T} \int_0^T \|\lambda_\theta^{(i)}(t)\|_2^2 dt = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \frac{1}{T} \int_0^T \lambda_{\theta,j}^{(i)}(t)^2 dt \\ &= \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \frac{1}{T} \int_0^T (\mathbb{H}_t \theta_j^\top)_i^2 dt = \sum_{j=1}^M \frac{1}{T} \int_0^T \theta_j \mathbb{H}_t^\top \mathbb{H}_t \theta_j^\top dt = \sum_{j=1}^M \|\mathbb{H}^{1/2} \theta_j^\top\|_2^2 = \|\mathbb{H}^{1/2} \theta^\top\|_F^2 \end{aligned} \quad (29)$$

où l'on rappelle qu'au vu de la définition de $\mathbb{H}$, il est clair que $\mathbb{H}$ est une matrice symétrique positive. Cette égalité montre directement le lien entre les propriétés de l'opérateur $\mathbb{H}$ par rapport aux normes moyennées $\|\cdot\|_T$ des intensités $\theta \mapsto \lambda_\theta$ dépendantes des échantillons.

Preuve (Proposition III.5 – 1) :

Nous allons enfin pouvoir utiliser l'hypothèse (H3) non pas sur $\mathbb{E}(\mathbb{H})$ mais sur $\mathbb{H}$ comme montré dans le Corollaire III.3 – 8, ce qui est une hypothèse plus faible que les hypothèses supplémentaires apparaissant dans [13] puisqu'elles ont une structure presque sûre. En effet, d'après l'équation (29) et (H3), nous avons l'inégalité suivante pour tout $\theta \in \mathcal{C}(S, \alpha)$:

$$\sum_{i=1}^n \|\lambda_\theta^{(i)}\|_T^2 = \|\mathbb{H}^{1/2} \theta^\top\|_F^2 \geq \left(n\phi_0 - \frac{16\sqrt{2}\|h\|_\infty^2}{a(\gamma)^2} \sqrt{n} \log(nM)^{5/2}\right) \|\theta\|_F^2, \quad (30)$$

avec probabilité d'au moins $1 - \frac{2+C(\gamma, T, m, s^*)}{n}$. Ainsi, l'inégalité obtenue depuis (28) devient, avec probabilité supérieure à $1 - \left(2 + 2C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4}\right) \frac{1}{n}$ d'après la Proposition II.3 – 4 :

$$2\phi_0 \|\hat{\theta} - \theta^*\|_F^2 \stackrel{(30)}{\lesssim} \frac{2}{n} \sum_{i=1}^n \|\lambda_{\hat{\theta}}^{(i)} - \lambda_{\theta^*}^{(i)}\|_T^2 \stackrel{(28)}{\lesssim} \left(1 + \frac{2}{\alpha - 1}\right) \frac{8\sqrt{2}\|h\|_\infty \log(n)^2}{a(\gamma)^{3/2} T} \|\hat{\theta} - \theta^*\|_{1, \text{supp}(\theta^*)}.$$

Or Cauchy-Schwarz permet d'établir l'inégalité $\|\hat{\theta} - \theta^*\|_{1, \text{supp}(\theta^*)} \leq \sqrt{(M+1)s^*} \|\hat{\theta} - \theta^*\|_{F, \text{supp}(\theta^*)}$ car nous sommes sous l'hypothèse (H2). Ainsi, puisque l'inégalité $\|\cdot\|_{F, S} \leq \|\cdot\|_F$ est évidente pour tout support S , nous obtenons la vitesse de convergence L^2 suivante de nos paramètres :

$$\frac{1}{\sqrt{M}} \|\hat{\theta} - \theta^*\|_F = \sqrt{\frac{1}{M} \sum_{j=1}^M \|\hat{\theta}_j - \theta_j^*\|_2^2} \lesssim \left(1 + \frac{2}{\alpha - 1}\right) \frac{4\sqrt{2}\|h\|_\infty \sqrt{s^*} \log(n)^2}{a(\gamma)^{3/2} T \phi_0} \frac{1}{\sqrt{n}},$$

$$\text{avec probabilité supérieure à } 1 - \left(2 + 2C(\gamma, T, m, s^*) + \frac{\pi^4}{18 \log(2)^4}\right) \frac{1}{n}.$$

On remarquera que la transformation de $M+1$ en M ne pose pas de problème ici car nous sommes toujours dans des inégalités asymptotiques et nous sommes toujours dans le régime de la grande dimension, c'est-à-dire $M \rightarrow +\infty$. On retrouve une dépendance en la racine carrée du support. $\square$

IV – Unité

1 - Résultats

Dans cette partie, nous allons étudier l'unicité de la solution au problème Lasso introduit dans (16). Pour cela, nous allons avoir besoin d'ajouter une hypothèse de régularité sur la fonction noyau h au voisinage de 0.

⊙ **Hypothèse (H5) :**

26

Sur un voisinage de 0 qui n'inclut pas 0, le noyau h est de classe $\mathcal{C}^1$ et strictement monotone.

Récrivons la fonction de perte quadratique $R_n(\theta)$ en fonction des opérateurs $\mathbb{H}_t$. Pour tout $\theta \in \mathbb{R}^{M \times (M+1)}$, on peut observer depuis la relation (29) :

$$\sum_{i=1}^n \left\| \lambda_{\theta}^{(i)} \right\|_T^2 = \sum_{j=1}^M \left\| \mathbb{H}^{1/2} \theta_j \right\|_2^2 = \left\| \mathbb{H}^{1/2} \theta^\top \right\|_F^2.$$

De plus, remarquons que si on écrit θ en fonction de ses colonnes : $\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_M \end{pmatrix} \in \mathbb{R}^{M \times (M+1)}$, on trouve pour tout $t \geq 0$,

$$\mathbb{H}_t \theta^\top = (\mathbb{H}_t \theta_1^\top \quad \cdots \quad \mathbb{H}_t \theta_M^\top) = \left(\left(\lambda_{\theta,1}^{(i)}(t) \right)_{1 \leq i \leq n} \quad \cdots \quad \left(\lambda_{\theta,M}^{(i)}(t) \right)_{1 \leq i \leq n} \right) = \left(\lambda_{\theta,j}^{(i)}(t) \right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}}.$$

De quoi on peut déduire l'égalité suivante pour tous $i \in \llbracket 1, n \rrbracket$ et $j \in \llbracket 1, M \rrbracket$:

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} \int_0^T \lambda_{\theta,j}^{(i)}(t) dN_j^{(i)}(t) = \int_0^T \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} (\mathbb{H}_t \theta^\top)_{i,j} (dN(t))_{i,j} = \int_0^T \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M \\ 0 \leq k \leq M}} \theta_{j,k} \int_0^T (\mathbb{H}_t)_{i,k} dN_j^{(i)}(t) = \left\langle \theta^\top \mid \int_0^T \mathbb{H}_t^\top dN(t) \right\rangle_F,$$

où l'on définit $\int_0^T \mathbb{H}_t dN(t)$ comme étant la matrice aléatoire composée des coefficients $\int_0^T \sum_{i=1}^n (\mathbb{H}_t^\top)_{ki} dN_j^{(i)}(t)$ pour $k \in \llbracket 0, M \rrbracket$ et $j \in \llbracket 1, M \rrbracket$. On peut alors enfin trouver l'expression suivante pour le risque $R_n(\theta)$:

$$\begin{aligned} R_n(\theta) &= \frac{1}{n} \sum_{i=1}^n \left(\left\| \lambda_{\theta}^{(i)} \right\|_T^2 - \frac{2}{T} \sum_{j=1}^M \int_0^T \lambda_{\theta,j}^{(i)}(t) dN_j^{(i)}(t) \right) \\ &= \frac{1}{n} \left\| \mathbb{H}^{1/2} \theta^\top \right\|_F^2 - \frac{2}{nT} \left\langle \theta^\top \mid \int_0^T \mathbb{H}_t^\top dN(t) \right\rangle_F. \end{aligned}$$

Pour pouvoir avancer dans cette expression en l'écrivant sous une même norme quadratique, nous allons utiliser le Lemme suivant qui sera prouvé ultérieurement.

• **Lemme IV.1-1 : Valable sous l'hypothèse (H5).**

27

Si $\min_{\substack{1 \leq j \leq M \\ 1 \leq i \leq n}} N_j^{(i)}(T) \geq 1$, alors la matrice $\mathbb{H}$ est symétrique définie positive presque sûrement.

La condition énoncée dans ce lemme est assez contraignante car elle impose à toutes les observations d'avoir des événements selon chaque dimension. Cette hypothèse intervient dans la preuve de ces résultats lorsque l'on s'intéresse au caractère continu des lois des temps d'activation des $(N^{(i)})_{1 \leq i \leq n}$. En effet, les quantités apparaissant dans la matrice $\mathbb{H}_t$ sont des sommes de $h(t - \tau)$ où τ parcourt l'ensemble des temps d'activation. Si ce dernier ensemble est vide avec probabilité strictement positive alors certaines coordonnées de $\mathbb{H}_t$ seraient égales à 0 sur ces événements non négligeables, ce qui ajouterait un atome à la loi jointe de $\mathbb{H}_t$ et impliquerait sa non continuité absolue par rapport à la mesure de Lebesgue.

Nous pouvons alors utiliser le [Lemme IV.1 – 1](#), tout en supposant qu’aucun des $N_j^{(i)}$ ne soit nul, pour écrire la relation suivante vérifiée par $R_n(\theta)$:

$$\begin{aligned} R_n(\theta) &= \frac{1}{n} \left\| \mathbb{H}^{1/2} \theta^\top \right\|_F^2 - \frac{2}{nT} \left\langle \theta^\top \mid \int_0^T \mathbb{H}_t^\top dN(t) \right\rangle_F \\ &= \frac{1}{n} \left(\left\| \mathbb{H}^{1/2} \theta^\top - \mathbb{H}^{-1/2} \frac{1}{T} \int_0^T \mathbb{H}_t^\top dN(t) \right\|_F^2 + \left\| \mathbb{H}^{-1/2} \frac{1}{T} \int_0^T \mathbb{H}_t^\top dN(t) \right\|_F^2 \right). \end{aligned}$$

Ce qui donne la relation suivante sur l’estimateur $\hat{\theta}$:

$$\hat{\theta} \in \underset{\theta \in \mathbb{R}^{M \times (M+1)}}{\operatorname{argmin}} \left(\frac{1}{n} \left\| \mathbb{H}^{1/2} \theta^\top - \mathbb{H}^{-1/2} \frac{1}{T} \int_0^T \mathbb{H}_t^\top dN(t) \right\|_F^2 + \kappa \|\theta\|_1 \right).$$

Nous allons pouvoir déduire immédiatement de cette relation l’unicité de $\hat{\theta}$ puisque la la fonction

$$\theta \mapsto \left\| \mathbb{H}^{1/2} \theta^\top - \mathbb{H}^{-1/2} \frac{1}{T} \int_0^T \mathbb{H}_t^\top dN(t) \right\|_F^2$$

est fortement convexe de paramètre $2\lambda_{\min}(\mathbb{H}) > 0$ d’après le [Lemme IV.1 – 1](#). Ainsi, puisque $\theta \mapsto \|\theta\|_1$ est convexe, on obtient bien que $\hat{\theta}$ minimise une fonction fortement convexe et donc est unique.

Notre cas est au final plus simple à obtenir que dans le cas d’une matrice de design de taille $n \times (M + 1)$, nous nous sommes ramené à une matrice carré de taille $M + 1$, qui est l’opérateur $\mathbb{H}$. Nous aurions pu, malgré l’utilisation d’une matrice de paramètres θ , utiliser les résultats de [\[32\]](#) car les conditions KKT s’appliquent sur nos colonnes et donc au final sur la matrice θ entière.

2 - Preuve du Lemme

Preuve (Lemme IV.1 – 1) : Pour montrer l’inversibilité de la matrice

$$\mathbb{H} := \frac{1}{T} \int_0^T \mathbb{H}_t^\top \mathbb{H}_t dt,$$

nous allons remarquer que pour tout $t \geq 0$, les matrices intégrandes $\mathbb{H}_t^\top \mathbb{H}_t$ sont symétriques positives, pour montrer que $\mathbb{H}$ est définie positive il suffit alors de montrer qu’il existe $A \subseteq [0, T]$ non Lebesgue-négligeable tel que pour tout $t \in A$, $\mathbb{H}_t$ soit de rang plein. Pour cela, on remarque que la matrice $\mathbb{H}_t$ peut s’écrire sous la forme

$$\mathbb{H}_t = \begin{pmatrix} 1 \\ \vdots \\ H(t) \\ 1 \end{pmatrix} \in \mathbb{R}^{n \times (M+1)} \quad \text{avec} \quad H(t) = \left(\int_0^{t^-} h(t-s) dN_j^{(i)}(s) \right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}}.$$

Nous allons commencer par montrer que $H(t)$ est de rang plein pour t assez petit, puis nous en déduirons l’inversibilité de $\mathbb{H}$ en montrant que la première colonne composée uniquement de 1 est linéairement indépendante des autres. D’après l’hypothèse **(H5)**, nous disposons de $\eta > 0$ tel que h soit de classe $\mathcal{C}^1$ sur $]0, \eta[$, et nous pouvons choisir η assez petit pour que h' ne s’annule pas sur $]0, \eta[$. Notons alors Φ l’application suivante, où $\mathbb{R}^{(\mathbb{N})}$ est l’ensemble des suites réelles finies :

$$\Phi : \tau \in \mathbb{R}^{(\mathbb{N})} \mapsto \sum_{t \in \tau} h(\eta - t) \in \mathbb{R},$$

et remarquons que si l’on note $\mathcal{T}_j^{(i)}$ l’ensemble des temps de sauts du processus de Hawkes selon la dimension j du i -ième échantillon pour des temps strictement inférieurs à η , nous avons presque sûrement $|\mathcal{T}_j^{(i)}| < +\infty$ puisque $N_j^{(i)}(\eta) \leq N_j^{(i)}(T)$ a un moment exponentiel fini, donc est en particulier fini presque sûrement. Ainsi, nous obtenons la relation suivante, ce pourquoi nous avons introduit Φ :

$$\Phi(\mathcal{T}_j^{(i)}) = H_j^{(i)}(\eta) = \int_0^{t^-} h(\eta - s) dN_j^{(i)}.$$

Or, pour tous $K \in \mathbb{N}$ et $(T_k)_{1 \leq k \leq K} \in \mathbb{R}^K$, le gradient de $\Phi_K := \Phi|_{\mathbb{R}^K}$ s’écrit comme suit :

$$\nabla \Phi_K \left((T_k)_{1 \leq k \leq K} \right) = (-h'(\eta - T_k))_{1 \leq k \leq K}.$$

Plaçons nous sur l'évènement $\mathcal{U} := \left(\min_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}} N_j^{(i)}(\eta) \geq 1 \right)$. Dès lors, nous avons pour tout i, j , $\nabla \Phi_K \left(\mathcal{T}_j^{(i)} \right) \neq 0$ car h' garde le même signe stricte sur $]0, \eta[$. Ainsi, puisque le gradient de Φ_K ne s'annule donc pas sur l'ouvert

$$U_K := \bigcup_{l=1}^K \left\{ (T_k)_{1 \leq k \leq K} \mid T_l < \eta \right\},$$

on peut utiliser la [Proposition VII.3 – 1](#) pour en déduire que pour tout $i \in \llbracket 1, n \rrbracket$, la ligne $\left(H_j^{(i)}(\eta) \right)_{1 \leq j \leq M}$ ait une loi qui soit absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^M$. En effet, considérons la fonction tensorisée Ψ de Φ suivante :

$$\Psi : (\tau_j)_{1 \leq j \leq M} \in \left(\mathbb{R}^{(\mathbb{N})} \right)^M \longmapsto (\Phi(\tau_j))_{1 \leq j \leq M} \in \mathbb{R}^M,$$

et notons pour tous $K = (K_1, \dots, K_M) \in (\mathbb{N}^*)^M$, la fonction restreinte $\Psi_K := \Psi|_{\prod_{j=1}^M \mathbb{R}^{K_j}}$. Calculons alors la différentielle de Ψ_K en fonction des gradients déjà connus des Φ_{K_j} .

$$d\Psi \left((T_k^j)_{\substack{1 \leq j \leq M \\ 1 \leq k_j \leq K_j}} \right) = \begin{pmatrix} \nabla \Phi \left((T_{k_1}^1)_{1 \leq k_1 \leq K_1} \right)^\top & 0 & \dots & 0 \\ 0 & \nabla \left(\Phi(T_{k_2}^2)_{1 \leq k_2 \leq K_2} \right)^\top & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \nabla \left(\Phi(T_{k_M}^M)_{1 \leq k_M \leq K_M} \right)^\top \end{pmatrix} \in \mathbb{R}^{M \times \sum_{j=1}^M K_j},$$

qui est de rang plein si les gradients de Φ y apparaissant sont tous non nuls, ce qui est le cas sur $\prod_{j=1}^M U_{K_j}$ puisque nous avons $K_j \geq 1$ pour tout $j \in \llbracket 1, M \rrbracket$ ici. Nous nous retrouvons alors avec l'identité désirée qui relie les temps d'activation de notre processus avec les variables $H_j^{(i)}$:

$$\left(H_j^{(i)}(\eta) \right)_{1 \leq j \leq M} = \Psi \left(\left(\mathcal{T}_j^{(i)} \right)_{1 \leq j \leq M} \right).$$

Or, nous savons que $\mathcal{T}^{(i)} := \left(\mathcal{T}_j^{(i)} \right)_{1 \leq j \leq M}$ est presque sûrement fini pour tout i et, la loi de $\mathcal{T}^{(i)}$ sachant les $N_j^{(i)}$ pour $j \in \llbracket 1, M \rrbracket$ ont des lois continues par rapport aux mesures de Lebesgues sur $\mathbb{R}^{N_j^{(i)}}$, d'après la [Proposition VII.3 – 4](#). Ainsi, puisque les restrictions de Ψ aux $\prod_{j=1}^M \mathbb{R}^{K_j}$ ont des gradients non-nuls pour $K \in (\mathbb{N}^*)^M$, le [Corollaire VII.3 – 2](#) montre alors que $\left(H_j^{(i)}(\eta) \right)_{1 \leq j \leq M}$ a une loi continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^M$. Ainsi, puisque les $\left(H_j^{(i)}(\eta) \right)_{1 \leq j \leq M}$ sont mutuellement indépendants pour $i \in \llbracket 1, n \rrbracket$, on observe les relations suivantes qui prouvent que $H(\eta)$ admet une loi absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n \times M}$:

$$\mathbb{P}_{H(\eta)} = \mathbb{P}_{\left(H_j^{(i)}(\eta) \right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq M}}} = \bigotimes_{i=1}^n \mathbb{P}_{\left(H_j^{(i)}(\eta) \right)_{1 \leq j \leq M}} \ll \bigotimes_{i=1}^n \text{Leb}_M = \text{Leb}_{n \times M}.$$

De cette propriété, on en déduit que $H(\eta)$ est de rang plein d'après le [Lemme VII.3 – 3](#). Pour en déduire que

$$\mathbb{H}_\eta = \begin{pmatrix} 1 \\ \vdots \\ H(\eta) \\ 1 \end{pmatrix} \in \mathbb{R}^{n \times (M+1)}$$

est inversible il suffit alors de montrer que le vecteur $(1, \dots, 1)$ est linéairement indépendant des colonnes de $H(\eta)$ presque sûrement. En effet, supposons le contraire, on dispose alors de $(\kappa_j)_{1 \leq j \leq M} \neq 0$ tel que pour tout $i \in \llbracket 1, n \rrbracket$,

$$1 = \sum_{j=1}^M \kappa_j H_j^{(i)}(\eta) = \sum_{j=1}^M \kappa_j \Phi \left(\mathcal{T}_j^{(i)} \right).$$

On procède alors comme précédemment, on considère l'application Υ définie comme suit :

$$\Upsilon : (\tau_j)_{1 \leq j \leq M} \in \left(\mathbb{R}^{(\mathbb{N})}\right)^M \mapsto \sum_{j=1}^M \kappa_j \Phi(\tau_j) \in \mathbb{R},$$

et on calcul son gradient restreint pour $K \in (\mathbb{N}^*)^M$ avec $\Upsilon_K := \Upsilon \Big|_{\prod_{j=1}^M \mathbb{R}^{\kappa_j}}$:

$$\nabla \Upsilon_K(\tau) = (\kappa_j \nabla \Phi_{K_j}(\tau_j))_{1 \leq j \leq M}.$$

Or, on a supposé que $\kappa \neq 0$, donc puisque $\nabla \Phi_{K_j}(\tau_j)$ est toujours non nul d'après l'hypothèse **(H5)** sur $\prod_{j=1}^M U_{K_j}$, on sait que le gradient $\nabla \Upsilon_K$ ne s'annule donc jamais sur $\prod_{j=1}^M U_{K_j}$, où il y est en plus continue. Ainsi, on peut appliquer le **Corollaire VII.3 – 2** pour en déduire que la loi de

$$\Upsilon(\mathcal{T}^{(i)}) = \sum_{j=1}^M \kappa_j H_j^{(i)}(\eta)$$

est continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Par indépendance, on en déduit comme précédemment que pour $i_1 \neq i_2 \in \llbracket 1, n \rrbracket$,

$$\mathbb{P}\left(\sum_{j=1}^M \kappa_j H_j^{(i_1)}(\eta), \sum_{j=1}^M \kappa_j H_j^{(i_2)}(\eta)\right) \ll \text{Leb}_2.$$

Or, l'ensemble $\{(x, y) \in \mathbb{R}^2 \mid x = y\}$ est de mesure nulle pour Leb_2 donc l'est aussi par absolue continuité pour la loi de $\left(\sum_{j=1}^M \kappa_j H_j^{(i_1)}(\eta), \sum_{j=1}^M \kappa_j H_j^{(i_2)}(\eta)\right)$, ce que l'on peut traduire par le fait que l'évènement

$$\sum_{j=1}^M \kappa_j H_j^{(i_1)}(\eta) = \sum_{j=1}^M \kappa_j H_j^{(i_2)}(\eta)$$

est de probabilité nulle et donc $\kappa = 0$ presque sûrement, ce qui permet de conclure que $(1, \dots, 1)$ est bien presque sûrement linéairement indépendant de $H(\eta)$ et donc que presque sûrement, la matrice $\mathbb{H}_\eta$ est de rang plein.

Pour conclure, il suffit de remarquer que le raisonnement précédent reste vrai pour tout autre $0 < t \leq \eta$ puisque nous avons initialement pris η quelconque assez petit, et donc, pour tout $t \in]0, \eta[$, la matrice $\mathbb{H}_t$ est presque sûrement de rang plein. Il va nous rester à montrer que l'on peut intervertir le presque sûrement avec le quantificateur en t . Or, l'hypothèse **(H5)** montre en particulier que h est mesurable sur $]0, \eta[$ et donc, pour tout $i, j \in \llbracket 1, n \rrbracket \times \llbracket 1, M \rrbracket$, l'application

$$(t, \omega) \mapsto H_{i,j}(t)(\omega) = \sum_{\substack{\tau_j^{(i)} \in \mathcal{T}_j^{(i)}(\omega) \\ \tau_j^{(i)} < t}} h(t - \tau_j^{(i)})$$

est mesurable pour la tribu produit et donc l'application $(t, \omega) \mapsto \det(H(t)(\omega))$ l'est aussi par continuité du déterminant. On en déduit, d'après Fubini, par mesurabilité et positivité les égalités suivantes :

$$\mathbb{E}\left(\int_0^\eta \mathbb{1}_{\det(H(t))=0} dt\right) = \int_0^\eta \mathbb{P}(\det(H(t))=0) dt = 0,$$

ce qui implique que presque sûrement, pour presque tout $t \in]0, \eta[$, la matrice $H(t)$ est inversible. Cette propriété montre finalement que, presque sûrement, la matrice

$$\mathbb{H} = \frac{1}{T} \int_0^T \mathbb{H}_t^\top \mathbb{H}_t dt$$

est symétrique définie positive. □

V – Données synthétiques

1 - Exemples choisis

Pour illustrer notre approche, nous allons d’abord modéliser un processus de Hawkes de la même forme que ceux introduits dans la partie 3, avec des paramètres connus et un noyau h qui sera exponentiel, c’est-à-dire que pour tout $t \geq 0$, $h(t) = \beta e^{-\beta t}$ pour un certain $\beta > 0$. Pour ce faire, nous utiliserons le package Python **Sparklen** introduit dans [20]. Concernant la génération des données, nous considérons un algorithme utilisant la représentation en clusters du processus de Hawkes [27], qui peut être vu comme un processus de Poisson à clusters grâce à sa structure de branchement. Nous considérons dans nos simulations les choix $\beta = 3$ et $T = 5$.

Nous allons introduire plusieurs structures de paramètres. La première sera celle d’une matrice d’interactions A^* par blocs uniformes, c’est-à-dire :

$$A_b^* := \begin{pmatrix} a_1 J_{k_1} & 0 & \cdots & 0 \\ 0 & a_2 J_{k_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_p J_{k_p} \end{pmatrix} \quad \text{où} \quad J_k = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k},$$

avec $M = \sum_{j=1}^q k_j$ et $a_k \in [0, 1/k[$ pour assurer $\rho(A^*) = \|A^*\|_{\infty \rightarrow \infty} < 1$. Cette structure en blocs de A^* modélise les interactions entre groupes d’individus tout en assurant une auto-excitation des composants. Nous considérons un vecteur de baseline associé qui aura exactement une valeur non-nulle par coordonnées associées aux blocs J_{k_l} pour $l \in \llbracket 1, q \rrbracket$. Cette structure admet un paramètre de parcimonie valant $s_b^* = q \vee \max_{1 \leq l \leq q} k_l$. Nous noterons par la suite θ_b cette famille d’exemples.

Nous remarquons que les coordonnées associées à chaque bloc J_{k_l} sont indépendantes selon l puisque la matrice d’adjacence A_b^* est celle d’une union disjointe de p graphes complets. C’est pourquoi nous allons étudier l’exemple suivant. Nous considérons un certain nombre $K > 0$ de graphes aléatoires $(G_k)_{1 \leq k \leq K}$ indépendant de loi respective celle d’un graphe d’Erdos-Renyi de paramètres (M_k, p) avec $(M_k)_{1 \leq k \leq K}$ des entiers relatifs non nuls de somme M . On considérera alors l’union disjointe de ces graphes

$$G = \bigsqcup_{k=1}^K G_k$$

dont les sommets formeront l’ensemble $\llbracket 1, M \rrbracket$ de manière croissante en k , c’est-à-dire que pour tout $k < l$, tout $(j, j') \in G_k \times G_l$, on a $j < j'$. Nous noterons par la suite θ_G cette famille d’exemples. Nous tenons à noter que cet exemple inclut le cas d’un graphe d’Erdos-Renyi classique puisqu’il suffira de choisir le paramètre $K = 1$.

2 - Détails sur l’optimisation

Nous utiliserons le package **Sparklen** pour approcher l’estimateur Lasso que l’on a introduit en (16). Dans notre cadre il est d’abord préférable d’approcher la constante κ qui serait optimale pour l’estimateur associé. Pour cela, nous utiliserons le critère de sélection EBIC introduit par [9] de paramètre $\gamma = 1$ pour les exemples de type θ_b et nous utiliserons la procédure de validation croisée développée dans [1] pour les exemples de type θ_G . Ces deux méthodes ont déjà été implémentées dans le package **Sparklen**. Ce dernier procédera à une sorte de descente de gradient pour obtenir une approximation d’une solution au problème (16), plus précisément, l’algorithme se basera sur une descente de gradient proximale. La relation de récurrence suivie par le paramètre $(\theta_k)_{k \geq 0}$ de la descente de gradient sera la suivante :

$$\theta_{k+1} = \text{prox}_{\eta\kappa, \|\cdot\|_1} (\theta_k - \eta \nabla R_n(\theta_k)),$$

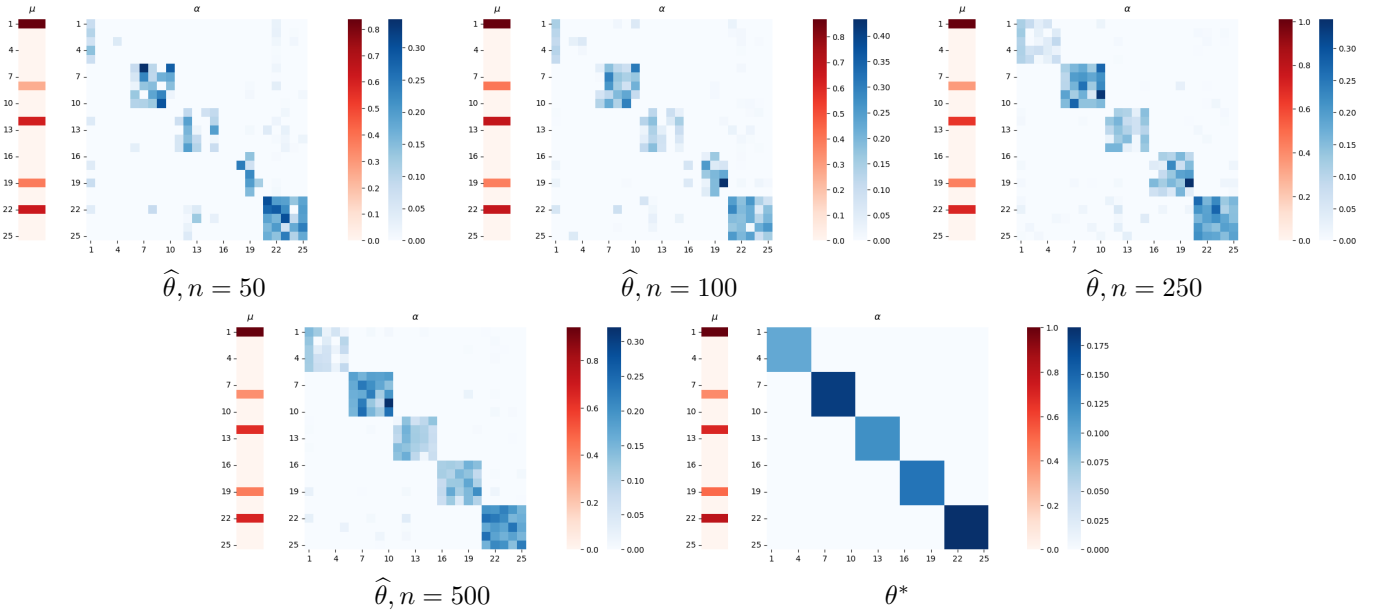
où $\text{prox}_{\eta\kappa, \|\cdot\|_1}(\cdot) = \text{sgn}(\cdot) \max(\|\cdot\|_1 - \eta\kappa, 0)$ est l’opérateur proximal associé au Lasso. **Sparklen** permet aussi d’utiliser un contraste différent du contraste quadratique définit en (15), plus précisément d’utiliser la *log-vraisemblance négative* définie comme suit :

$$\forall \theta \in \Theta, L_n(\theta) := \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}} \left(\int_0^T \lambda_{\theta,j}^{(i)}(t) dt - \int_0^T \log \left(\lambda_{\theta,j}^{(i)}(t) \right) dN_j^{(i)}(t) \right), \quad (31)$$

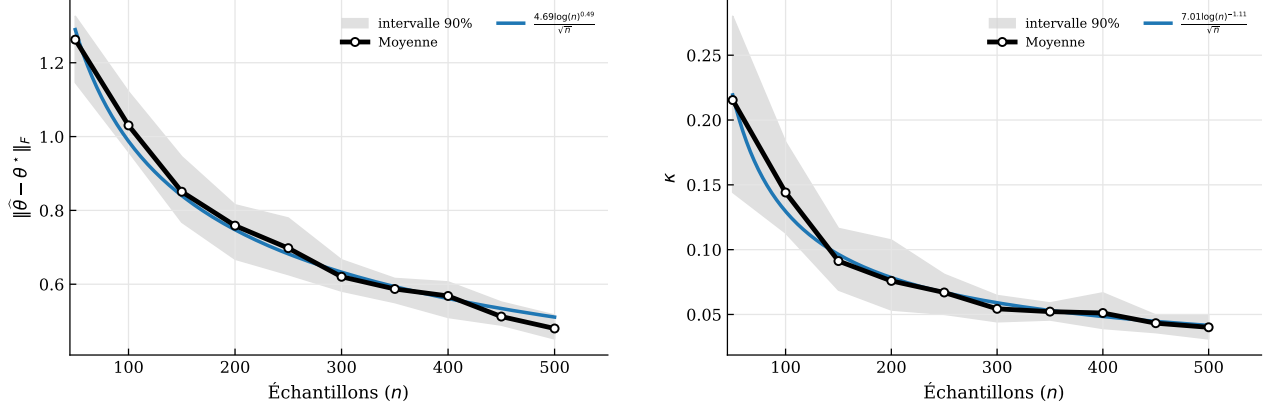
qu’il faudra alors minimiser selon $\theta \in \Theta$.

3 - Résultats expérimentaux

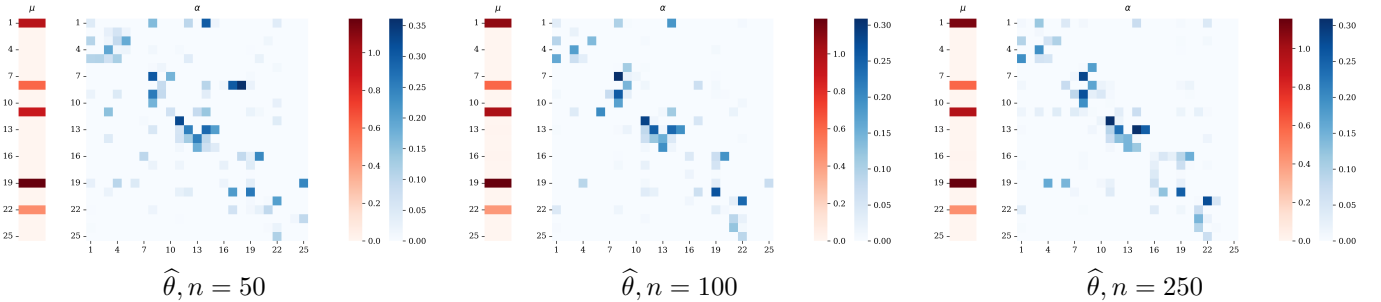
Pour le premier exemple θ_b , nous obtenons les estimateurs suivants :

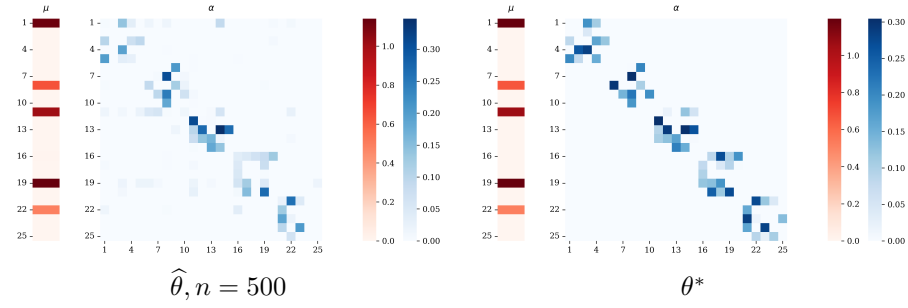


De ces estimateurs, nous pouvons calculer l'écart en norme L^2 avec les paramètres réels. Nous calculons la distance selon le nombre de données utilisées lors du calcul de l'estimateur, on observe les figures ci-dessous après 10 estimations indépendantes. Nous remarquerons que nous sommes bien dans un cadre de grande dimension puisque n va de 10 à 500 et $M = 25$. On trouve donc bien que le nombre de paramètres à estimer vérifie $M(M+1) > n$. Nous obtenons les graphes suivants après 10 répétitions.

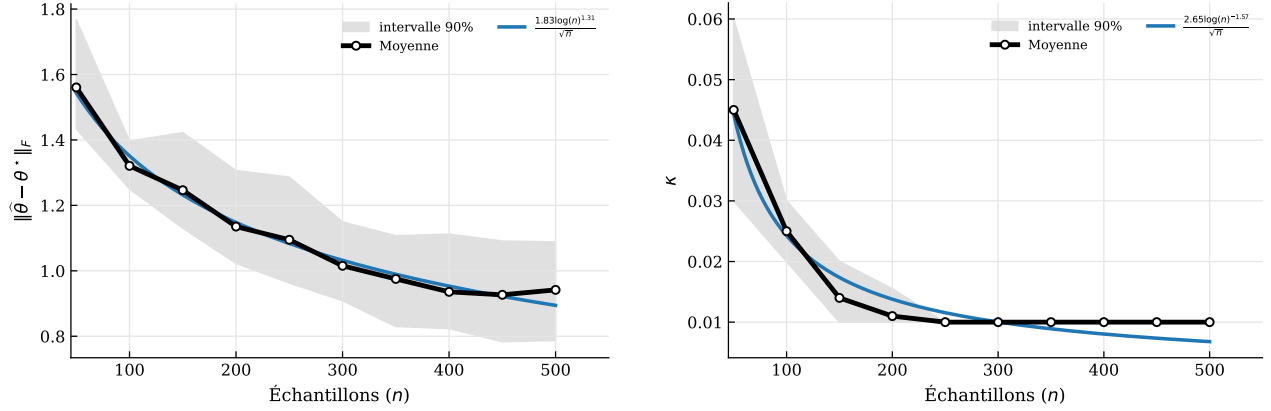


Pour la seconde famille d'exemple θ_G avec les paramètres $K = 5$, $p = 0.4$ et $k = (5, 5, 5, 5, 5)$, on trouve les estimateurs suivants :

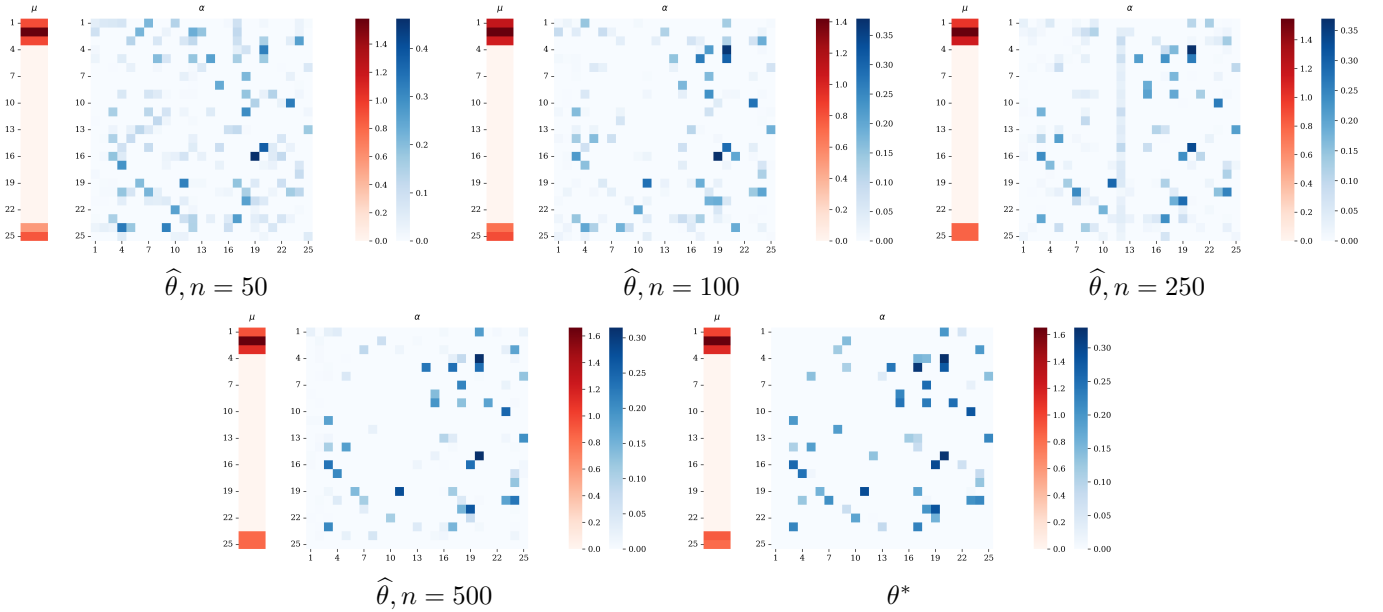




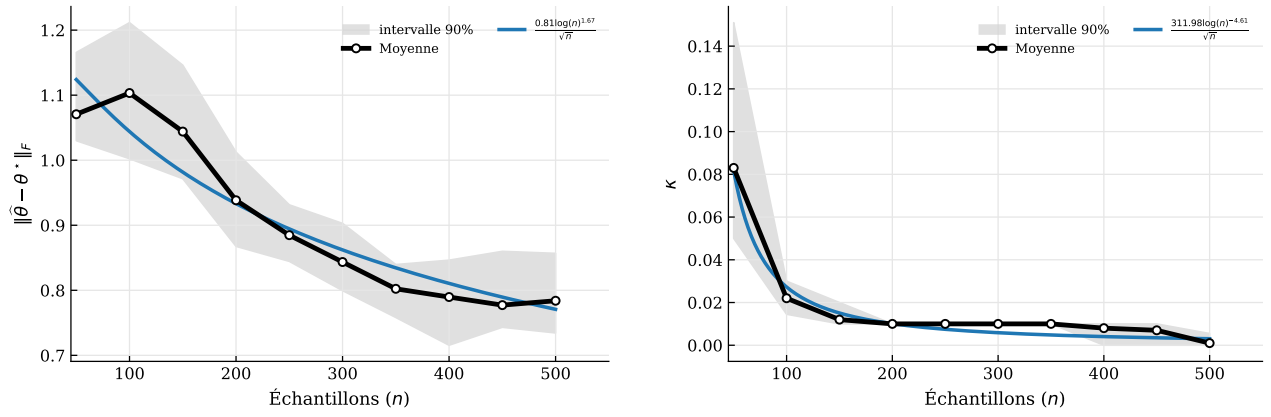
On peut de la même manière calculer les risques L^2 et les pénalités Lasso utilisées lors de l'optimisation. Nous obtenons les graphes suivants après 10 répétitions.



Dans la même famille θ_G avec les paramètres $K = 1$, $p = 0.6$ et $k = (25)$, on trouve les estimateurs suivants :



qui donnent lieu aux distances L^2 suivantes après 10 répétitions :



Dans les trois exemples précédents nous avons pu faire 10 répétitions mais il faudrait idéalement en faire un plus grand nombre. Malheureusement la machine utilisée pour faire les calculs n'est pas très puissante et ces 10 répétitions ont pris une dizaine d'heures pour chaque exemple.

Maintenant nous allons comparer les estimateurs obtenus avec un contraste quadratique et avec un modèle ayant une architecture de transformer que l'on peut retrouver dans l'article [25]. Or, ce dernier ne renvoie pas exactement de paramètres $\hat{\theta}$ comme le fait notre modèle avec le package **Sparklen**, nous allons donc devoir introduire une métrique pour comparer ce transformer avec nos méthodes. En effet, ce modèle de transformer permet d'approcher l'intensité du processus λ . Pour étudier son efficacité nous allons donc évaluer les intensités que le modèle renvoie à différentes coordonnées temporelles, que l'on notera $\mathcal{T}$ et qui sera l'ensemble $\{0.1, 0.2, \dots, 4.9\}$. Ainsi, nous pourrions considérer la distance empirique suivante :

$$\Delta_{\mathcal{T}}(\hat{\theta}, \theta) = \sqrt{\frac{1}{k|\mathcal{T}|} \sum_{i=1}^k \sum_{t \in \mathcal{T}^{(i)}} \left\| \lambda_{\hat{\theta}}^{(i)}(t) - \lambda_{\theta}^{(i)}(t) \right\|_2^2},$$

où i fait référence au i -ième échantillon de test. Nous obtenons le tableau suivant avec $T = 5$, $k = 500$ et $\mathcal{T}$ choisi précédemment.

Nous avons entraîné nos modèles sur une base de donnée constituée de processus de Hawkes de dimension 50 suivant l'exemple θ_b avec $n = 1000$ répétitions. Nous notons $\hat{\theta}_{\text{Lasso}}$ l'estimateur Lasso introduit dans le rapport et calculé à partir des données précédentes et $\hat{\theta}_{\text{Transformer}}^1$ et $\hat{\theta}_{\text{Transformer}}^2$ les estimateurs obtenus avec l'algorithme de [25] avec 5 epochs et avec comme paramètres respectifs (dont la description peut se trouver dans [25]) :

	$d_{\text{modèle}}$	$d_{\text{cachée}}$	Couches	Têtes	d_k	d_v	Nombre de paramètres
$\hat{\theta}_{\text{Lasso}}$	-	-	-	-	-	-	2550
$\hat{\theta}_{\text{Transformer}}^1$	4	8	1	1	4	8	1673
$\hat{\theta}_{\text{Transformer}}^2$	8	16	2	2	8	16	3705

Tableau 1 : Paramètres des transformers considérés

Nous obtenons alors les résultats suivant selon la métrique $\Delta_{\mathcal{T}}$:

θ	$\hat{\theta}_{\text{Lasso}}$	$\hat{\theta}_{\text{Transformer}}^1$	$\hat{\theta}_{\text{Transformer}}^2$
$\Delta_{\mathcal{T}}(\theta, \theta^*)$	0.354	2.624	2.206

Tableau 2 : Score obtenus par les différents modèles

Nous observons que notre estimation performe beaucoup mieux que qu'un transformer dans un régime de haute dimension.

VI – Conclusion et ouverture

Les résultats de convergence obtenus au cours de ce rapport sont classiques dans le contexte d'estimations Lasso de processus stochastiques en grande dimension et sont attendus, cependant les hypothèses considérées sont faibles par rapport aux jeux d'hypothèses présentées dans la littérature. En effet, l'étude de nos estimateurs sous une métrique L^2 nous a permis de nous rapprocher d'une structure matricielle pour les représenter aux côtés de nos données. De nombreuses inégalités de concentration ont été utilisées, et notamment une concentration en norme d'opérateur L^2 , caractéristique à notre étude, a permis d'affaiblir l'hypothèse classique des valeurs propres restreintes en une hypothèse déterministe. Cette hypothèse est cruciale dans notre étude puisqu'en plus de permettre de passer d'une inégalité oracle à une vitesse de convergence des paramètres estimés, elle permet d'assurer l'identifiabilité du modèle statistique. Sous ce point de vue matriciel, nous avons pu étudier l'inversibilité de la matrice $\mathbb{H}$ de par l'étude de la stabilité du caractère continu des lois de probabilités des temps de sauts de processus de Hawkes multivariés par des fonctions de classe C^1 dont la différentielle est de rang plein. Cette inversibilité implique immédiatement l'unicité de l'estimateur Lasso introduit dans la présente note.

La condition présentée ici pour l'unicité de l'estimateur pourrait être revue et probablement améliorée. Des algorithmes plus efficaces de sélection de modèle (choix du paramètre de pénalisation Lasso κ) pourrait être choisi pour diminuer le temps de calcul et améliorer les résultats expérimentaux. Une étude plus approfondie sur les dépendances des résultats sur les hypothèses d'horizon temporel T fini ou sur la borne m uniforme sur le vecteur de baseline pourraient être menées pour généraliser les résultats dans des contextes où ces valeurs seraient négligeables devant des ordres de grandeurs du type $\log(n)$.

VII – Annexe

1 - Moyenne glissante

On prouve dans cette section un lemme utile dans la preuve de la [Proposition III.1 – 2](#).

• Lemme VII.1-1 :

28

Soit $h \geq 0$ et d'intégrale 1 sur $[0, T]$. Alors, pour $\delta \in [0, T/2]$ assez grand, une quantité non Lebesgue-négligeable de $u \in [0, \delta]$ vérifient

$$\int_u^{\delta+u} h > 0.$$

Preuve (Lemme VII.1 – 1) : Tout d'abord, remarquons que $\delta \mapsto \int_0^{2\delta} h$ est une fonction croissante de δ sur $[0, T/2]$ valant 1 en $T/2$, on peut donc choisir $\delta > 0$ pour lequel $\int_0^{2\delta} h$ soit strictement positif. Notons $F_\delta : u \in [0, \delta] \mapsto \int_u^{\delta+u} h$. Cette fonction est continue donc intégrable, elle est de plus positive. D'après Fubini,

$$\begin{aligned} \int_0^\delta F_\delta(u) du &= \int_0^\delta \int_0^T h(t) \mathbb{1}_{[u, \delta+u]}(t) dt du = \int_0^T h(t) \int_0^\delta \mathbb{1}_{[t-\delta, t]}(u) du dt \\ &= \int_0^T h(t) \text{Leb}([t-\delta, t] \cap [0, \delta]) dt = \int_0^{2\delta} h(t) \ell_\delta(t) dt \end{aligned}$$

avec

$$\ell_\delta(t) = \text{Leb}([t-\delta, t] \cap [0, \delta]) = \begin{cases} t & \text{si } t \in [0, \delta] \\ 2\delta - t & \text{si } t \in]\delta, 2\delta] \end{cases}.$$

Ainsi, ℓ_δ est strictement positive sur l'intervalle $]0, 2\delta[$ et par hypothèse sur $\delta > 0$, on sait que $\int_0^{2\delta} h > 0$ donc h est strictement positive sur un ensemble non négligeable de $]0, 2\delta[$. Ainsi, $\int_0^{2\delta} h(t) \ell_\delta(t) dt > 0$, ce qui montre que $\int_0^\delta F_\delta > 0$ et donc que sur une partie non négligeable de $[0, \delta]$, F_δ est strictement positive, c'est exactement le résultat recherché. $\square$

2 - Inégalité de Bernstein

Nous présentons dans cette partie une preuve simple du cas particulier présenté ci-dessous de l'inégalité de Bernstein dont nous avons besoin pour prouver le [Lemme III.3 – 7](#).

★ Théorème VII.2-1 :

29

Soit $(X_i)_{1 \leq i \leq n}$ des variables aléatoires réelles indépendantes, centrées et bornées par une constante C . Alors, pour tout $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2 \exp\left(-\frac{t^2}{2nC^2 + \frac{2}{3}tC}\right).$$

Preuve (Théorème VII.2 – 1) : On a pour tout $\lambda > 0$, en notant $S_n = \sum_{i=1}^n X_i$, par l'inégalité de Markov :

$$\mathbb{P}(S_n \geq t) \leq e^{-xt} \mathbb{E}(\exp(xS_n)) = \exp\left(-xt + \sum_{i=1}^n \log(\mathbb{E}(e^{xX_i}))\right).$$

Or, puisque X_i est centré, on a

$$\mathbb{E}(e^{xX_i}) = 1 + \sum_{k \geq 2} \frac{\mathbb{E}(X_i^k)}{k!} \leq 1 + \frac{x^2 C^2}{2} \sum_{k \geq 0} \frac{2(xC)^k}{(k+2)!}.$$

Mais on sait que pour tout $k \geq 2$, on a $(k+2)! \geq 2 \times 3^k$ donc, pour $x < \frac{3}{M}$,

$$\mathbb{E}(e^{xX_i}) \leq 1 + \frac{x^2 C^2}{2} \frac{1}{1 - \lambda C/3} \leq \exp\left(\frac{x^2 C^2}{2(1 - \lambda C/3)}\right),$$

par convexité de l'exponentielle. Ainsi, on a pour tout $t > 0$, avec le choix de $x = \frac{t}{nC^2 + Mt/3}$:

$$\mathbb{P}(S_n > t) \leq \exp\left(-\frac{t^2}{2nC^2 + \frac{2}{3}Ct}\right),$$

et on peut faire de même en considérant $(-X_i)_{1 \leq i \leq n}$ et on obtient finalement le résultat attendu :

$$\forall t > 0, \mathbb{P}(|S_n| > t) \leq 2 \exp\left(-\frac{t^2}{2nC^2 + \frac{2}{3}Ct}\right).$$

□

3 - Continuité absolue par rapport à Lebesgue

► Proposition VII.3-1 :

30

Soit $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ de classe $\mathcal{C}^1$ sur $U \subseteq \mathbb{R}^m$ ouvert où sa différentielle est de rang plein. Alors, si $m \geq n$ et μ une mesure finie dont le support inclu dans U , on a l'implication suivante :

$$\mu \ll \text{Leb}_m \implies g_{\#}\mu \ll \text{Leb}_n.$$

Preuve (Proposition VII.3-1) : Nous allons procéder comme si $U = \mathbb{R}^m$ pour améliorer la clarté de la preuve, aucune difficulté ne s'ajoute dans le cas général.

La continuité absolue de μ par rapport à Leb_m nous octroie une application mesurable positive f vérifiant pour tout A mesurable :

$$\mu(A) = \int_A f d\text{Leb}_m.$$

Nous allons utiliser la formule de la Co-aire, que l'on peut retrouver par exemple dans [16] au Théorème 3.11. Nous remarquerons que l'hypothèse du Théorème 3.11 fait référence à une fonction sommable, mais sa positivité et sa mesurabilité sont suffisantes, ce qui sera notre cas d'application. En effet, puisque g est de classe $\mathcal{C}^1$, nous savons que dg est continue. De plus, dg est aussi de rang plein, ce qui fait de $(dg)^\top dg$ une fonction à valeurs dans les matrices symétriques définies positives, nous avons donc $x \mapsto \det((dg(x))^\top dg(x))$ strictement positive sur $\mathbb{R}^m$. Ainsi, la fonction

$$x \mapsto \frac{f(x) \mathbb{1}_{g^{-1}(A)}(x)}{\det((dg(x))^\top dg(x))}$$

est positive et mesurable de par la continuité de dg , la mesurabilité de A et $f \geq 0$. Ainsi, nous obtenons avec la formule de la Co-aire la relation suivante, où $\mathcal{H}^s$ est la mesure de Hausdorff en dimension s :

$$\begin{aligned} \int_{\mathbb{R}^n} \left(\int_{g^{-1}(\{y\})} \frac{f(x) \mathbb{1}_{g^{-1}(A)}(x)}{\det((dg(x))^\top dg(x))} d\mathcal{H}^{m-n}(x) \right) dy &= \int_{\mathbb{R}^m} \frac{f(x) \mathbb{1}_{g^{-1}(A)}(x)}{\det((dg(x))^\top dg(x))} \det((dg(x))^\top dg(x)) dx \\ \int_A \left(\int_{g^{-1}(\{y\})} \frac{f(x) \mathbb{1}_{g^{-1}(A)}(x)}{\det((dg(x))^\top dg(x))} d\mathcal{H}^{m-n}(x) \right) dy &= \int_{g^{-1}(A)} f(x) d\text{Leb}_m(x) = g_{\#}\mu(A) \end{aligned}$$

la fonction $\tilde{f} : y \mapsto \int_{g^{-1}(\{y\})} \frac{f(x) \mathbb{1}_{g^{-1}(A)}(x)}{\det((dg(x))^\top dg(x))} d\mathcal{H}^{m-n}(x)$ étant mesurable positive et d'intégrale 1 sur $\mathbb{R}^n$, on peut donc conclure que $g_{\#}\mu$ admet bien une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ et elle est donc absolument continue par rapport à celle-ci. □

Pour l'énoncé du corollaire suivant, nous rappelons la définition de $\mathbb{R}^{(\mathbb{N})}$ comme ensemble des suites réelles finies :

$$\mathbb{R}^{(\mathbb{N})} := \bigsqcup_{n \geq 1} \mathbb{R}^n.$$

— **Corollaire VII.3-2 :**

31

Soient $m \geq n$ et $g : (\mathbb{R}^{(\mathbb{N})})^m \rightarrow \mathbb{R}^n$ telle que pour tout $K \in (\mathbb{N}^*)^m$, $g_K := g|_{\prod_{j=1}^m \mathbb{R}^{K_j}}$ soit de classe $\mathcal{C}^1$ sur un ouvert U_K où sa différentielle y est aussi de rang plein. Considérons μ une mesure finie dont les restrictions aux $\prod_{j=1}^m \mathbb{R}^{K_j}$ aient un support dans U_K . Alors l'implication suivante est vérifiée :

$$\left(\forall K \in (\mathbb{N}^*)^m, \mu_K \ll \bigotimes_{j=1}^m \text{Leb}_{K_j} \right) \implies g_{\#} \mu \ll \text{Leb}_n$$

$$\text{où } \mu_K : A \mapsto \mu \left(A \cap \prod_{j=1}^m \mathbb{R}^{K_j} \right).$$

Preuve (Corollaire VII.3 – 2) : Nous allons procéder comme si $U_K = \prod_{j=1}^m \mathbb{R}^{K_j}$ pour améliorer la clarté de la preuve, aucune difficulté ne s'ajoute dans le cas général.
Remarquons la décomposition de l'espace suivante :

$$(\mathbb{R}^{(\mathbb{N})})^m = \bigsqcup_{K_1, \dots, K_m \geq 1} \prod_{j=1}^m \mathbb{R}^{K_j}$$

avec laquelle on peut procéder à la décomposition de μ suivante :

$$\mu = \sum_{K \in (\mathbb{N}^*)^m} \mu_K, \text{ avec } \mu_K : A \mapsto \mu \left(A \cap \prod_{j=1}^m \mathbb{R}^{K_j} \right).$$

De plus, la mesure image de μ par g va pouvoir se décomposer comme suit :

$$g_{\#} \mu = \sum_{K \in (\mathbb{N}^*)^m} (g_K)_{\#} \mu_K \text{ où } g_K \text{ est la restriction de } g \text{ à } \prod_{j=1}^m \mathbb{R}^{K_j}.$$

Or, par hypothèse, $\mu_K \ll \bigotimes_{j=1}^m \text{Leb}_{K_j} = \text{Leb}_{\sum_{j=1}^m K_j}$ donc la Proposition VII.3 – 1 nous assure que $(g_K)_{\#} \mu_K$ est absolument continue par rapport à Leb_n . Ceci suffit à conclure car si on considère A mesurable négligeable devant Leb_n alors,

$$g_{\#} \mu(A) = \sum_{K \in (\mathbb{N}^*)^m} (g_K)_{\#} \mu_K(A) = 0.$$

donc $g_{\#} \mu$ est bien absolument continue par rapport à Leb_n . □

• **Lemme VII.3-3 :**

32

Une matrice aléatoire dont la loi est absolument continue par rapport à la mesure de Lebesgue est presque sûrement de rang plein.

Preuve (Lemme VII.3 – 3) : Soit $A \in \mathbb{R}^{m \times n}$ telle que $\mathbb{P}_A \ll \text{Leb}_{mn}$ où $\mathbb{P}_A$ est la loi de A . Remarquons d'abord que le rang d'une matrice est la taille de la plus grande matrice extraite inversible de cette matrice. Ainsi, on peut écrire la relation suivante, en notant R l'ensemble des matrices de taille $m \times n$ de rang non plein, on a en notant $r = \min(n, m)$,

$$R = \bigcap_{\substack{I \subseteq \llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket \\ |I| = r}} (\det(A_I) = 0).$$

Or, pour tout $I \subseteq \llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket$ non vide, l'application $B \in \mathbb{R}^{m \times n} \mapsto \det(B_I)$ est un polynôme non nul en les coefficients de B . Montrons alors que les zéros d'un polynôme à plusieurs variables est toujours un ensemble négligeable pour la

mesure de Lebesgue, ceci permettra de conclure immédiatement. On procède par récurrence sur le nombre de variables du polynôme. Pour une variable, on sait qu'il n'y a qu'un nombre fini de racines, c'est donc de mesure nulle. Si $P \in \mathbb{R}[X_1, \dots, X_N]$, alors on peut l'écrire de cette façon :

$$P(X_1, \dots, X_N) = \sum_{k=0}^{+\infty} a_k(X_1, \dots, X_{N-1}) X_N^k \text{ avec } (a_k)_{k \geq 0} \in (\mathbb{R}[X_1, \dots, X_{N-1}])^{(\mathbb{N})}.$$

Alors, l'ensemble des zéros de P peut s'écrire de la manière suivante :

$$\underbrace{\{(y, x_N) \mid a_k(y) = 0 \text{ pour tout } k \geq 0\}}_{=:D} \cup \underbrace{\{(y, x_N) \mid x_N \text{ racine du polynôme non nul } P(y, \cdot)\}}_{=:R}$$

Or, puisqu'un polynôme non nul a un nombre fini de racines, on sait que

$$\begin{aligned} \text{Leb}_N(R) &= \int_{\mathbb{R}^{N-1}} \int_{\mathbb{R}} \mathbf{1}_{a \neq 0}(y) \mathbf{1}_{P(y, \cdot)=0}(x_N) dx_N dy \\ &= \int_{\mathbb{R}^{N-1}} \mathbf{1}_{a \neq 0}(y) \left(\underbrace{\int_{\mathbb{R}} \mathbf{1}_{P(y, \cdot)=0}(x_N) dx_N}_{=0} \right) dy = 0 \end{aligned}$$

De plus, $\text{Leb}_N(D) = 0$ par hypothèse de récurrence. On en déduit bien le résultat et cela conclut. $\square$

► **Proposition VII.3-4 :**

33

Soient $t \geq 0$ et N un processus de Hawkes de dimension M . Pour tout $K \in (\mathbb{N}^*)^M$, la loi de $(\mathcal{T}_j)_{1 \leq j \leq M}$ intersectée avec $(N_j(t) = K_j)$ pour tout $j \in \llbracket 1, M \rrbracket$ est absolument continue par rapport à la mesure de Lebesgue.

Preuve (Proposition VII.3 – 4) : Rappelons que si on note λ_j les intensités associées et π_j des procesus de poisson indépendant sur $\mathbb{R}^2$ d'intensités unitaire, alors on peut écrire

$$N_j(t) = \int_0^t \int_{\mathbb{R}} \mathbf{1}_{x \leq \lambda_j(s)} d\pi_j(dx, ds).$$

Alors, le processus global $N = \sum_{j=1}^M N_j$ est un hawkes d'intensité $\Lambda = \sum_{j=1}^M \lambda_j$. En notant $(\mathcal{F}_t)_{t \geq 0}$ la tribu associée, on peut calculer la loi conditionnelle du $k+1$ -ième temps de saut de N selon le k -ième :

$$\mathbb{P}(T_{k+1} > t \mid \mathcal{F}_{T_k}) = \exp \left(- \int_{T_k}^t \Lambda(s \mid \mathcal{F}_{s-}) ds \right).$$

Si on note $(Z_k)_{k \geq 0}$ le type du k -ième saut, on trouve

$$\mathbb{P}(Z_{k+1} = j \mid T_{k+1} = t, \mathcal{F}_{T_k}) = \frac{\lambda_j(t \mid \mathcal{F}_{t-})}{\Lambda(t \mid \mathcal{F}_{t-})},$$

on trouve alors la densité du type et du temps jointes :

$$f_{T_{k+1}, Z_{k+1} \mid T_k}(t, j) = \lambda_j(t \mid \mathcal{F}_{t-}) \exp \left(- \int_{T_k}^t \Lambda(s \mid \mathcal{F}_{s-}) ds \right).$$

On peut ainsi calculer la densité jointe des n premiers temps et types :

$$f_{(T_1, Z_1, \dots, T_n, Z_n)}(t_1, z_1, \dots, t_n, z_n) = \prod_{k=1}^n f_{T_k, Z_k \mid \mathcal{F}_{T_{k-1}}}(t_k, z_k) = \exp \left(- \int_0^{t_n} \Lambda(s \mid \mathcal{F}_{s-}) ds \right) \prod_{k=1}^n \lambda_{z_k}(t_k \mid \mathcal{F}_{t_k-}).$$

On peut aussi procéder au calcul suivant pour obtenir le nombre de sauts des types désirés, si on note

$$Z(K_1, \dots, K_M) = \{z \in \llbracket 1, M \rrbracket^n \mid \forall j \in \llbracket 1, M \rrbracket, |z^{-1}(\{j\})| = K_j\},$$

on trouve au final :

$$\begin{aligned}
f_{(T_1, \dots, T_n, N_1, \dots, N_M)}(t_1, \dots, t_n, K_1, \dots, K_M) &= \exp \left(- \int_0^{t_n} \Lambda(s | \mathcal{F}_{s^-}) ds \right) \sum_{z \in Z(K_1, \dots, K_M)} \prod_{k=1}^n \lambda_{z_k}(t_k | \mathcal{F}_{t_k^-}) \\
&= \exp \left(- \int_0^{t_n} \sum_{j=1}^M \left(\mu_j^* + \sum_{l=1}^M a_{k,l}^* \sum_{k: t_k < s} h(t - t_k) \right) ds \right) \\
&\quad \times \sum_{z \in Z(K_1, \dots, K_M)} \prod_{k=1}^n \left(\mu_{z_k}^* + \sum_{j=1}^M a_{z_k,j}^* \sum_{\substack{1 \leq l < k \\ z_l = j}} h(t - t_l) \right).
\end{aligned}$$

Or, d'après l'hypothèse **(H4)** la fonction h est mesurable positive et les coefficients $a_{i,j}^*$ et μ_j^* sont tous positifs, on en déduit que la fonction de densité est bien mesurable positive et bien sûr d'intégrale 1. Or, pour tout $K \in (\mathbb{N}^*)^M$, d'après ce qui précède sur la loi des temps de sauts intersectée avec le nombre de sauts de chaque type, on obtient l'égalité suivante :

$$(\mathbb{P}_{\mathcal{T}})_K : A \mapsto \int_A f \left(T_1, \dots, T_{\sum_{j=1}^M K_j}, K_1, \dots, K_M \right) (t, K) d\text{Leb}_{\sum_{j=1}^M K_j}(t),$$

ce qui montre que $(\mathbb{P}_{\mathcal{T}})_K$ admet une densité par rapport à $\bigotimes_{j=1}^M \text{Leb}_{K_j}$ et donc est bien absolument continue par rapport à la mesure de Lebesgue. □

VIII – Biobliographie

- [1] Sylvain ARLOT et Alain CELISSE. “A survey of cross-validation procedures for model selection”. In : *Statistics Surveys* 4.none (2010), p. 40-79. DOI : [10.1214/09-SS054](https://doi.org/10.1214/09-SS054). URL : <https://doi.org/10.1214/09-SS054>.
- [2] Emmanuel BACRY et al. *Sparse and low-rank multivariate Hawkes processes*. 2020. arXiv : [1501.00725 \[stat.ML\]](https://arxiv.org/abs/1501.00725). URL : <https://arxiv.org/abs/1501.00725>.
- [3] Emmanuel BACRY et al. “tick: a Python Library for Statistical Learning, with an emphasis on Hawkes Processes and Time-Dependent Models”. In : *Journal of Machine Learning Research* 18.214 (2018), p. 1-5. URL : <http://jmlr.org/papers/v18/17-381.html>.
- [4] Ali BAOUAN et al. *What should clubs monitor to predict future value of football players*. 2022. arXiv : [2212.11041 \[stat.AP\]](https://arxiv.org/abs/2212.11041). URL : <https://arxiv.org/abs/2212.11041>.
- [5] Anna BONNET et al. “Neuronal network inference and membrane potential model using multivariate Hawkes processes”. In : *Journal of Neuroscience Methods* 372 (2022), p. 109550. ISSN : 0165-0270. DOI : <https://doi.org/10.1016/j.jneumeth.2022.109550>. URL : <https://www.sciencedirect.com/science/article/pii/S0165027022000772>.
- [6] Pierre BRÉMAUD et Laurent MASSOULIÉ. “Stability of nonlinear Hawkes processes”. In : *The Annals of Probability* 24 (juill. 1996). DOI : [10.1214/aop/1065725193](https://doi.org/10.1214/aop/1065725193).
- [7] Peter BÜHLMANN et Sara GEER. *Statistics for High-Dimensional Data: Method, Theory and Applications*. Jan. 2011. ISBN : 978-3-642-20191-2. DOI : [10.1007/978-3-642-20192-9](https://doi.org/10.1007/978-3-642-20192-9).
- [8] Biao CAI, Jingfei ZHANG et Yongtao GUAN. *Latent Network Structure Learning from High Dimensional Multivariate Point Processes*. 2021. arXiv : [2004.03569 \[stat.ME\]](https://arxiv.org/abs/2004.03569). URL : <https://arxiv.org/abs/2004.03569>.
- [9] Jiahua CHEN et Zehua CHEN. “Extended Bayesian information criteria for model selection with large model spaces”. In : *Biometrika* 95.3 (sept. 2008), p. 759-771. ISSN : 0006-3444. DOI : [10.1093/biomet/asn034](https://doi.org/10.1093/biomet/asn034). eprint : <https://academic.oup.com/biomet/article-pdf/95/3/759/578041/asn034.pdf>. URL : <https://doi.org/10.1093/biomet/asn034>.
- [10] Scott Shaobing CHEN, David L. DONOHO et Michael A. SAUNDERS. “Atomic Decomposition by Basis Pursuit”. In : *SIAM Journal on Scientific Computing* 20.1 (1998), p. 33-61. DOI : [10.1137/S1064827596304010](https://doi.org/10.1137/S1064827596304010). eprint : <https://doi.org/10.1137/S1064827596304010>. URL : <https://doi.org/10.1137/S1064827596304010>.
- [11] Shizhe CHEN et al. *The Multivariate Hawkes Process in High Dimensions: Beyond Mutual Excitation*. 2019. arXiv : [1707.04928 \[stat.ME\]](https://arxiv.org/abs/1707.04928). URL : <https://arxiv.org/abs/1707.04928>.
- [12] Christophe DENIS et al. “Bats monitoring: a classification procedure of bats behaviours based on Hawkes processes”. In : *Journal of the Royal Statistical Society Series C: Applied Statistics* 73 (juin 2024), p. 1025-1041. DOI : [10.1093/jrssc/qlae024](https://doi.org/10.1093/jrssc/qlae024).
- [13] Christophe DENIS et al. “Lasso-type estimator and classification algorithm for high-dimensional multivariate Hawkes processes”. working paper or preprint. Nov. 2025. DOI : [10.48550/arXiv.2407.11455](https://arxiv.org/abs/10.48550/arXiv.2407.11455). URL : <https://hal.science/hal-04646888>.
- [14] Sophie DONNET, Vincent RIVOIRARD et Judith ROUSSEAU. *Nonparametric Bayesian estimation of multivariate Hawkes processes*. 2018. arXiv : [1802.05975 \[math.ST\]](https://arxiv.org/abs/1802.05975). URL : <https://arxiv.org/abs/1802.05975>.
- [15] Paul EMBRECHTS, Thomas LINIGER et Lu LIN. “Multivariate Hawkes processes: an application to financial data”. In : *Journal of Applied Probability* 48.A (2011), p. 367-378. DOI : [10.1239/jap/1318940477](https://doi.org/10.1239/jap/1318940477).
- [16] Lawrence C. EVANS et Ronald F. GARIÉPY. *Measure theory and fine properties of functions, revised version*. 2015. Chapman and Hall/CRC : [9780429161483](https://doi.org/10.1201/b18333) (Mathematics, Statistics). URL : <https://doi.org/10.1201/b18333>.
- [17] Trevor HASTIE, Robert TIBSHIRANI et Martin WAINWRIGHT. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Mai 2015, p. 1-337. ISBN : 9780429171581. DOI : [10.1201/b18401](https://doi.org/10.1201/b18401).
- [18] ALAN G. HAWKES. “Spectra of some self-exciting and mutually exciting point processes”. In : *Biometrika* 58.1 (avr. 1971), p. 83-90. ISSN : 0006-3444. DOI : [10.1093/biomet/58.1.83](https://doi.org/10.1093/biomet/58.1.83). eprint : <https://academic.oup.com/biomet/article-pdf/58/1/83/602628/58-1-83.pdf>. URL : <https://doi.org/10.1093/biomet/58.1.83>.
- [19] Alexander KREISS, Enno MAMMEN et Wolfgang POLONIK. *Common Drivers in Sparsely Interacting Hawkes Processes*. 2025. arXiv : [2504.03916 \[math.ST\]](https://arxiv.org/abs/2504.03916). URL : <https://arxiv.org/abs/2504.03916>.

- [20] Romain E. LACOSTE. “Sparklen: A Statistical Learning Toolkit for High-Dimensional Hawkes Processes in Python”. In : (2025). arXiv : [2502.18979](https://arxiv.org/abs/2502.18979) [stat.ME]. URL : <https://arxiv.org/abs/2502.18979>.
- [21] Théo LEBLANC. “Exponential moments for Hawkes processes under minimal assumptions”. In : *Electronic Communications in Probability* 29.none (jan. 2024). ISSN : 1083-589X. DOI : [10.1214/24-ECP625](https://doi.org/10.1214/24-ECP625). URL : <http://dx.doi.org/10.1214/24-ECP625>.
- [22] Myrto LIMNIOS et Niels R. HANSEN. “Chapter Nine - Selective Review of Penalized Learning Methods for Event Processes”. In : *Stochastic Modeling and Statistical Methods*. Sous la dir. d’Ioannis S. TRIANTAFYLLOU, Sonia MALEFAKI et Alex KARAGRIGORIOU. Advances in Reliability Science. Academic Press, 1^{er} jan. 2025, p. 159-189. DOI : [10.1016/B978-0-44-331694-4.00014-1](https://doi.org/10.1016/B978-0-44-331694-4.00014-1). URL : <https://www.sciencedirect.com/science/article/pii/B9780443316944000141> (visité le 05/07/2025).
- [23] Antoine LOTZ. *A sparsity test for multivariate Hawkes processes*. 2024. arXiv : [2405.08640](https://arxiv.org/abs/2405.08640) [math.ST]. URL : <https://arxiv.org/abs/2405.08640>.
- [24] Nicolai MEINSHAUSEN et Peter BÜHLMANN. “High-dimensional graphs and variable selection with the Lasso”. In : *The Annals of Statistics* 34.3 (2006). ISSN : 0090-5364. DOI : [10.1214/009053606000000281](https://doi.org/10.1214/009053606000000281). URL : <http://dx.doi.org/10.1214/009053606000000281>.
- [25] Zizhuo MENG et al. *Interpretable Transformer Hawkes Processes: Unveiling Complex Interactions in Social Networks*. 2024. arXiv : [2405.16059](https://arxiv.org/abs/2405.16059) [cs.SI]. URL : <https://arxiv.org/abs/2405.16059>.
- [26] George MOHLER et al. “Self-Exciting Point Process Modeling of Crime”. In : *Journal of the American Statistical Association* 106 (mars 2011), p. 100-108. DOI : [10.1198/jasa.2011.ap09546](https://doi.org/10.1198/jasa.2011.ap09546).
- [27] Jesper MØLLER et Jakob G. RASMUSSEN. “Perfect simulation of Hawkes processes”. In : *Advances in Applied Probability* 37.3 (2005), p. 629-646. DOI : [10.1239/aap/1127483739](https://doi.org/10.1239/aap/1127483739).
- [28] Lisa NICVERT et al. “Using the multivariate Hawkes process to study interactions between multiple species from camera trap data”. In : *Ecology* 105.4 (), e4237.
- [29] Yosihiko OGATA. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In : *Journal of the American Statistical Association* 83.401 (1988), p. 9-27. ISSN : 01621459, 1537274X. URL : <http://www.jstor.org/stable/2288914> (visité le 27/08/2026).
- [30] Patricia REYNAUD-BOURET et Sophie SCHBATH. “Adaptive estimation for Hawkes processes; application to genome analysis”. In : *The Annals of Statistics* 38.5 (oct. 2010). ISSN : 0090-5364. DOI : [10.1214/10-AOS806](https://doi.org/10.1214/10-AOS806). URL : <http://dx.doi.org/10.1214/10-AOS806>.
- [31] Robert TIBSHIRANI. “Regression Shrinkage and Selection Via the Lasso”. In : *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (jan. 1996), p. 267-288. ISSN : 0035-9246. DOI : [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x). eprint : https://academic.oup.com/jrsssb/article-pdf/58/1/267/49098631/jrsssb_58_1_267.pdf. URL : <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- [32] Ryan J. TIBSHIRANI. *The Lasso Problem and Uniqueness*. 2012. arXiv : [1206.0313](https://arxiv.org/abs/1206.0313) [math.ST]. URL : <https://arxiv.org/abs/1206.0313>.
- [33] Joel TROPP. “Just relax: Convex programming methods for identifying sparse signals in noise”. In : *IEEE Transactions on Information Theory* 52 (avr. 2006), p. 1030-1051. DOI : [10.1109/TIT.2005.864420](https://doi.org/10.1109/TIT.2005.864420).
- [34] Rui ZHANG et al. *Efficient Non-parametric Bayesian Hawkes Processes*. 2022. arXiv : [1810.03730](https://arxiv.org/abs/1810.03730) [cs.LG]. URL : <https://arxiv.org/abs/1810.03730>.