



HAL
open science

Discrete time Hawkes processes with inhibition

Anthony Muraro

► **To cite this version:**

Anthony Muraro. Discrete time Hawkes processes with inhibition. Probability [math.PR]. Université de Toulouse, 2024. English. NNT : 2024TLSES103 . tel-04816617

HAL Id: tel-04816617

<https://theses.hal.science/tel-04816617v1>

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Processus de Hawkes en temps discret avec inhibition

Thèse présentée et soutenue, le 29 août 2024 par

Anthony MURARO

École doctorale

EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Spécialité

Mathématiques et Applications

Unité de recherche

IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par

Pascal MAILLARD et Manon COSTA

Composition du jury

Mme Laure COUTIN, Présidente, Université Toulouse III - Paul Sabatier

M. Bertrand CLOEZ, Rapporteur, INRAE Occitanie-Montpellier

M. Eric LUÇON, Rapporteur, Université Paris Cité

M. Mathieu ROSENBAUM, Examineur, Institut Polytechnique de Paris

M. Pascal MAILLARD, Directeur de thèse, Université Toulouse III - Paul Sabatier

Mme Manon COSTA, Co-directrice de thèse, Université Toulouse III - Paul Sabatier

*I turned to look but it was gone,
I cannot put my finger on it now,
The child is grown, the dream is gone.
I have become, comfortably numb.*

Pink Floyd
Comfortably Numb - The Wall

Remerciements

Je souhaite commencer en remerciant chaleureusement ma directrice et mon directeur de thèse, Manon et Pascal. Merci à tous les deux pour votre patience et votre habileté à effectuer des dessins en trois dimensions, et votre adaptabilité à toute épreuve. Je mesure la chance que j'ai eu d'avoir un duo comme le vôtre, avec des interactions sereines. Votre calme et votre enthousiasme ont fortement contribué à faire de ces trois ans une période très agréable et appréciable. Vous aimez votre métier et vous le communiquez très bien, merci d'être des chercheurs incroyables et des humains géniaux.

Je remercie également mon jury de thèse. Bertrand Cloez et Eric Luçon, merci d'avoir pris le temps de relire mon manuscrit et pour vos commentaires. Laure Coutin et Mathieu Rosenbaum, merci beaucoup d'avoir accepté de faire partie de mon jury.

Un grand merci à Marie-Hélène et à l'école doctorale MITT en général pour m'avoir permis contre vents et marées de pouvoir soutenir à la fin du mois d'août. Merci d'accompagner comme vous le faites les doctorant.e.s tout au long de leur thèse.

Qu'il me soit permis ici de remercier chaleureusement le personnel administratif, notamment Céline pour sa bonne humeur et Papa pour sa disponibilité et sa sympathie. Merci Jean-Marc de m'avoir accompagné sur la dernière ligne droite, jusqu'à l'impression de ce manuscrit.

Du côté des enseignants de l'IMT ou d'ailleurs, merci à Laurent pour ton accessibilité et les discussions sur le métier d'enseignant, et à Xavier de m'avoir reçu un nombre incalculable de fois dans ton bureau pour parler de la dernière interro. Merci à Joan pour les échanges pour savoir comment expliquer au mieux la diagonalisation aux étudiants et merci à Olivia pour ta confiance. Merci à Fanny pour tes précieux conseils pour les dossiers de candidature. Un grand merci à Michel pour toutes les soirées et pour les nombreuses discussions, je te souhaite plein de bonheur avec Hélène et Simon.

Un immense merci aux étudiants que j'ai pu croiser et à qui j'ai eu l'honneur d'enseigner du mieux que je le pouvais. Merci pour les regards qui s'illuminent quand vous finissez par comprendre, merci pour les efforts consentis. Merci de m'avoir montré ma vocation.

Si ces trois dernières années ont été si merveilleuses, c'est en très grande partie grâce à l'excellente ambiance qui règne à l'IMT, qui abrite les fameux doctorants. Nico, merci pour ton rire contagieux et d'être l'un des seuls représentants du charmant accent du Sud. Merci pour le Quercy des îles, j'ai hâte de voir ce qu'il te reste à nous faire découvrir. Sophia, la harissa vissée sur le porte-clefs, je ne doute

pas que tu vas finir ta thèse en beauté, bon courage pour la suite à Montréal. Joachim, loin des yeux mais proche du coeur, les discussions, les manifs et les Mario Party sont autant de bons souvenirs. Armand, camarade, merci de partager du Pink Floyd tôt le matin ou très tard le soir entre des cartons. Merci à Niki pour sa bonne humeur. Nicolas, merci de ne pas me laisser seul avec les répliques de Kaamelott, même si je dois bien m'incliner devant ta maîtrise incroyable. Lucas C, merci pour ta sympathie et ta culture cinématographique, bon courage pour la suite de ton parcours. Anthony, cher homonyme, j'espère que tu te plairas dans ta carrière d'enseignant. Fanny, je m'attends à te voir débarquer dans mon lycée à tout moment, je surveillerai.

Merci aux nouveaux, Arnaud, Elie et Candice, pour nous prêter main forte au tarot. Merci aux encore plus nouveaux, les bébés Laurens, Mathis, Boris et Paul pour avoir embrassé la tradition du Biergarten. Merci Florian pour les montages pétés et les chansons. Merci Elio pour ton calme et ton magnifique JDR (c'est quand le prochain?), je te souhaite bon courage pour la suite de ta thèse. Bon courage également à Lucile pour le début de ta carrière! Merci à Benjamin de nous faire sortir ailleurs de temps en temps (mais pas trop quand même), et pour les (trop) nombreuses bouteilles de Chardonnay. Merci à Louis Devers (ou bien David?) de me faire mourir de rire avec tes expressions. Merci pour ta gentillesse et je te promets de t'accompagner écumer les mers un de ces quatre. Je souhaite bon courage aux nouveaux occupants du bureau 201, Louis C. et Thibault, je compte sur vous pour assurer la digne relève.

Merci aux anciens, Javi pour ta bonne humeur contagieuse, et les motivations à courir. Alain, Corentin, Lucas, Paola, Clément, merci pour lancer les conversations à l'Upsidum. Perla, tu pourrais nous laisser un peu gagner au tarot quelques fois. Bonne chance pour la suite de tes aventures à Dubaï. Mahmoud, merci pour les discussions autour des Hawkes et des accents arabes. Laetitia, grande soeur de thèse, merci pour l'accueil que tu m'as réservé à l'IMT, et félicitations pour ton poste! Maxime, merci pour garder la tête froide aux jeux coopératifs, et pour les moments partagés dans le bureau.

Alex, quelle chance de te connaître, merci pour ta gentillesse à toute épreuve et pour les nombreux jeux que tu nous as appris. Je garderai toujours un souvenir mémorable des Baraka Jeux et des verres pris en avance dans les bars.

Enfin Etienne et Erwanne, merci d'être des amis incroyables. Merci pour les séances de cinéma avant les sushis, et pour ces sessions de Towerfall et de Skull King (jeu de merde). Félicitations encore une fois le E, quand j'écris ces lignes tu viens tout juste de m'annoncer ton affectation. Si tu as quelques minutes entre deux copies... p'tit zombie?

Avant de débarquer en terre toulousaine, je le confesse, j'ai fait mes classes en région parisienne... pendant 5 ans, j'ai rencontré des gens exceptionnels, merci à Marion, Keivan (tft?), Philippe, Jordan, Jeanne et Donatien d'être tout simplement géniaux et de garder le contact malgré la distance. Contre toute attente, la colloc' aura finalement contribué à fournir des docteurs et futurs docteurs, merci à Gatan

et à Charlou pour votre belle amitié. Merci en particulier à Nico d'être toujours présent, pour les multiples fous rires, merci à Mathilde de l'accompagner.

Avant de débarquer en terre parisienne, je le confesse, j'ai grandi à Bordeaux. J'ai eu la chance d'y rencontrer Joévan, Loïc et Cécily, qui sont maintenant à Toulouse eux aussi (comme quoi. . .), merci à eux pour les verres innocents qui se transforment en longues soirées. Avant ça, mes jeunes années au lycée ont vu les rencontres de Yoann, Seb, Zac, merci d'être toujours dans ma vie, et pour les heures passées devant Rocket League. Charles, les fous rires partagés en cours de SI ne me semblent pas si loin que ça. Merci pour les discussions passionnantes et pour tout ce qui touche de près ou de loin à la bière. Encore avant, j'ai eu la chance de rencontrer Lisa, merci pour les Skype annuels de 3 heures et pour faire perdurer notre belle amitié au fil des années.

Corentin, mon frère de coeur, merci d'être cette personne fantastique à tout moment. Merci d'être inspirant et accessible, as-tu un défaut ? Je t'admire et je sais que rien ne nous séparera. Merci à Valeria pour les Catan et les loups-garous à 4.

Merci aux prix Nobel pour les soirées et les souvenirs. Même en grognant, on partage beaucoup entre nous. Notre amitié m'est chère.

Pour terminer, je me dois de remercier ma famille, merci à mon Papa et à ma Maman d'être des parents magiques qui m'ont toujours fait confiance. Merci pour votre soutien permanent depuis le tout début. Merci à mes belles-familles qui sont maintenant des familles à part entière : merci Lionel et Mathilde, merci Marina et Miléna. Merci à la famille Demoy pour leur accueil et leur gentillesse. Geoffrey, que dire que l'on ne s'est pas déjà dit. Merci d'être un grand frère incroyable, on a encore tant à partager. Si tu as un peu de temps je t'expliquerai ce qu'il y a d'écrit dans ce manuscrit.

Enfin, merci à Pauline, merci pour tout. Merci pour ces nombreuses années ensemble, merci de me supporter et merci de toujours me surprendre. Merci de rire à mes blagues et de faire comme si elles étaient drôles. Merci pour tous les moments partagés, je t'aime fort.

Table des matières

1	Introduction	1
1.1	Les processus de comptage : vers le processus de Hawkes avec inhibition	2
1.1.1	Le processus de Poisson	3
1.1.2	Processus de Hawkes	6
1.1.3	Interlude sur les équations de renouvellement	12
1.2	Processus de Hawkes discret	13
1.2.1	Introduction et définitions	13
1.2.2	Le cadre théorique développé : chaîne de Markov et critères de <i>drift</i>	16
1.2.3	Plan de la thèse et contributions	20
1.2.4	Perspectives	31
2	Explorations numériques	33
2.1	Des équations fonctionnelles intégrales	33
2.1.1	Équation intégrale de Volterra de seconde espèce	34
2.1.2	Équation fonctionnelle non-linéaire	35
2.2	Une classe particulière de fonctions h	40
2.2.1	Comparaison des solutions de l'équation de renouvellement linéaire et non-linéaire	40
2.2.2	Détermination de fonctions h ne satisfaisant pas le critère suffisant de stabilité	42
3	Discrete-time Hawkes process with length two memory	45
3.1	Introduction.	46
3.2	Notations, definitions and results	48
3.2.1	Definition and main result	48
3.2.2	The associated Markov chain	49
3.3	Proof of Theorem 3.2.2 : recurrence	50
3.3.1	Foster-Lyapounov drift criteria	51
3.3.2	Case \mathcal{R}_1	52
3.3.3	Case \mathcal{R}_2	53
3.3.4	Case \mathcal{R}_3	55
3.4	Proof of Theorem 3.2.2 : transience	57
3.4.1	Case T1	58
3.4.2	Case T2 : $0 \leq a < 2$ and $a + b > 1$ or $a \geq 2$ and $a^2 + 4b > 0$	59
3.5	Perspectives and open problems	64
3.A	Linear recurrence equations	65
3.B	Criteria for strong irreducibility	66

4	Discrete-time Hawkes process with length three memory	71
4.1	Introduction.	72
4.2	Main results	73
4.2.1	The general setting	73
4.2.2	Three-parameter discrete-time Hawkes process	74
4.2.3	Transience results	77
4.2.4	Recurrence results	78
4.2.5	Conjecture and numerical illustrations	78
4.3	Proofs of Recurrent cases	83
4.3.1	Lyapunov function and Drift criterion	83
4.3.2	Proof of Theorem 4.2.1	85
4.3.3	Proof of Theorem 4.2.3	87
4.4	Proofs of Transience cases	92
4.4.1	Case $a, b < 0, c > 1$	93
4.4.2	Case $b > 1, ab + c < 0$	94
4.A	Proof of transience when $a_1, \dots, a_p \geq 0$ and $a_1 + \dots + a_p > 1$	97
4.B	Proofs of technical lemmas	98
4.B.1	Proof of Lemma 4.2.4	98
4.B.2	Proof of Lemma 4.3.7	99
4.B.3	Proof of Lemma 4.B.1	99
4.C	Perspectives	100
5	Critical discrete-time Hawkes processes	105
5.1	Introduction.	105
5.2	The two-parameter case	107
5.2.1	A discrete martingale	107
5.2.2	Framework and theoretical results	109
5.2.3	Time and space rescaling	111
5.3	Perspectives	122

Introduction

Contents

1.1 Les processus de comptage : vers le processus de Hawkes avec inhibition	2
1.1.1 Le processus de Poisson	3
1.1.1.1 Processus de Poisson homogène	3
1.1.1.2 Processus de Poisson inhomogène	4
1.1.2 Processus de Hawkes	6
1.1.2.1 Processus de Hawkes linéaire	7
1.1.2.2 Le processus de Hawkes avec inhibition	10
1.1.3 Interlude sur les équations de renouvellement	12
1.2 Processus de Hawkes discret	13
1.2.1 Introduction et définitions	13
1.2.2 Le cadre théorique développé : chaîne de Markov et critères de <i>drift</i>	16
1.2.2.1 Stratégie pour démontrer l'ergodicité de (X_n)	16
1.2.2.2 Stratégie pour démontrer la transience de (X_n)	19
1.2.3 Plan de la thèse et contributions	20
1.2.3.1 Le cas $p = 2$	20
1.2.3.2 Le cas $p = 3$	26
1.2.3.3 Description de quelques cas critiques	29
1.2.4 Perspectives	31

Cette thèse porte sur les processus de Hawkes. Les processus de Hawkes sont des processus de comptage en temps continu, dont l'intensité est elle-même aléatoire et dépend de l'ensemble des points passés. Introduits pour la première fois par Hawkes [41] et Oakes [43] dans les années 1970, leur célébrité est dans un premier temps due à leur utilisation pour la modélisation des tremblements de terre par Hawkes [42] et Ogata [71, 70]. Aujourd'hui, les processus de Hawkes restent très largement utilisés dans des domaines variés comme par exemple la criminologie [68, 88], l'épidémiologie [10, 57], la finance mathématique [5, 40] ou encore les neurosciences [35, 37].

Nous avons fait le choix de nous concentrer sur une version discrète du processus de Hawkes, que nous détaillerons dans la suite, afin d'obtenir de nouveaux résultats sur la stabilité de ce processus discret lorsqu'on modélise un phénomène d'inhibition dans le système. Dans un premier temps, nous rappellerons donc progressivement la définition du processus de Hawkes continu ainsi que sa version discrète, avant de

décrire les résultats obtenus et les contributions de cette thèse pour l'étude de ces processus.

1.1 Les processus de comptage : vers le processus de Hawkes avec inhibition

Démarrons cette introduction avec une motivation décrite par Appel dans son livre [1]. Pour commencer, on souhaite répartir n points uniformément et indépendamment les uns des autres sur un intervalle $[a, b]$ donné, avec $0 < a < b < +\infty$. Jusque-là, pas de difficulté particulière, il suffit simplement de définir n variables aléatoires indépendantes et identiquement distribuées (i.i.d) de loi uniforme sur $[a, b]$, les réalisations de ces variables aléatoires formant nos points dans les conditions voulues.

Si maintenant on s'intéresse à une question un peu plus générale et abstraite, à savoir la répartition d'une **infinité** de points sur toute la demi-droite réelle \mathbb{R}_+ , dans les mêmes conditions. On veut dire par là qu'on souhaite que, pour deux intervalles $[a, b]$ et $[c, d]$ disjoints, le nombre de points dans $[a, b]$ soit indépendant du nombre de points dans $[c, d]$. De plus, par "répartition uniforme", on entend désormais ceci : introduisons un nombre réel $\lambda \geq 0$. Alors, pour tout intervalle borné $[a, b]$, le nombre de points sur cet intervalle est une variable aléatoire, et on impose que celle-ci soit d'espérance égale à $(b-a)\lambda$. Le nombre λ décrit en quelque sorte la densité de points qu'on s'attend, en moyenne, à trouver dans un intervalle donné. Ici, la densité de points est constante égale à λ , ce qui décrit l'uniformité de la distribution des points le long de la demi-droite réelle.

Formalisons tout cela dans le langage probabiliste : pour tout $t \geq 0$, on note $N(t)$ la variable aléatoire donnant le nombre de points de la construction décrite ci-dessus se situant dans l'intervalle $[0, t]$. On a en fait affaire à un processus de comptage :

Définition 1.1.1. *On considère un processus stochastique réel, c'est-à-dire une famille*

$$\{N(t), t \geq 0\}$$

*de variables aléatoires, définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs réelles. On suppose que ce processus permet de compter, au cours du temps, le nombre de réalisations d'un événement donné. Le processus $\{N(t), t \geq 0\}$ est appelé **processus de comptage** si les variables $N(t)$ sont toutes à valeurs entières et si, pour tout $\omega \in \Omega$, la fonction $t \mapsto N(t, \omega)$ est croissante. La quantité $N(t)$ compte alors le nombre d'événements, qu'on appelle **points**, ayant eu lieu durant l'intervalle de temps $[0, t]$.*

Cela signifie en particulier que le nombre de points dans l'intervalle $]s, t]$ est $N(t) - N(s)$. Finalement, nous aurons réponse à notre problème dès lors que nous construirons un processus de comptage $\{N(t), t \geq 0\}$ tel que, pour tout $a < b < c < d$,

- $N(b) - N(a)$ est indépendante de $N(d) - N(c)$.
- $N(b) - N(a)$ est d'espérance $(b - a)\lambda$.

On va voir qu'un exemple d'un tel processus est donné par le processus de Poisson.

1.1.1 Le processus de Poisson

1.1.1.1 Processus de Poisson homogène

Définition 1.1.2. *Un processus de comptage $\{N(t), t \geq 0\}$ est un **processus de Poisson homogène d'intensité $\lambda > 0$** si :*

1. $N(0) = 0$.
2. *Le processus est à **accroissements indépendants**, c'est-à-dire que pour tout intervalles bornés disjoints $I_1, I_2, \dots, I_n \subset \mathbb{R}_+$*

$$N(I_1) \perp\!\!\!\perp N(I_2) \perp\!\!\!\perp \dots \perp\!\!\!\perp N(I_n).$$

3. *Le processus est à **accroissements stationnaires**, c'est-à-dire que pour tout $s, t > 0$,*

$$N(s + t) - N(t) \stackrel{(\text{loi})}{=} N(s).$$

4. *Pour tout $t \geq 0, h > 0$, lorsque $h \rightarrow 0$:*

$$\begin{aligned} \mathbb{P}(N(t + h) - N(t) = 0) &= 1 - \lambda h + o(h) \\ \mathbb{P}(N(t + h) - N(t) = 1) &= \lambda h + o(h) \\ \mathbb{P}(N(t + h) - N(t) \geq 2) &= o(h), \end{aligned} \tag{1.1.1}$$

Le nom de processus de Poisson vient de la proposition suivante :

Proposition 1.1.3. *Pour tout $t > 0$, la variable aléatoire $N(t)$ suit une loi de Poisson de paramètre λt :*

$$\forall t > 0, \forall n \in \mathbb{N}, \quad \mathbb{P}(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Remarque 1.1.4. Pour plus de détails autour du processus de Poisson homogène sur \mathbb{R}_+ ainsi que des exemples d'applications, notamment en biologie, nous renvoyons au chapitre 5 du livre de Méléard [66].

Finalement, nous avons répondu à notre première question : sur la demi-droite \mathbb{R}_+ , les points d'un processus de Poisson homogène sont répartis aléatoirement et uniformément avec la densité λ désirée. En effet, le processus de Poisson est à accroissements indépendants, propriété qui correspond exactement à l'indépendance du nombre de points dans deux intervalles disjoints. De plus, ce processus étant à accroissements stationnaires, le nombre de points dans un intervalle $[a, b]$ est une variable aléatoire qui suit une loi de Poisson de paramètre $\lambda(b - a)$, et qui est donc d'espérance $\lambda(b - a)$.

En particulier, si l'on se fixe une fenêtre de temps de taille 1, on a la relation suivante :

$$\forall t \geq 0, \quad \mathbb{E}[N(t+1) - N(t)] = \lambda. \quad (1.1.2)$$

Nous venons de donner une première intuition de ce que l'**intensité** d'un processus de comptage représente : au cours du temps, c'est le nombre moyen de points que l'on s'attend à trouver durant une période de temps de longueur unitaire. Remarquons que le processus de Poisson homogène tire son nom de la relation (1.1.2) : ici, l'intensité est constante au cours du temps.

On peut désormais étendre notre question initiale et considérer un problème plus général, dans lequel la répartition des points sur la demi-droite \mathbb{R}_+ ne serait plus uniforme au cours du temps, mais suivrait une tendance décrite par une fonction déterministe. En d'autres termes, et pour utiliser le vocabulaire qui vient d'être introduit, on souhaite construire un processus de comptage dont l'intensité n'est plus constante mais qui est une fonction du temps.

1.1.1.2 Processus de Poisson inhomogène

Définition 1.1.5. *Un processus de comptage $\{N(t), t \geq 0\}$ est un **processus de Poisson inhomogène** d'intensité $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ si :*

1. $N(0) = 0$.
2. *Le processus est à accroissements indépendants.*
3. *Pour tout $t \geq 0$,*

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

Remarque 1.1.6. Plus généralement, l'intensité d'un processus de comptage est définie comme suit :

*Soit $\{N(t), t \geq 0\}$ un processus de comptage. L'**intensité** du processus N est la fonction définie par :*

$$\forall t \geq 0, \quad \lambda(t) := \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(N(t+h) - N(t) > 0)}{h}. \quad (1.1.3)$$

Le processus de Poisson inhomogène possède des propriétés similaires à son homologue homogène, notamment :

Proposition 1.1.7. *Si N est un processus de Poisson inhomogène d'intensité λ , alors pour tout $t \geq 0$ la variable aléatoire $N(t)$ suit une loi de Poisson de paramètre $\int_0^t \lambda(s) ds$, c'est-à-dire :*

$$\forall t > 0, \forall n \in \mathbb{N}, \quad \mathbb{P}(N(t) = n) = \frac{e^{-\int_0^t \lambda(s) ds} \left(\int_0^t \lambda(s) ds \right)^n}{n!}.$$

Les processus de Poisson ainsi construits sont un cas particulier de processus de Poisson généraux.

Définition 1.1.8. Soit ν une mesure borélienne sur (E, d) espace métrique complet séparable (typiquement un espace euclidien, par exemple \mathbb{R}^d), telle que $\nu(A) < \infty$ pour tout borélien A borné de E . Le processus $\{N(t), t \geq 0\}$ est un **processus ponctuel de Poisson général** sur E de mesure intensité ν si :

1. Pour tout A borélien borné, $N(A)$ suit une loi de Poisson de paramètre $\nu(A)$.
2. Pour tout A_1, \dots, A_k boréliens bornés disjoints, $N(A_1), \dots, N(A_k)$ sont indépendantes.

Notamment, la fonction d'intensité λ d'un processus de Poisson inhomogène peut-être vue comme la dérivée de Radon-Nikodym d'une certaine mesure intensité ν par rapport à la mesure de Lebesgue. Pour de plus amples détails autour des processus ponctuels de Poisson généraux, nous renvoyons au livre de Kingman [53].

Simulation d'un processus de Poisson inhomogène : le thinning Dans ce qui suit, on donne une méthode de simulation d'un processus de Poisson inhomogène utilisant un processus de Poisson homogène sous-jacent, qu'on utilise en pratique pour simuler numériquement ces processus. Ainsi, nous espérons rendre plus naturelle la construction du Processus de Hawkes par la suite.

Si l'on souhaite construire un processus de Poisson inhomogène d'intensité λ (ici λ est une fonction bornée de \mathbb{R}_+ dans \mathbb{R}_+) sur l'intervalle de temps $[0, T]$, $T > 0$, on peut en effet :

1. Se donner un majorant de la fonction λ sur $[0, T]$, qu'on note M .
2. Simuler un processus de Poisson homogène d'intensité M sur $[0, T]$. Pour chacun des points t_i de ce processus, on associe une variable aléatoire U_i de loi uniforme sur $[0, M]$.
3. Pour tous les points t_i , si la réalisation de la variable aléatoire uniforme associé U_i est supérieure à $\lambda(t_i)$, le point t_i est supprimé. Sinon, il est conservé.
4. Les points restants à l'issue de l'étape 3 forment un processus de Poisson inhomogène sur $[0, T]$.

La figure 1.1 illustre l'algorithme décrit ci-dessus. Plus formellement, la construction d'un processus de Poisson inhomogène *via* un processus de Poisson homogène sous-jacent provient du théorème suivant, issu de l'article de Lewis et Schedler [60] :

Théorème 1.1.9. On note $\bar{\lambda} := \sup_{0 \leq t \leq T} \lambda(t)$. Soit $(\bar{N}(t), t \geq 0)$ un processus de Poisson homogène d'intensité $\bar{\lambda}$, dont on note les points sur l'intervalle $[0, T]$: $\bar{t}_1, \bar{t}_2, \dots, \bar{t}_{\bar{N}(T)}$.

Pour tout $k \in \{1, \dots, \bar{N}(T)\}$, on supprime le point \bar{t}_k avec probabilité $1 - \frac{\lambda(\bar{t}_k)}{\bar{\lambda}}$. Alors les points restants forment un processus de Poisson inhomogène $\{N(t), t \geq 0\}$ satisfaisant (1.1.5).

Cette construction rend très intuitive la notion d'intensité d'un processus de comptage. Si l'on trace la fonction d'intensité d'un processus de Poisson inhomogène, au vu du théorème 1.1.9, il est naturel qu'en moyenne, on s'attende à trouver

beaucoup de points aux instants où la fonction d'intensité a des valeurs élevées, et vice-versa.

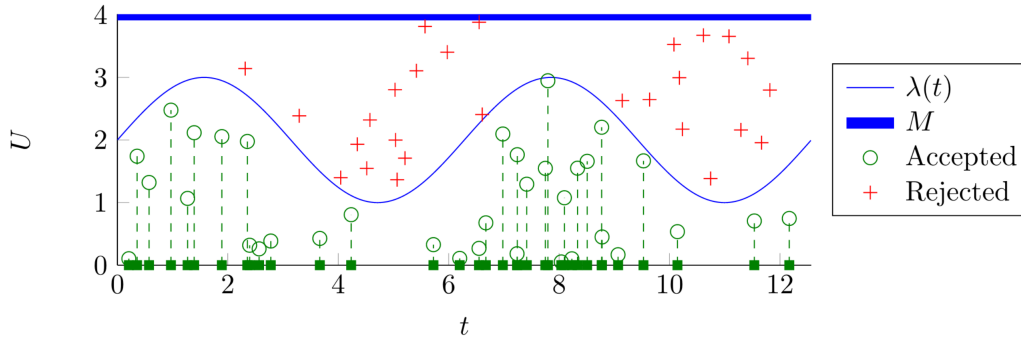


FIGURE 1.1 – Simulation d'un processus de Poisson inhomogène sur $[0, 12]$, d'intensité $\lambda(t) := 2 + \sin(t)$. Ici, $M = 4$ mais tout majorant de λ est admissible. Figure tirée de Laub *et al* [58].

1.1.2 Processus de Hawkes

Jusqu'à présent, l'intensité d'un processus telle qu'on l'a définie est une fonction déterministe du temps. D'un point de vue modélisation, c'est une hypothèse évidemment très forte que l'on fait. Si les points des processus précédemment décrits représentent par exemple des occurrences de phénomènes naturels, ces derniers dépendent bien souvent de leurs occurrences passées. Néanmoins, l'utilisation de processus de Markov est très largement répandue, d'une part puisque les calculs ainsi que les simulations de ces modélisations s'en trouvent grandement facilitées, mais également parce que les résultats mêmes de leurs interprétations sont globalement déjà satisfaisants. Pour certains phénomènes cependant, où une propriété d'auto-excitation émerge, comme c'est le cas par exemple des tremblements de terre ou des bulles sur les marchés financiers, on perd beaucoup d'informations pertinentes en modélisant ces phénomènes à l'aide de processus de Markov. C'est pourquoi nous allons introduire une généralisation des processus de Poisson, prenant en compte l'influence des événements passés sur l'intensité du processus dans le futur. D'un point de vue probabiliste nous allons avoir besoin de définir de nouveaux outils afin de définir le dernier processus de cette introduction : le processus de Hawkes.

Remarque 1.1.10. Il existe également des processus dont l'intensité est aléatoire sans dépendre du passé du processus. Un exemple est donné par le processus de Cox [15], aussi connu sous le nom de double processus stochastique de Poisson. Il s'agit d'un processus ponctuel de Poisson tel que défini à la définition 1.1.8, pour lequel la mesure intensité est une mesure aléatoire.

1.1.2.1 Processus de Hawkes linéaire

On commence par élargir la notion d'intensité précédemment introduite. Les processus de Poisson de la section précédente sont des processus sans mémoire : le fait de connaître le comportement du processus à un instant donné ne modifie pas les prévisions pour le futur. On souhaite justement faire évoluer cela en définissant un processus qui va posséder une mémoire de ses points passés, de telle manière que ces derniers ont une influence sur la probabilité d'apparition des points futurs. On commence par quelques définitions :

Définition 1.1.11. Soit $\{N(t), t \geq 0\}$ un processus de comptage. La **filtration naturelle** $\mathcal{F}^N = (\mathcal{F}_t^N, t \geq 0)$ du processus est définie par :

$$\forall t \geq 0, \quad \mathcal{F}_t^N := \sigma \{N([0, s]), s \leq t\}.$$

La filtration \mathcal{F}_t^N est donc la σ -algèbre engendrée par N jusqu'au temps t , et représente l'historique interne de N jusqu'à ce temps. Elle modélise la mémoire que possède le processus de ses points passés, avant le temps t . On peut maintenant définir une généralisation de la fonction d'intensité : puisqu'on souhaite inclure la mémoire du processus, nous allons conditionner la probabilité infinitésimale décrite par (1.1.3) par la filtration naturelle \mathcal{F}_t^N introduite dans la définition 1.1.11, pour tout $t \geq 0$. Nous allons donc obtenir une famille de variable aléatoire indexée par $t \geq 0$, donc un processus, ce qui justifie la définition suivante :

Définition 1.1.12. Soit $\{N(t), t \geq 0\}$ un processus de comptage muni de sa filtration naturelle \mathcal{F}^N . Le processus défini par :

$$\forall t \geq 0, \quad \lambda^*(t) := \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(N(t+h) - N(t) > 0 \mid \mathcal{F}_t^N)}{h}$$

est appelé **intensité conditionnelle** du processus N .

Remarque 1.1.13. Pour le processus de comptage $\{N(t), t \geq 0\}$, cela signifie en particulier que :

$$\forall t \geq 0, \quad \mathbb{P}(\text{Un saut se produit dans l'intervalle } (t, t + dt) \mid \mathcal{F}_t^N) = \lambda^*(t)dt.$$

On peut alors voir λ^* comme une densité conditionnelle décrivant la distribution du temps d'arrivée du prochain saut.

Plus généralement, pour plus de détails sur l'intensité conditionnelle d'un processus ponctuel le lecteur est renvoyé à la section 7.2 du livre de Daley et Vere-Jones [16].

Maintenant que l'on a construit un objet mathématique représentant la mémoire d'un processus, nous allons quantifier la manière dont les points passés ont une influence sur les points futurs. Pour cela, partons d'une situation concrète : on se place à l'instant du premier point de notre processus. On veut que celui-ci ait une influence sur la probabilité d'apparition des points suivants : on modélise cela *via*

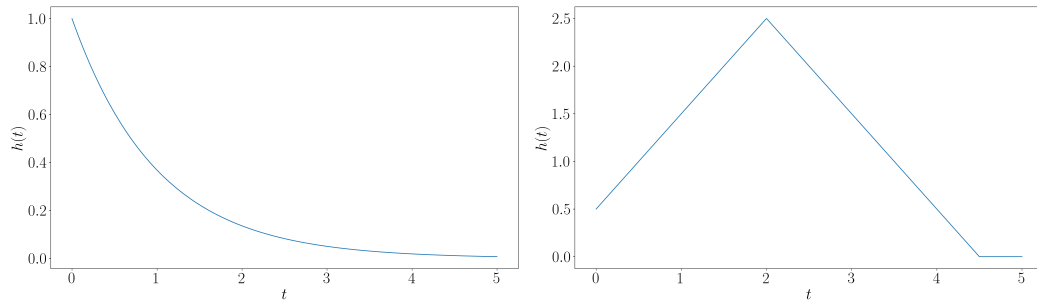


FIGURE 1.2 – Exemples de noyau h pour un processus de Hawkes. À gauche, l'arrivée d'un point augmente instantanément l'intensité conditionnelle avant de décroître : l'influence du point décroît exponentiellement vite. À droite, un saut provoque une augmentation de l'intensité conditionnelle, et cette influence augmente encore sur l'intervalle $[0, 2]$, avant de décroître et de valoir 0 : à partir d'un temps $t = 4.5$, un saut n'a plus d'influence sur le futur.

une fonction $h : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, qui décrira l'évolution de l'influence de chaque nouveau point sur l'intensité conditionnelle du processus : on donne des exemples à la figure 1.2.

Cette influence, encodée dans la fonction h , vaudra pour chacun des points du processus. Dans le cas où plusieurs points apparaissent durant une même période d'influence d'un point précédent, il suffit d'additionner les participations de ces nouveaux points à l'intensité du processus. On appelle couramment une telle fonction h un **noyau**. Cela nous amène à la définition suivante :

Définition 1.1.14. *Un processus de comptage $\{N(t), t \geq 0\}$ est un **processus de Hawkes linéaire** d'intensité conditionnelle λ^* , de noyau $h : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ si :*

- $N(0) = 0$
- L'intensité conditionnelle du processus est λ^* , et celle-ci peut s'écrire :

$$\lambda^*(t) = \lambda + \int_{(0,t)} h(t-s)N(ds) = \lambda + \sum_{t_i < t} h(t-t_i),$$

où $\lambda > 0$ et les $\{t_i\}$ sont les points de N .

Remarque 1.1.15. • On voit que chaque point participe à l'augmentation de l'intensité conditionnelle, au travers de la fonction h qui est positive. On dit souvent des processus de Hawkes linéaires qu'ils sont des processus "auto-excitants".

- Au vu de la définition 1.1.14, il est clair qu'*a priori* le processus de Hawkes linéaire n'est pas un processus de Markov, puisque ses réalisations dépendent de l'ensemble du passé du processus. Pour certaines classes de fonctions h cependant, il est possible de retrouver une propriété d'absence de mémoire du processus, typiquement pour des noyaux de la forme $h(t) = \alpha e^{-\beta t}$, où $\alpha, \beta > 0$.

Pour rendre cette définition plus intuitive et mieux visualiser ce qu'est un processus de Hawkes, il est plus aisé de ne se focaliser que sur son intensité conditionnelle. En effet, cette dernière caractérise le processus (proposition 7.2.IV de [16]). On fera souvent l'amalgame entre le processus de Hawkes et son intensité conditionnelle associée, vue comme un processus à part entière. Sur la figure 1.3, on comprend que le processus de Hawkes n'est qu'une généralisation d'un processus de Poisson inhomogène, où l'intensité change instantanément à chaque nouvelle occurrence d'événement. On peut de la même manière simuler celui-ci en utilisant un processus de Poisson homogène sous-jacent, la difficulté résidant souvent dans la définition d'un majorant de l'intensité sur toute la plage de temps considérée.

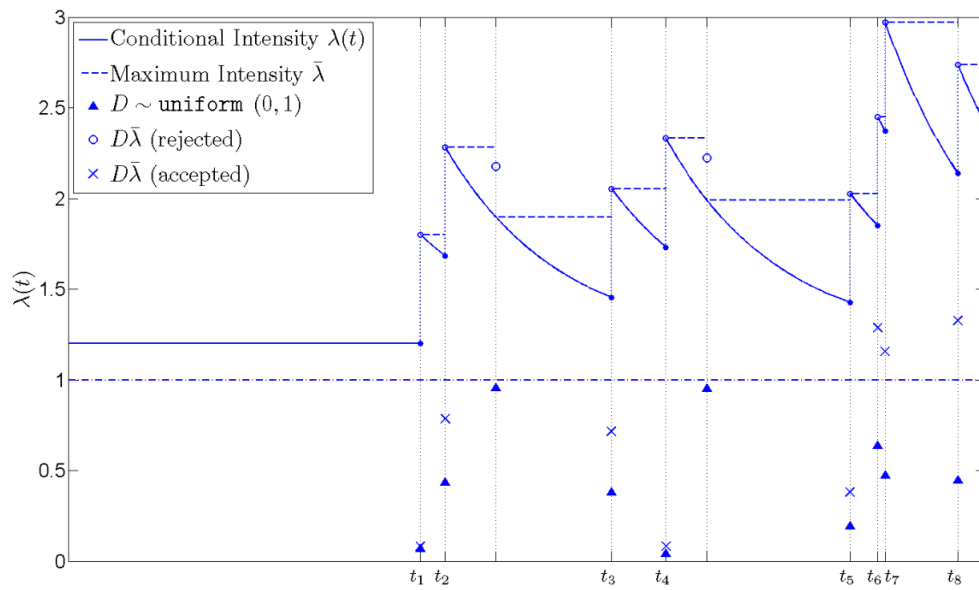


FIGURE 1.3 – Une simulation d'un processus de Hawkes linéaire, où $\lambda = 1.2$ et $h(t) = 0.6e^{-0.8t}$. Dans le cas d'une fonction h décroissante, on peut redéfinir le paramètre $\bar{\lambda}$ du processus de Poisson homogène sous-jacent à chaque occurrence d'un de ses points, résultant en un gain considérable de temps computationnel (méthode dite d'Ogata). Figure créée par Yuanda Chen.

Lorsqu'on étudie le processus de Hawkes, il est naturel de s'intéresser à son comportement asymptotique, et notamment à la question de l'existence d'une version stationnaire de ce processus satisfaisant la définition 1.1.14. On dit dans ce cas que le processus de Hawkes est stable : cette propriété dépend du noyau h , et une condition suffisante de stabilité est donnée par le théorème 1.1.16.

Théorème 1.1.16. *Soit $\lambda > 0$ et $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Il existe une version stationnaire du processus de Hawkes linéaire de noyau h si :*

$$\|h\|_1 := \int_0^{+\infty} h(s)ds < 1. \tag{1.1.4}$$

On doit ce résultat à Hawkes et Oakes [43] qui ont tiré parti de la représentation naturelle du processus de Hawkes comme un processus de Galton-Watson avec immigration : chaque point du processus est soit un immigrant, soit le descendant d'un point passé. La condition (1.1.4) correspond alors au cas sous-critique de ce processus de branchement.

Le processus de Hawkes linéaire a été beaucoup étudié et est aujourd'hui bien compris : la stabilité de ce processus, mais également des théorèmes limites de type loi des grands nombres et théorème central limite (qu'on peut trouver dans l'article de Bacry *et al.* [4]) ainsi que des principes de grandes déviations (notamment dûs à Bordenave et Torrisi [6]) ont été prouvés. Nous nous intéressons plus particulièrement au processus de Hawkes non-linéaire que nous présentons maintenant.

1.1.2.2 Le processus de Hawkes avec inhibition

Une généralisation du processus de Hawkes linéaire consiste à rajouter une dépendance en temps de l'intensité conditionnelle λ^* *via* une fonction déterministe $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ qu'on appellera **noyau extérieur** (la fonction h étant désormais nommée plus précisément le **noyau intérieur**) :

Définition 1.1.17. *Un processus de comptage $\{N(t), t \geq 0\}$ est un **processus de Hawkes non-linéaire** d'intensité conditionnelle λ^* , de noyau extérieur $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ et de noyau intérieur $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ si :*

- $N(0) = 0$
- L'intensité conditionnelle du processus est λ^* , et celle-ci peut s'écrire :

$$\lambda^*(t) = \phi \left(\int_{(0,t)} h(t-s)N(ds) \right) = \phi \left(\sum_{t_i < t} h(t-t_i) \right),$$

où les $\{t_i\}$ sont les points de N .

Remarque 1.1.18. Si on pose

$$\forall x \in \mathbb{R}_+, \quad \phi(x) := \lambda + x,$$

où $\lambda > 0$, on retrouve la définition 1.1.14 d'un processus de Hawkes linéaire.

Les processus de Hawkes non-linéaires ont été moins étudiés que leurs homologues linéaires, principalement à cause des complications techniques et computationnelles induites par l'ajout du noyau extérieur ϕ . En effet, avec la perte de la linéarité de l'intensité conditionnelle s'ajoute la perte de la propriété de branchement de ce processus. Cependant on peut citer de nombreuses avancées dans la compréhension de ces processus : Brémaud et Massoulié [8] ont donné un critère de stabilité pour ce processus sous une hypothèse de lipschitzianité du noyau extérieur, et Massoulié [9] a explicité la vitesse de convergence vers l'état stationnaire de ce processus. Pour des noyaux de type Erlang, des résultats de stabilité et de

convergence vers l'équilibre ont également été démontrés par Duarte *et al.* [20]. Enfin, Karabash [50] a démontré la stabilité d'une classe plus large de processus de Hawkes non-linéaire, relâchant l'hypothèse d'un noyau Lipschitz. Des TCL ainsi que des principes de grandes déviations concernant ces processus ont été démontrés par Zhu [86, 87].

Pour certaines applications, il peut être intéressant de modéliser un effet d'inhibition, décrivant le fait qu'un événement peut diminuer la probabilité de survenue du prochain saut. Cet effet peut être modélisé en autorisant le noyau intérieur h à prendre des valeurs négatives. Ainsi, avec un tel noyau intérieur, un point du processus peut éventuellement participer négativement à l'intensité conditionnelle du processus, c'est-à-dire diminuer celle-ci. Comme le noyau extérieur ϕ est à valeurs positives, l'intensité conditionnelle d'un tel processus sera positive et donc bien définie. Afin de modéliser correctement l'effet de l'inhibition induite par le noyau intérieur h , on supposera également que le noyau extérieur ϕ est une fonction croissante. En reprenant la définition 1.1.17, si $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ est une fonction réelle croissante à valeurs positives, et $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ une fonction à valeurs réelles, éventuellement négatives, on parle de **processus de Hawkes avec inhibition**.

Par exemple, Costa *et al.* [13] pour des résultats d'existence et de stabilité et Cattiaux *et al.* [12] pour des théorèmes limites sur les processus de Hawkes avec inhibition utilisent la fonction ReLU, définie par :

$$\forall x \in \mathbb{R}, \quad \phi(x) := \max(0, \lambda + x), \quad \lambda > 0.$$

La fonction ReLU modélise très simplement l'effet de l'inhibition, et permet de comparer facilement la dynamique du processus avec le modèle linéaire associé. On retrouve d'autres choix de fonctions ϕ dans la littérature, notamment dans l'article de Mei et Eisner [65], dans lequel $\phi(x) = s \log(1 + \exp(x/s))$, $s > 0$.

Remarque 1.1.19. Il est possible de modéliser différemment l'effet de l'inhibition. Par exemple, on peut suivre l'évolution d'une population d'individus, dont chaque comportement individuel est décrit par un processus de Hawkes, en interaction avec les autres individus : c'est le **processus de Hawkes multivarié**. Cette population peut être divisée en deux catégories, l'une excitatrice et l'autre inhibitrice, comme étudié par Raad et Löcherbach dans [73]. Dans ce sens, Duval, Luçon et Pouzat [22] ont travaillé sur des processus de Hawkes avec inhibition avec une approche champ-moyen, en modélisant l'inhibition de façon multiplicative : l'intensité conditionnelle d'une population (telle que décrite dans la définition 1.1.17) est multipliée par une fonction positive, par exemple décroissante vers 0 en la taille de la population inhibitrice du système.

Maintenant que l'on s'autorise à considérer des fonctions h signées, il est naturel de se demander s'il est possible de trouver un noyau h tel que la condition (1.1.4) ne soit pas satisfaite, mais dont résulterait quand même un processus de Hawkes stable. En d'autres termes, les valeurs négatives de la fonction h sont-elles suffisantes à récupérer de la stabilité dans un cas où le processus est instable en ne considérant

que les valeurs positives de h ? Une première réponse est donnée dans [13], où il est prouvé qu'une condition suffisante de stabilité est :

$$\|h^+\|_1 < 1. \quad (1.1.5)$$

Bien que la condition (1.1.5) soit moins restrictive que (1.1.4), elle ne prend pas en compte les éventuelles valeurs négatives de la fonction h . L'objectif premier de cette thèse était de raffiner ce critère suffisant de stabilité, en particulier afin d'incorporer véritablement l'effet de l'inhibition induite h . Des avancées récentes [73] ont permis d'obtenir de nouvelles pistes. Toutefois, cette question semble actuellement extrêmement difficile à traiter en toute généralité. Cette difficulté nous a conduit à nous pencher dans cette thèse sur un processus analogue en temps discret, afin de travailler dans un cadre théorique plus propice à l'étude rigoureuse du comportement asymptotique d'une certaine classe de processus qui se veut proche du processus de Hawkes continu. L'objectif était également de tirer des enseignements de la compréhension du comportement d'un processus simplifié pour porter un éclairage sur les résultats envisageables à obtenir dans le cas continu, ainsi que les outils probabilistes nécessaires à leur démonstrations.

1.1.3 Interlude sur les équations de renouvellement

Dans le cas d'un processus de Hawkes linéaire, tel que définit à la définition 1.1.14, l'espérance de l'intensité conditionnelle du processus satisfait une **équation de renouvellement**. En effet, pour tout $t \geq 0$,

$$f(t) := \mathbb{E}[\lambda^*(t)] = \mathbb{E}\left[\lambda + \int_{(0,t)} h(t-s)N(ds)\right] = \lambda + \int_0^t h(t-s)f(s)ds. \quad (1.1.6)$$

Les équations de renouvellement sont des cas particuliers des **équations intégrales de Volterra** (voir à ce sujet le livre de Brunner [11]). Sous des hypothèses de régularité de la fonction h , on peut montrer que l'équation

$$\forall t \geq 0, \quad f(t) = \lambda + \int_0^t h(t-s)f(s)ds \quad (1.1.7)$$

possède une unique solution. En général, il n'est pas possible d'écrire explicitement la solution de (1.1.7). Cependant, on peut en donner une approximation numérique, détaillée par exemple dans le livre de Delves et Walsh [18].

Dans le cas d'un processus de Hawkes avec inhibition d'intensité λ^* , l'égalité (1.1.6) ne tient plus. On a cependant, d'après l'inégalité de Jensen :

$$f(t) := \mathbb{E}[\lambda^*(t)] \geq \left(\lambda + \int_0^t h(t-s)f(s)ds\right)_+.$$

Nous avons alors étudié les éventuelles solutions de l'équation non linéaire :

$$f(t) = \left(\lambda + \int_0^t h(t-s)f(s)ds\right)_+. \quad (1.1.8)$$

En les comparant avec les solutions de l'équation (1.1.7). Dans le cas général, cette équation ne possède pas de solution explicite, c'est pourquoi nous avons considéré une classe particulière de fonction h pour lesquelles la solution de l'équation (1.1.7) est explicite, à savoir les fonctions h de la forme :

$$\forall x \in \mathbb{R}, \quad h(x) = e^{-bx} (\gamma C_a(x) + \delta S_a(x)), \quad (1.1.9)$$

où

$$C_a(x) := \begin{cases} \cos(\sqrt{ax}) & \text{si } a \geq 0, \\ \cosh(\sqrt{-ax}) & \text{si } a < 0, \end{cases}, \quad S_a(x) := \begin{cases} \frac{\sin(\sqrt{ax})}{\sqrt{a}} & \text{si } a > 0 \\ \frac{\sinh(\sqrt{ax})}{\sqrt{-a}} & \text{si } a < 0 \\ 1 & \text{si } a = 0 \end{cases},$$

et $a, b, \gamma, \delta \in \mathbb{R}$.

Le chapitre 2 est dédié à l'étude numérique des solutions de l'équation non-linéaire (1.1.8) ainsi que leur comparaison avec les solutions de l'équation linéaire (1.1.7). Pour ces deux équations, un schéma numérique d'approximation des solutions est proposé, ainsi qu'une discussion autour de la comparaison entre ces solutions et la moyenne empirique des intensités conditionnelles des processus de Hawkes associés.

1.2 Processus de Hawkes discret

1.2.1 Introduction et définitions

Après avoir présenté le processus de Hawkes en temps continu, nous nous tournons maintenant vers son équivalent en temps discret. À notre connaissance, la littérature ne propose pas de définition précise d'un processus de Hawkes discret. Néanmoins, une approche consiste à considérer un processus de Hawkes en temps continu et à en donner une approximation discrète. Kirchner illustre cette méthode dans son article [54] en utilisant les séries temporelles INAR(∞) (*Integer-valued AutoRegressive*), où il décrit la discrétisation d'un processus de Hawkes linéaire. Nous résumons ici cette démarche.

Notation 1.2.1. Dans la suite du manuscrit, la lettre caligraphiée $\mathcal{P}(\cdot)$ décrit la loi de Poisson avec son paramètre.

Définition 1.2.2. Soient $\alpha_0, \alpha_1, \dots \geq 0$ des nombres réels positifs. On définit une suite de variables aléatoires $(\varepsilon_n)_{n \in \mathbb{Z}}$ i.i.d de loi $\mathcal{P}(\alpha_0)$, ainsi qu'une suite $(\xi_\ell^{(n,k)})$ où $n \in \mathbb{Z}, k \in \mathbb{N}^*, \ell \in \mathbb{N}$, de variables aléatoires i.i.d de loi $\mathcal{P}(\alpha_k)$, indépendantes de (ε_n) .

Une suite $(X_n)_{n \in \mathbb{Z}}$ est une **série temporelle INAR(∞)** si elle est solution du système d'équations stochastiques suivant :

$$\varepsilon_n = X_n - \sum_{k=1}^{\infty} \sum_{\ell=1}^{X_{n-k}} \xi_\ell^{(n,k)}, \quad n \in \mathbb{Z}.$$

Remarque 1.2.3. Si (X_n) est une série temporelle $\text{INAR}(\infty)$, on a alors pour tout $n \in \mathbb{Z}$:

$$X_n \mid X_{n-1}, X_{n-2}, \dots \sim \mathcal{P} \left(\alpha_0 + \sum_{k=1}^{\infty} \alpha_k X_{n-k} \right).$$

Le résultat principal de cet article est le

Théorème 1.2.4. Soit $\{N(t), t \geq 0\}$ un processus de Hawkes linéaire. On construit une famille spécifique de processus $\left\{ \left(X_n^{(\Delta)} \right)_{n \in \mathbb{Z}} \right\}_{\Delta > 0}$ $\text{INAR}(\infty)$. De plus, pour tout $\Delta > 0$ on définit le processus de comptage suivant :

$$N^{(\Delta)}((a, b]) := \sum_{n: \Delta n \in (a, b]} X_n^{(\Delta)}.$$

Alors, on a la convergence en loi : $N^{(\Delta)} \xrightarrow{\Delta \rightarrow 0} N$.

De ce fait, le processus de Hawkes linéaire peut être vu comme une version continue d'une série temporelle $\text{INAR}(\infty)$, et vice-versa. Notons qu'une approche similaire ainsi que des résultats analogues ont été obtenus pour les processus de Hawkes non linéaires par Huang et Khabou [46], lorsque le noyau du processus de Hawkes est de type Erlang.

Le modèle $\text{INAR}(\infty)$ est une extension du modèle de séries temporelles $\text{INAR}(p)$ où $p \in \mathbb{N}$. Ces modèles sont dit auto-régressifs, et sont particulièrement utiles pour les données discrètes où les observations sont des nombres entiers, comme le nombre d'événements survenant dans des intervalles de temps donnés. Lorsque la distribution sous-jacente est une loi de Poisson, ces séries temporelles s'inscrivent plus généralement dans le modèle **INGARCH** (pour *INteger-valued Generalized Auto-Regressive Conditional Heteroscedasticity models*).

Définition 1.2.5. Soit $p \in \mathbb{N}^*$ et $q \in \mathbb{N}$. Alors (\tilde{X}_n) est un processus **INGARCH**(p, q) si pour tout $n \geq 1$, conditionnellement à $\tilde{X}_{n-1}, \tilde{X}_{n-2}, \dots$, la variable aléatoire \tilde{X}_n suit une loi de Poisson de paramètre M_n , où

$$M_n := \lambda + \sum_{i=1}^p a_i \tilde{X}_{n-i} + \sum_{j=1}^q \beta_j M_{n-j},$$

où $\lambda > 0, a_1, \dots, a_p, \beta_1, \dots, \beta_q \geq 0$.

Ce processus, notamment étudié par Ferland *et al.* [26] et Fokianos [28], a la particularité de dépendre de façon aléatoire de ses réalisations passées mais également de ses intensités passées. Pour une revue détaillée de différents types de processus auto-régressifs de Poisson utilisés dans la littérature, nous renvoyons à Liu [61].

Dans la suite, nous allons étudier un analogue discret du processus de Hawkes continu, dont l'intensité dépend des points passés. Plus précisément, nous nous intéresserons à un cas particulier du processus **INGARCH** pour $q = 0$. Par ailleurs, afin de modéliser un effet inhibiteur dans notre système, nous considérerons ce processus

avec des paramètres a_1, \dots, a_p à valeurs réelles. Dans ce contexte, pour assurer la bonne définition du processus, nous ajusterons le paramètre M_n de la loi de Poisson de la définition 1.2.5 de manière à ce qu'il reste positif pour tout $n \in \mathbb{N}$. Pour cela, nous utiliserons à nouveau la fonction ReLU définie précédemment. Voici plus précisément le processus sur lequel portent nos travaux de recherche :

Définition 1.2.6. Soit $p \in \mathbb{N}^*$, $a_1, \dots, a_p \in \mathbb{R}$ et $\lambda > 0$. Le processus $(\tilde{X}_n)_{n \in \mathbb{N}}$ est un **processus de Hawkes discret avec une mémoire de taille p** si, étant donnée une condition initiale $(\tilde{X}_{-p+1}, \dots, \tilde{X}_0) \in \mathbb{N}^p$, pour tout $n \geq 1$,

$$\text{conditionnellement à } \tilde{X}_{n-1}, \dots, \tilde{X}_{n-p}, \quad \tilde{X}_n \sim \mathcal{P} \left(\left(a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda \right)_+ \right).$$

Dans cette thèse, on s'intéresse au comportement asymptotique de ce processus, lorsque n tend vers $+\infty$, en fonction des paramètres du modèle. Plus précisément, on souhaite classifier la stabilité du processus de la définition 1.2.6, c'est-à-dire, distinguer les cas où :

- Le processus est **instable**, dans le sens où :

$$\lim_{n \rightarrow +\infty} [\tilde{X}_{n+1} + \tilde{X}_{n+2} + \dots + \tilde{X}_{n+p}] = +\infty, \quad \text{ps.}$$

- Au contraire, le processus est **stable**, dans le sens où (\tilde{X}_n) converge en loi vers une variable aléatoire finie lorsque n tend vers $+\infty$.

Notons que dans [26], dans le cas linéaire, c'est-à-dire pour lequel $a_1, \dots, a_p \geq 0$, un critère nécessaire et suffisant de stabilité du processus (\tilde{X}_n) est donné par :

$$\sum_{i=1}^p a_i < 1. \quad (1.2.1)$$

Remarque 1.2.7. • La condition (1.2.1) rappelle immédiatement la condition (1.1.4) concernant le processus de Hawkes linéaire, nous confortant dans l'idée que la comparaison entre le modèle discret et continu est pertinente.

- Le critère (1.2.1) implique notamment que lorsque l'un des paramètres a_i est strictement supérieur à 1, alors le processus de Hawkes discret est instable, ce qui restreint fortement les paramètres admissibles : pour ce modèle, une condition nécessaire de stabilité est : $a_1, \dots, a_p < 1$. En reprenant les questionnements laissés ouverts concernant le processus de Hawkes continu, une question s'impose : si l'on relâche l'hypothèse de positivité des paramètres a_1, \dots, a_p , peut-on fixer un paramètre strictement supérieur à 1 et choisir les autres de telle manière à obtenir un processus stable ? De façon plus générale, quelles sont les valeurs des paramètres qui induisent un processus stable ou instable ?

Dans le cas général $p \in \mathbb{N}^*$ quelconque, les questions soulevées dans la remarque 1.2.7 semble être très dures à résoudre, et pour le moment hors de portée. Nous

sommes néanmoins parvenus à classifier le comportement de ce processus pour le cas $p = 2$, à donner de nombreuses conditions suffisantes de stabilité ou d'instabilité dans le cas $p = 3$, et à donner quelques résultats sur le cas $p \in \mathbb{N}^*$. Tout au long de ce manuscrit, nous nous attacherons à montrer que l'ajout de la non-linéarité et la prise en compte de l'effet de l'inhibition ont pour conséquences de nombreuses complexifications mathématiques, parfois contre-intuitives au premier abord.

Stratégie générale d'étude Pour étudier ces processus, la stratégie générale a été de se ramener à l'étude de la chaîne de Markov naturellement associée à (\tilde{X}_n) , à savoir :

$$X_n := (\tilde{X}_n, \dots, \tilde{X}_{n-p}). \quad (1.2.2)$$

En effet, considérer cette chaîne de Markov a permis de se placer dans un cadre théorique convenable : pour le processus (\tilde{X}_n) , la non-linéarité induite par la fonction ReLU fait perdre la structure de branchement naturellement vérifiée par la version linéaire du processus. Ainsi, du problème posé par la classification du comportement asymptotique du processus (\tilde{X}_n) , nous nous sommes ramenés au problème de la classification de la chaîne de Markov (X_n) , en cherchant à déterminer les paramètres du modèle pour lesquels (X_n) est transiente ou ergodique. On parlera donc indifféremment de transience et d'instabilité, ainsi que de récurrence/ergodicité et de stabilité. Dans le prochain paragraphe, nous détaillons le cadre théorique dans lequel nous avons effectué nos recherches, ainsi que les outils dont nous nous sommes servis pour prouver les résultats énoncés dans cette thèse.

1.2.2 Le cadre théorique développé : chaîne de Markov et critères de *drift*

1.2.2.1 Stratégie pour démontrer l'ergodicité de (X_n)

Afin de démontrer l'ergodicité de la chaîne de Markov (X_n) , nous avons utilisé de manière extensive les critères de dérive (*drift criteria*) via des fonctions de Lyapunov. Les critères de dérive sont de puissants outils qui ont été introduits par Foster en 1953 [31], et profondément étudiés et popularisés par Meyn et Tweedie [67] dans les années 90. Ces critères permettent de prouver la convergence de chaînes de Markov vers leur mesure invariante et fournissent des taux de convergence explicites. Nous utilisons dans cette thèse le formalisme développé par Douc, Moulines, Priouret et Soulier [19], plus adapté à l'étude de notre chaîne (X_n) . Nous en rappelons ici les définitions et résultats importants, qui seront beaucoup utilisées tout au long de ce manuscrit.

Définition 1.2.8 (7.1.1 de [19]). *Soit \mathcal{X} un espace d'états quelconque et X une chaîne de Markov sur \mathcal{X} . On note P^n la n -ième matrice de transition de X .*

*Un état $y \in \mathcal{X}$ est **accessible** si pour tout $x \in \mathcal{X}$, il existe un entier $n \geq 1$ tel que $P^n(x, y) > 0$.*

On suppose maintenant que l'espace d'états \mathcal{X} est discret. On dit que :

- La chaîne X est **irréductible** (*irreducible*) si elle admet un état accessible.
- La chaîne X est **fortement irréductible** (*strongly irreducible*) si tous les états sont accessibles.

Remarque 1.2.9. • Dans notre cas, au vu de la définition (1.2.2), puisqu'une variable aléatoire de Poisson a toujours une probabilité non nulle d'être égale à 0, l'état $\mathbf{0} := (0, \dots, 0) \in \mathbb{N}^p$ est accessible puisque, quelque soit $x \in \mathbb{N}^p$,

$$P^p(x, \mathbf{0}) > 0.$$

La chaîne de Markov (X_n) est donc irréductible au sens de la définition 1.2.8.

- En général dans la littérature, la définition d'irréductibilité d'une chaîne de Markov correspond à celle de forte irréductibilité de la définition 1.2.8. On prendra donc garde ici à ne pas confondre ces deux définitions. La distinction est importante : la chaîne de Markov (X_n) n'est effectivement pas fortement irréductible en général. On peut trouver une preuve de ce fait pour le cas $p = 2$ dans la section 3.B, qui consiste à exhiber un état non accessible.

Nous présentons maintenant plus spécifiquement les critères de *drift* pour des chaînes de Markov ergodiques. Premièrement, nous introduisons la notion d'ensemble d'états *petite* :

Définition 1.2.10 (9.4.1 de [19]). *Soit \mathcal{X} un espace d'états discret. Un ensemble d'états $C \subset \mathcal{X}$ est dit **petite** s'il existe un état $x_0 \in \mathcal{X}$ et une distribution de probabilité $(p_n)_{n \in \mathbb{N}}$ sur \mathbb{N} tels que*

$$\inf_{x \in C} \sum_{n \in \mathbb{N}} p_n P^n(x, x_0) > 0,$$

où $P^n(x, x_0)$ est la probabilité de transition en n étapes de x à x_0 .

Un ensemble d'états *petite* est donc un ensemble sur lequel on peut minorer uniformément la probabilité d'atteindre un certain état $x_0 \in \mathcal{X}$, de telle sorte que cette dernière soit strictement positive. Dans notre cas, l'état x_0 de la définition 1.2.10 sera sans surprise au vu de la remarque 1.2.9 : $\mathbf{0} = (0, \dots, 0) \in \mathbb{N}^p$. Bien que l'on ait une probabilité strictement positive d'atteindre en temps fini cet état, rien n'assure *a priori* la conservation de l'inégalité stricte dans le passage à l'inf de la définition 1.2.10 lorsque l'ensemble C est infini. On a donc une propriété forte sur les états constituant un tel ensemble *petite*.

Nous poursuivons avec la définition du *drift* d'une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$.

Définition 1.2.11. Soit \mathcal{X} un espace d'états discret, et X une chaîne de Markov irréductible sur \mathcal{X} . Soit $V : \mathcal{X} \rightarrow [1, \infty)$ une fonction, $\varepsilon \in (0, 1]$, $K < \infty$ deux nombres réels et $C \subset \mathcal{X}$ un ensemble d'états. On dit que la **condition de dérive** (drift condition) $D(V, \varepsilon, K, C)$ est satisfaite si

$$\Delta V(x) := \mathbb{E}_x[V(X_1) - V(X_0)] \leq -\varepsilon V(x) + K\mathbf{1}_C, \quad (1.2.3)$$

où $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot | X_0 = x]$. La fonction ΔV ainsi définie est appelée **drift** en V de la chaîne X . Si la condition de dérive est satisfaite, on dit que la fonction V est une **fonction de Lyapounov** pour la chaîne X .

Ainsi, pour une fonction V définie sur l'espace d'états \mathcal{X} , le *drift* en V de la chaîne (X_n) est une fonction donnant la tendance moyenne de la chaîne sur un pas de temps, au travers de la fonction V , pour tous les états $x \in \mathcal{X}$. La notion de *drift* est à ce titre souvent intuitée comme une "dérivée probabiliste", au vu de l'analogie frappante entre les critères de *drift* et les théorèmes de stabilité des systèmes dynamiques utilisant des fonctions de Lyapounov. Comme la fonction V est à valeurs positives, la condition (1.2.3) signifie qu'en dehors d'un certain ensemble C , le drift en V de (X_n) peut être majoré par une certaine quantité strictement négative dépendant de V , signifiant une tendance de cette dernière à quitter, en moyenne, ces états. On s'attend donc, selon la forme de l'ensemble C , à retrouver une propriété de récurrence, voire d'ergodicité de la chaîne (X_n) sous la condition (1.2.3). Cette intuition est confirmée par la proposition suivante, où l'on utilise la définition 1.2.10 d'un ensemble *petite* :

Proposition 1.2.12. Supposons que la condition de dérive $D(V, \varepsilon, K, C)$ est vérifiée pour certains V, ε, K et C comme ci-dessus et supposons que C est petite. Alors il existe $\beta > 1$ et une mesure de probabilité π sur \mathcal{X} , telle que, pour tout état initial $x \in \mathcal{X}$,

$$\beta^n \times \sum_{y \in \mathcal{X}} V(y) |\mathbb{P}_x(X_n = y) - \pi(y)| \rightarrow 0, \quad n \rightarrow \infty.$$

En particulier, pour tout état initial $x \in \mathcal{X}$,

$$\beta^n d_{VT}(\text{Loi}(X_n), \pi) \rightarrow 0, \quad n \rightarrow \infty,$$

où $d_{VT}(\cdot, \cdot)$ est la distance en variation totale des deux mesures. On a alors que π est une mesure de probabilité invariante pour la chaîne de Markov $(X_n)_{n \geq 0}$, et que cette dernière est ergodique.

Lorsqu'il s'agira de démontrer l'ergodicité de la chaîne (X_n) , on cherchera à appliquer la proposition 1.2.12. Néanmoins, dans le cas général en pratique, il n'existe pas de méthode permettant de construire explicitement la fonction V , les réels K, ε ainsi que l'ensemble *petite* C convenables. De ce fait, durant notre travail de recherche autour de ces processus, un temps considérable a été consacré à la recherche des éléments convenables décrits ci-dessus.

1.2.2.2 Stratégie pour démontrer la transience de (X_n)

Concernant les preuves de la transience de la chaîne de Markov (X_n) , grâce à de nombreuses simulations numériques, réalisées en amont, nous avons repérés que les trajectoires de la chaîne (X_n) semblaient se trouver "piégées" à l'intérieur de certains types d'états bien précis. Ces observations nous ont poussées à nous appuyer sur le résultat suivant :

Lemme 1.2.13. *Soit $p \geq 2$ un entier, S_1, S_2, \dots une séquence de sous-ensembles de \mathbb{N}^p et $0 < m_1 < m_2 < \dots$ une suite croissante d'entiers. On notera $\mathbf{0}$ l'état $(0, \dots, 0) \in \mathbb{N}^p$. Supposons que :*

1. *Sur l'événement $\bigcap_{n \geq 1} X_{m_n} \in S_n$, on a $X_n \neq 0$ pour tout $n \geq 1$,*
2. *$\mathbb{P}_{\mathbf{0}}(X_{m_1} \in S_1) > 0$ et pour tout $n \geq 1$ et tout $x \in S_n$, on a $\mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) > 0$.*
3. *Il existe $(p_n)_{n \geq 1}$ prenant des valeurs dans $[0, 1]$ et tel que $\sum_{n \geq 1} (1 - p_n) < \infty$, tel que*

$$\forall n \geq 1 : \forall x \in S_n : \mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) \geq p_n.$$

Alors la chaîne de Markov $(X_n)_{n \geq 0}$ est transiente.

Pour les cas de transience, il s'est donc essentiellement agi de déterminer les suites de nombres (m_n) et d'ensembles (S_n) qui convenaient. Notamment, la propriété de concentration de la loi de Poisson nous a permis d'utiliser des inégalités type Bienaymé-Tchebychev afin de parvenir à nos fins.

Remarque 1.2.14. • Il existe également des critères de *drift* permettant de démontrer la transience d'une chaîne de Markov.

En reprenant les notations de la définition 1.2.11, on dit qu'une fonction $V : \mathcal{X} \rightarrow \mathbb{R}_+$ est **sur-harmonique** si :

$$\forall x \in \mathcal{X}, \quad \Delta V(x) \leq 0. \tag{1.2.4}$$

Dans le livre de Douc *et al.* [19] (on trouvera les détails dans le chapitre 10), il est prouvé que, si (X_n) est irréductible et récurrente, alors toute fonction sur-harmonique positive est constante. Ainsi, si l'on exhibe une fonction V strictement positive et non-constante sur \mathcal{X} , satisfaisant (1.2.4), alors (X_n) est transiente.

- On pourra trouver un grand nombre de critères de *drift* dans le livre de Fayolle *et al.* [25], avec notamment des méthodes pour exhiber de telles fonctions dans certains cas particuliers.

Nous détaillons maintenant les éléments présents dans les différents chapitres de ce cette thèse, en insistant sur les résultats obtenus ainsi que les contributions apportées.

1.2.3 Plan de la thèse et contributions

1.2.3.1 Le cas $p = 2$

Le chapitre 3 de cette thèse est dédié à l'étude du processus (\tilde{X}_n) pour le cas $p = 2$, dont le comportement asymptotique a été entièrement classifié, à l'exception des cas frontières.

Plus précisément, on étudie le processus défini comme suit :

Définition 1.2.15. *Soit $a, b \in \mathbb{R}$, $\lambda > 0$, et $\tilde{X}_0, \tilde{X}_{-1} \in \mathbb{N}$. Alors pour tout $n \geq 1$,*

$$\tilde{X}_n \sim \mathcal{P} \left(\left(a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + \lambda \right)_+ \right).$$

Les paramètres a et b représentent donc des poids que l'on affecte sur les réalisations passées du processus \tilde{X}_n , respectivement la dernière et l'avant-dernière. Ainsi, une valeur négative de a aura pour conséquence que la réalisation la plus récente de \tilde{X}_n inhibera le processus, tandis qu'une valeur positive de b amènera l'avant-dernière réalisation à l'exciter, et vice-versa. Ces remarques concernant la modélisation de ces excitations/inhibitions ainsi que leurs temporalités par rapport aux réalisations passées de (\tilde{X}_n) sont primordiales pour la suite de cette thèse et vont jouer un rôle déterminant pour la compréhension du comportement asymptotique de ce processus.

Comme annoncé dans la section précédente, nous sommes d'abord placés dans un cadre plus propice à l'étude du processus (\tilde{X}_n) en considérant la chaîne de Markov (X_n) naturellement associée à ce processus, définie par :

$$\forall n \geq 0, \quad X_n := (\tilde{X}_n, \tilde{X}_{n-1}). \quad (1.2.5)$$

On cherche alors à déterminer les paramètres $a, b \in \mathbb{R}$ pour lesquels la chaîne (X_n) est récurrente positive (ergodique), ou transiente. On verra que dans le cas ergodique, on parvient même à montrer l'ergodicité géométrique de la chaîne, c'est-à-dire que la loi de (X_n) converge à une vitesse exponentielle vers sa distribution stationnaire. Le résultat principal de ce chapitre est le théorème 1.2.16, séparant l'espace des paramètres $(a, b) \in \mathbb{R}^2$ en deux régions \mathcal{R} et \mathcal{T} , pour lesquelles la chaîne de Markov (X_n) est respectivement ergodique et transiente :

Théorème 1.2.16. *Soit la fonction définie par :*

$$b_c(a) = \begin{cases} 1 & a \leq 0 \\ 1 - a & a \in (0, 2) \\ -\frac{a^2}{4} & a \geq 2 \end{cases}$$

ainsi que les ensembles de \mathbb{R}^2 :

$$\begin{aligned} \mathcal{R} &= \{(a, b) \in \mathbb{R}^2 : b < b_c(a)\} \\ \mathcal{T} &= \{(a, b) \in \mathbb{R}^2 : b > b_c(a)\}. \end{aligned}$$

- Si $(a, b) \in \mathcal{R}$, alors la chaîne de Markov (X_n) est géométriquement ergodique.
- Si $(a, b) \in \mathcal{T}$, alors la chaîne de Markov (X_n) est transiente.

En plus de donner une illustration des ensembles \mathcal{R} et \mathcal{T} décrits dans le théorème 1.2.16, la figure 1.4 fournit des trajectoires typiques de la chaîne de Markov selon le choix des paramètres du modèle.

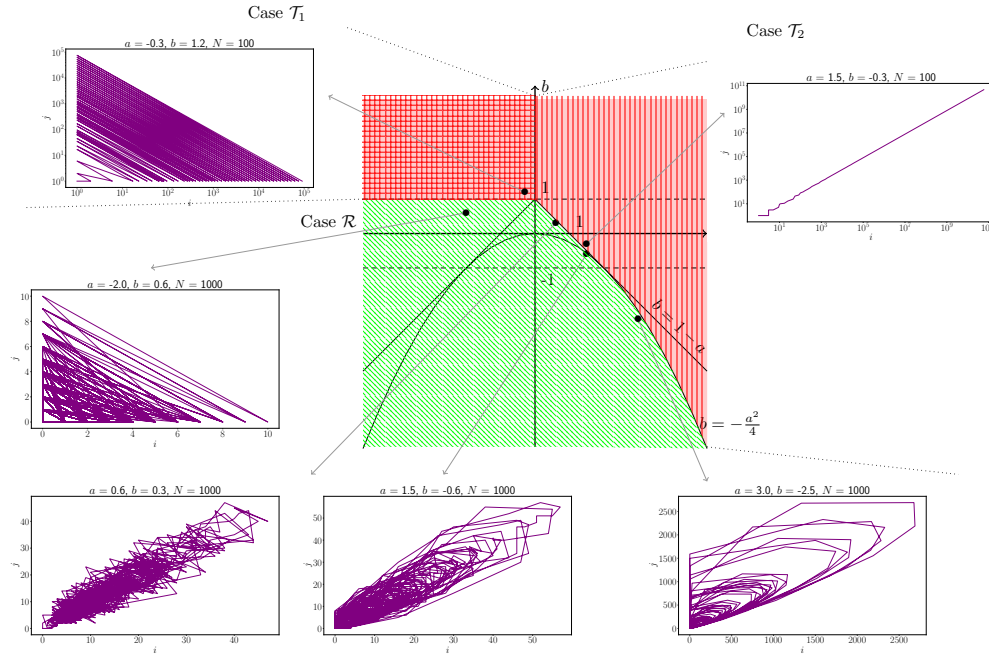


FIGURE 1.4 – Illustration du théorème 1.2.16. Les différentes trajectoires affichées sont celles de la chaîne de Markov $(X_n) \in \mathbb{R}^2$, pour lesquelles deux points consécutifs sont reliés.

Du théorème 1.2.16 découle un phénomène étonnant et contre-intuitif à première vue. Comme on l’a vu précédemment (voir (1.2.1)), dans le cas linéaire (*i.e.*, si $a, b \geq 0$), un critère nécessaire et suffisant de stabilité lorsque $p = 2$ est donné par $a + b < 1$: dans ce cas, les rôles de a et b sont symétriques. D’après le théorème 1.4, on perd cette propriété : dans le cas non-linéaire, la modélisation de l’inhibition dans le processus (\tilde{X}_n) de la définition 1.2.15 induit une asymétrie dans les rôles des paramètres du modèle.

De plus, une conséquence immédiate du théorème 1.2.16 est :

- Pour tout $a > 1$, il est possible de sélectionner le paramètre b de telle sorte que la chaîne de Markov soit ergodique. Il s’agira alors de prendre $b < 0$ suffisamment petit.
- À l’inverse, dès lors que $b > 1$, quelque soit la valeur du paramètre a , la chaîne (X_n) est transiente.

Concernant le processus (\tilde{X}_n) , ce qui vient d'être décrit est à mettre en regard de la remarque 1.2.7. Il est en effet possible de passer outre la condition nécessaire de stabilité du processus (\tilde{X}_n) (i.e. $a, b < 1$) lorsqu'on s'autorise à considérer des valeurs négatives pour les paramètres du modèle, mais cela ne vaut pas pour tous les paramètres. D'un point de vue modélisation, cela signifie que les temporalités de l'excitation et de l'inhibition induites par le choix des paramètres a et b sont primordiales pour saisir le comportement asymptotique de ce processus : une excitation sur les événements récents peut être compensée par une inhibition sur les événements plus anciens, mais pas l'inverse.

Afin de démontrer le théorème 1.2.16, nous avons découpé l'espace des paramètres en sous-cas, en traitant chacun d'entre eux indépendamment. Le découpage en question a été motivé par de nombreuses expérimentations numériques qui nous ont permis de constater que le processus possédait une très grande variété de comportement selon les paramètres a, b choisis, comme affiché sur la figure 1.4. Ces expérimentations ont été d'une grande aide pour conjecturer dans un premier temps les résultats du théorème 1.2.16, mais également pour déterminer les outils et méthodes que nous pouvions utiliser pour les démontrer.

Preuves des cas récurrents Comme annoncé dans la section 1.2.2.1, nous avons utilisé des critères de *drift via* des fonctions de Lyapounov pour mener à bien les démonstrations des cas récurrents. Dans ce but, la zone délimité en vert (l'ensemble \mathcal{R} du théorème 1.2.16) a été redécoupé en trois sous-ensemble de la manière représentée dans la figure 1.5.

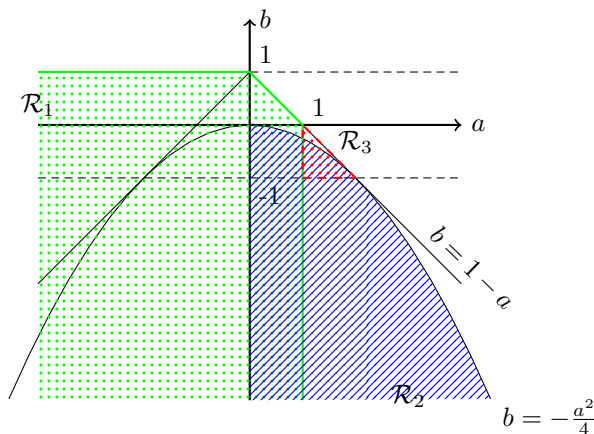


FIGURE 1.5 – Recouvrement de l'ensemble \mathcal{R} par trois sous-ensembles $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$.

Pour chacun de ces sous-cas, la difficulté a résidé dans la détermination des éléments convenables pour appliquer la proposition 1.2.12. Plus particulièrement, un temps considérable a été nécessaire pour trouver les fonctions de Lyapounov. En effet, à notre connaissance, il n'existe pas de méthode générale permettant de construire des fonctions de Lyapounov, excepté pour quelques cas décrits par exemple dans [25]. Nous donnons ici les fonctions de Lyapounov dont nous nous

sommes servis afin de prouver l'ergodicité de la chaîne de Markov (X_n) dans les cas sus-nommés.

- Cas $\mathcal{R}_1 = \{(a, b) \in \mathbb{R}^2, a, b < 1 \text{ et } a + b < 1\}$. Dans ce cas, on s'attend intuitivement à ce que le processus (\tilde{X}_n) soit stable. En effet, dans le cas linéaire, la condition $a + b < 1$ correspond au critère de stabilité du processus. Ici, on conserve cette condition tout en autorisant les paramètres a, b à prendre des valeurs négatives : on ne fait que rajouter de l'inhibition dans un système déjà stable. Ici, la fonction de Lyapounov utilisée est une forme linéaire :

$$V : (i, j) \in \mathbb{N}^2 \mapsto \alpha i + \beta j + 1,$$

où $\alpha, \beta \in \mathbb{R}_+$ dépendent des paramètres a, b .

- Cas $\mathcal{R}_2 = \{(a, b) \in \mathbb{R}^2, a \geq 0 \text{ et } b < -a^2/4\}$. On utilise une fonction quotient :

$$V : (i, j) \in \mathbb{N}^2 \mapsto \frac{i}{j+1} + 1.$$

Dans ce cas, il est notable que la fonction de Lyapounov utilisée ne dépend pas des paramètres du modèle, et convient pour tout $(a, b) \in \mathcal{R}_2$.

- Cas $\mathcal{R}_3 = \{(a, b) \in \mathbb{R}^2, 1 < a < 2 \text{ et } -1 < b < 1 - a\}$. Ici, on utilise une forme quadratique du type :

$$V : (i, j) \in \mathbb{N}^2 \mapsto \alpha i^2 + \beta j^2 + \gamma ij + 1,$$

où $\alpha, \beta, \gamma \in \mathbb{R}$ dépendent des paramètres a, b et satisfont $4\alpha\beta - \gamma^2 > 0$. L'utilisation d'une forme quadratique en tant que fonction de Lyapounov est classique comme on peut le constater dans [25].

Au vu de ces différents cas, notons que dans le cadre de notre travail de recherche, il a été plus commode de déterminer les fonctions de Lyapounov pour les cas \mathcal{R}_1 et \mathcal{R}_3 : en effet, en prenant des fonctions génériques que l'on peut paramétrer, il a été plus aisé de déterminer des conditions nécessaires portant sur de tels paramètres afin d'obtenir une fonction V convenable. A l'inverse, il été plus délicat de construire la fonction V du cas \mathcal{R}_2 malgré son apparente simplicité.

Concernant les ensembles *petite* utilisés afin d'appliquer la proposition 1.2.12, ces derniers sont de natures différentes selon les cas. Par exemple pour le cas \mathcal{R}_1 , l'ensemble *petite* considéré ne se réduit qu'à un ensemble fini de points. Cependant, pour les cas \mathcal{R}_2 et \mathcal{R}_3 , les ensembles *petite* utilisés sont des ensembles infinis, pour lesquels la propriété définissant de tels ensembles a été une nouvelle contrainte. Dans ces cas, une première observation a été que l'ensemble

$$A := \{(i, j) \in \mathbb{N}^2, s_{ij} = 0\} = \{(i, j) \in \mathbb{N}^2, ai + bj + \lambda \leq 0\}$$

est un ensemble *petite*. Afin de fournir une explication intuitive de ce fait, plaçons nous dans le cas \mathcal{R}_2 et représentons en rouge l'ensemble A . On fournit également une trajectoire de la chaîne de Markov (X_n) dans ce cas.

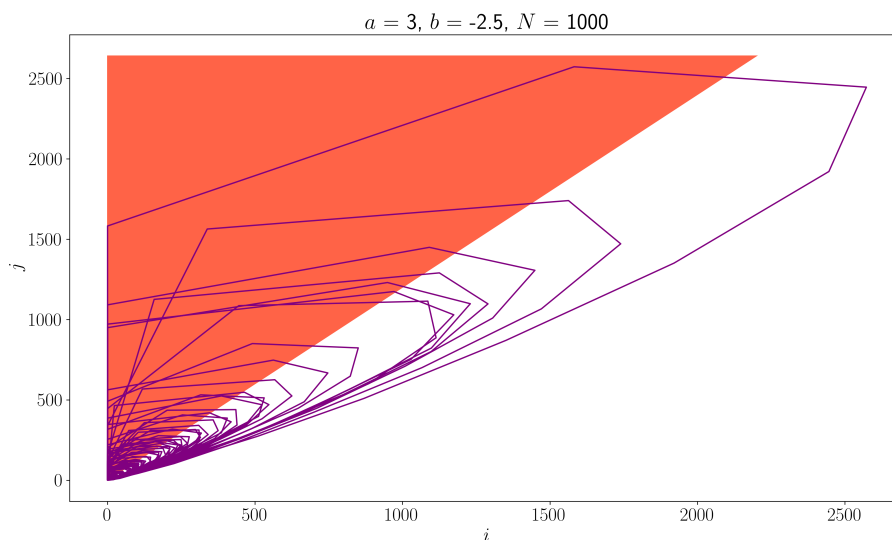


FIGURE 1.6 – Une trajectoire de $N = 1000$ pas de la chaîne de Markov $(X_n) \in \mathbb{N}^2$ dans le cas où $a = 3, b = -2.5$. En rouge, l'ensemble $A \subset \mathbb{N}^2$. Par définition de (X_n) , partant de cet ensemble d'états, le prochain pas de la chaîne se trouve sur l'axe j du graphe dessiné, le paramètre de la loi de Poisson de la définition du processus étant $s_{ij} = 0$.

Si la chaîne atteint l'un des états (i, j) de l'ensemble A , alors la prochaine réalisation de la loi de Poisson associée sera 0, et pour peu que j soit suffisamment grand, on s'attend à ce que la chaîne atteigne l'état accessible $\mathbf{0} = (0, 0)$ en deux temps partant de l'ensemble A , avec une probabilité strictement positive : c'est un ensemble *petite*.

La preuve de l'ergodicité de la chaîne de Markov (X_n) dans ce cas repose alors sur le fait que la chaîne possède un *drift* strictement négatif en dehors de l'ensemble A , dénotant une tendance de (X_n) à rejoindre l'ensemble A en temps fini, avant d'atteindre l'état $\mathbf{0}$, et ce une infinité de fois : la chaîne effectue des excursions de tailles finies puis revient en l'origine.

Preuve des cas transients Pour la démonstration de la transience de la chaîne de Markov (X_n) , nous avons partitionné l'ensemble \mathcal{T} en deux sous-ensembles \mathcal{T}_1 et \mathcal{T}_2 comme indiqué sur la figure 1.4. Pour chacun de ces sous-cas, nous avons utilisé le lemme 1.2.13, pour lequel il a fallu déterminer la suite d'ensemble $(S_n)_{n \in \mathbb{N}}$ qui convenait. Là encore, nous avons tiré une précieuse aide de la simulation numérique de la chaîne (X_n) : en observant plusieurs trajectoires de celle-ci dans les cas \mathcal{T}_1 et \mathcal{T}_2 , nous avons été en mesure d'identifier des espaces d'états dans lesquels la chaîne semblait coincée à partir d'un certain temps :

- Cas $\mathcal{T}_1 = \{(a, b) \in \mathbb{R}^2, a < 0, b > 1\}$. Ici, nous avons utilisé le lemme 1.2.13 en

posant pour $1 < r < b$:

$$m_n := 2n - 1, \quad S_n = \{(i, 0), i \geq r^n\}.$$

Afin de motiver ce choix, supposons que la chaîne (X_n) ait atteint un état $X_k = (0, i)$. Alors son prochain pas sera $X_{k+1} = (j, 0)$, où j est la réalisation d'une variable aléatoire de loi de Poisson de paramètre $bi + \lambda > i$, puisque $b > 1$. En moyenne, on s'attend donc à ce que $j > i$. Pour le pas suivant, si on suppose que j est suffisamment grand, le paramètre de la loi de Poisson est $aj + \lambda < 0$, puisque $a < 0$, et donc le prochain état de la chaîne de Markov sera $X_{k+2} = (0, j)$, où $j > i$ avec grande probabilité. Ainsi, on voit qu'en deux pas, la chaîne est cantonnée à des états du type $(0, i)$, avec en moyenne une croissance de sa deuxième coordonnée par un facteur multiplicatif $b > 1$.

La méthode employée pour démontrer la transience dans le cas \mathcal{T}_1 fournit également une preuve du caractère oscillatoire des trajectoires de la chaîne (X_n) dans ce cas, comme illustré sur la figure 1.4. A partir d'un certain temps, les trajectoires de la chaîne consistent en une succession de points appartenant aux axes i et j , alternant entre ces deux derniers à chaque pas de temps.

• Cas $\mathcal{T}_2 = \{(a, b) \in \mathbb{R}^2, 0 < a < 2 \text{ et } a+b > 1\} \cup \{(a, b) \in \mathbb{R}^2, a \geq 2 \text{ et } a^2+4b > 0\}$. Pour ce cas, nous avons tiré parti de la comparaison de notre processus avec la suite récurrente affine déterministe qui lui est naturellement associée, à savoir :

$$\forall n \geq 0, y_{n+1} = ay_n + by_{n-1} + \lambda. \quad (1.2.6)$$

Le polynôme caractéristique de cette suite récurrente est noté

$$P(X) := X^2 - aX - b.$$

Pour $(a, b) \in \mathcal{T}_2$, les racines du polynôme P sont simples et réelles, et la plus grande, notée θ , vérifie $\theta > 1$. Il est notoire que dans ce cas, les solutions de la récurrence (1.2.6) divergent et ne sont pas bornées. Nous avons alors traité distinctement le cas \mathcal{T}_2 selon le signe du paramètre b . En effet, lorsque $b > 0$, le paramètre de la loi de Poisson définissant le processus (\tilde{X}_n) est toujours positif : on se trouve en fait dans le cas linéaire. On s'attend alors à ce que le processus (\tilde{X}_n) se comporte de façon similaire aux solutions de (1.2.6), ayant une croissance exponentielle à chaque pas de temps. C'est pourquoi on utilise le lemme 1.2.13 avec :

$$\forall n \geq 1, \quad m_n := n, \quad S_n := \{(i, j) \in \mathbb{N}^2, i \geq r^n, j \geq r^{n-1}\}, \text{ où } 1 < r < \theta.$$

Afin de démontrer que la chaîne se trouvait bien dans les ensembles S_n décrits précédemment avec grande probabilité, nous avons utilisé une borne inférieure du paramètre, aléatoire et dépendant des réalisations passées, de la loi de Poisson définissant le processus (\tilde{X}_n) . Lorsque $b < 0$, cette minoration a été plus délicate et nous avons dû montrer un résultat plus fort pour parvenir à nos fins. En effet, nous avons utilisé plus spécifiquement que le facteur multiplicatif dans la croissance exponentielle des solutions de 1.2.6 est exactement θ . Ainsi, nous avons posé pour $\varepsilon > 0$

suffisamment petit :

$$\forall n \geq 1, \quad m_n := n, \quad S_n := \left\{ (i, j) \in \mathbb{N}^2, i \geq r^n, j \geq r^{n-1}, \left| \frac{i}{j} - \theta \right| < \varepsilon \right\},$$

où $1 < r < \theta$.

1.2.3.2 Le cas $p = 3$

Forts de ces observations et de ces résultats, nous avons poursuivi l'étude du processus (\tilde{X}_n) en augmentant la mémoire de celui-ci. Nous espérons que l'étude réussie du cas $p = 2$ nous donnerait plus de facilités pour l'étude de (\tilde{X}_n) dans le cas $p = 3$, notamment dans la détermination de fonctions de Lyapounov adaptées à ce cas. En effet, nous avons l'intuition que les résultats susceptibles d'être démontrés dans le cas $p = 3$ seraient des extensions naturelles des résultats du chapitre 2, et que leur démonstrations nécessiteraient d'employer les mêmes outils et méthodes.

L'étude du processus de Hawkes discret dans le cas $p = 3$ s'est finalement avérée plus difficile que ce que nous anticipions, l'augmentation de la mémoire du processus résultant en une forte complexification de son étude analytique. Nous sommes néanmoins parvenus à établir des résultats intéressants et très encourageants pour la suite de ce travail, consistant en des critères suffisants de stabilité et d'instabilité du processus. Cependant, nous n'avons pas réussi à classifier intégralement le comportement asymptotique de ce processus : nous anticipons que cette classification complète est difficile et nécessite un travail très minutieux, dans le sens où le processus (\tilde{X}_n) avec une mémoire de longueur 3 possède une variété plus large de comportement selon le choix des paramètres du modèle. L'identification de ces différents types de comportements et la détermination des outils nécessaires à les traiter convenablement s'inscrit certainement dans un projet de longue durée.

Synthèse des résultats obtenus Dans ce chapitre, on étudie le processus (\tilde{X}_n) suivant :

Définition 1.2.17. *Soit $a, b, c \in \mathbb{R}$, $\lambda > 0$, et $\tilde{X}_0, \tilde{X}_{-1}, \tilde{X}_{n-2} \in \mathbb{N}$. Alors pour tout $n \geq 1$,*

$$\tilde{X}_n \sim \mathcal{P} \left(\left(a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + c\tilde{X}_{n-3} + \lambda \right)_+ \right).$$

Nous énumérons maintenant les différents résultats que nous avons démontré concernant la stabilité et l'instabilité de ce processus. De manière analogue au chapitre 3, nous avons étudié la chaîne de Markov (X_n) définie par :

$$\forall n \geq 0, \quad X_n := (\tilde{X}_n, \tilde{X}_{n-1}, \tilde{X}_{n-2}). \quad (1.2.7)$$

Comme précédemment, nous parlerons indifféremment de la stabilité/l'instabilité du processus (\tilde{X}_n) et l'ergodicité/transience de la chaîne (X_n) , respectivement.

Le résultat principal de ce chapitre, donnant un critère suffisant d'ergodicité de (X_n) , s'appuie sur le discriminant du polynôme P de la définition 1.2.18.

Définition 1.2.18. Soit $(a, b, c) \in \mathbb{R}^3$, on définit le polynôme P de la façon suivante :

$$P(X) := X^3 - aX^2 - bX - c,$$

ainsi que son *discriminant* :

$$\text{Disc}(P) := a^2b^2 + 4b^3 - 4a^3c - 18abc - 27c^2.$$

Le discriminant d'un polynôme donne des informations sur le nombre de racines complexes de celui-ci. Plus précisément, pour le polynôme P :

- Si $\text{Disc}(P) \geq 0$, alors le polynôme P ne possède que des racines réelles, éventuellement multiples.
- Si $\text{Disc}(P) < 0$, alors le polynôme P possède deux racines complexes conjuguées, et une racine réelle simple.

Dans la figure 4.1, nous proposons une représentation graphique de la surface de \mathbb{R}^3 délimitée par $\text{Disc}(P) = 0$.

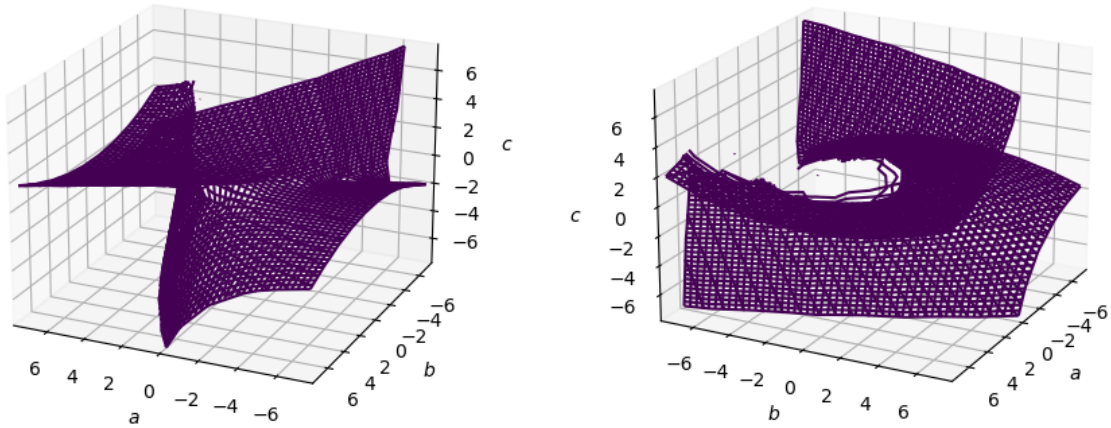


FIGURE 1.7 – Représentation graphique de la surface de \mathbb{R}^3 délimité par $\text{Disc}(P) = 0$. Les axes représentent les différentes valeurs des paramètres $a, b, c \in \mathbb{R}$. De plus, $\text{Disc}(P) < 0$ correspond à la région de \mathbb{R}^3 en dehors des deux surfaces.

Notons que le polynôme P est le polynôme caractéristique de la suite récurrente affine $(\tilde{u}_n)_{n \in \mathbb{N}}$ définie par :

$$\tilde{u}_n := a\tilde{u}_{n-1} + b\tilde{u}_{n-2} + c\tilde{u}_{n-3} + \lambda. \quad (1.2.8)$$

Le signe du discriminant de P donne de ce fait une information qualitative sur le comportement asymptotique des solutions de (1.2.8). En effet, une de l'espace vectoriel dirigeant l'espace affine des solutions de (1.2.8) est donnée par les racines du polynôme P . Ainsi, dans le cas où $\text{Disc}(P) < 0$, (*i.e.* deux racines de P sont complexes) de telles solutions auront un comportement oscillatoire, avec possiblement une divergence exponentielle. Dans ce dernier cas, ces solutions prendront donc au bout d'un certain temps des valeurs négatives. Or, dans la définition du processus

(\tilde{X}_n) , lorsque le paramètre de la loi de Poisson est négatif, celui-ci est tronqué à 0 via la fonction ReLU, et on peut espérer que cette troncature induise une certaine stabilité. Cette intuition est confirmée par le théorème 1.2.19.

Théorème 1.2.19. *Soit $a > 1$. Si $b, c < 0$ sont tels que $\text{Disc}(P) < 0$, alors la chaîne de Markov (X_n) est géométriquement ergodique.*

La preuve de ce résultat repose une fois de plus sur l'utilisation de la proposition 1.2.12 adaptée au cadre $p = 3$, pour laquelle nous avons déterminé une fonction de Lyapounov V ainsi qu'un ensemble *petite* C convenables qui sont proches des éléments utilisés pour le cas \mathcal{R}_2 du chapitre 3, puisqu'on a posé :

$$V : (i, j, k) \in \mathbb{N}^3 \mapsto \frac{i + \alpha j}{j + \alpha k + 1} + 1, \quad A := \{(i, j, k) \in \mathbb{N}^3, ai + bj + ck + \lambda < 0\},$$

où $\alpha > 0$ est l'unique racine réelle du polynôme $P(-X)$. L'ensemble C *petite* considéré est l'union de A et d'un ensemble fini de points de \mathbb{N}^3 .

On retrouve alors une fonction de Lyapounov qui est un quotient de formes linéaires, mais qui ici dépend des paramètres du modèle via la valeur de α . Afin de démontrer que la fonction V est une fonction de Lyapounov pour la chaîne de Markov (X_n) , nous nous sommes appuyés sur la théorie des formes quadratiques et notamment l'algorithme de décomposition de Gauss. En effet, le calcul du *drift* de cette fonction donne, pour $\varepsilon > 0$ et $(i, j, k) \in \mathbb{N}^3$ tels que $ai + bj + ck + \lambda > 0$:

$$\Delta V(i, j, k) + \varepsilon V(i, j, k) = \frac{q(i, j, k) + L(i, j, k)}{(i + \alpha j + 1)(j + \alpha k + 1)},$$

où q est une forme quadratique sur \mathbb{R}^3 , ici évaluée seulement en des points (i, j, k) à valeurs entières, et L un polynôme de degré 1 en (i, j, k) . Le choix de α comme étant la racine réelle du polynôme $P(-X)$ rend la forme quadratique q strictement négative pour les (i, j, k) considérés, satisfaisant ainsi la condition (1.2.3) de la proposition 1.2.12.

Remarque 1.2.20. Nous conjecturons que les hypothèses du théorème 1.2.19 peuvent être relâchées de la façon suivante :

Soit $a > 1$. Si $b \leq 1$ et $c < 0$ sont tels que $\text{Disc}(P) < 0$, alors la chaîne de Markov (X_n) est géométriquement ergodique.

Nous ne sommes pas parvenus à démontrer ce résultat, car nous nous heurtons à une difficulté technique concernant l'ensemble *petite* à considérer. Cependant, nous donnons dans le chapitre 4 une étude numérique détaillée illustrant cette conjecture.

Dans ce chapitre, nous donnons également deux critères suffisant d'instabilité du processus, décrit dans le théorème 1.2.21.

Théorème 1.2.21. *Si $(a, b, c) \in \mathbb{R}^3$ sont tels que :*

- $a, b < 0$ et $c > 1$, ou
- $ab + c < 0$ et $b > 1$,

alors la chaîne de Markov (X_n) est transiente.

La preuve du théorème 1.2.21 repose sur l'utilisation du lemme 1.2.13. Elle tire parti d'une observation analogue à celle effectuée dans le cas $p = 2$: pour le premier cas du théorème 1.2.21, en trois pas de temps, la chaîne (X_n) passe d'un état du type $X_k = (i, 0, 0)$ à un état $X_{k+3} = (j, 0, 0)$ où, avec grande probabilité, j est proche de $ci > i$. Ainsi, on pose, pour $1 < r < c$:

$$\forall n \geq 1, \quad m_n = 3n - 1, \quad S_n = \{(i, 0, 0) \in \mathbb{N}^3 : i \geq r^n\}.$$

De même, pour le deuxième cas du théorème 1.2.21, cette périodicité a lieu tous les deux pas de temps, sur des états du type $(0, i, 0)$, c'est pourquoi on utilise dans ce cas :

$$\forall n \geq 1, \quad m_n = 2n - 1, \quad S_n = \{(0, i, 0) \in \mathbb{N}^3 : i \geq r^n\}.$$

À travers les résultats de ce chapitre, nous retrouvons la propriété d'asymétrie des rôles des paramètres du modèle. En effet, d'après les théorèmes 1.2.19 et 1.2.21 :

- Pour tout $a > 1$, il est possible de choisir $b, c < 0$ suffisamment petits tels que le processus soit stable.
- Cependant, dès lors que $b > 1$ (resp. $c > 1$), alors quelque soit $a, c < 0$ (resp. $a, b < 0$), le processus (\tilde{X}_n) est instable.

Résultats pour le cas $p \in \mathbb{N}^*$ Nous avons également démontré le résultat suivant pour le processus de Hawkes discret avec une mémoire de taille $p \in \mathbb{N}^*$ arbitraire, introduit dans la définition 1.2.6 :

Théorème 1.2.22. *Soit $a_1, \dots, a_p \in \mathbb{R}$ tels que $(a_1)_+ + \dots + (a_p)_+ < 1$, et $\lambda > 0$ (on rappelle que pour tout $x \in \mathbb{R}$, $(x)_+ = \max(0, x)$). Alors la chaîne de Markov associée définie par (1.2.2) est géométriquement ergodique.*

Le théorème 1.2.22 étend naturellement la condition (1.2.1) au cas où les paramètres du modèle peuvent prendre des valeurs négatives, et donne un critère de stabilité semblable à (1.1.5) concernant le processus de Hawkes non-linéaire. C'est également l'extension du cas \mathcal{R}_1 du chapitre 3, tant pour son énoncé que pour sa démonstration. En effet, pour prouver le théorème 1.2.22, nous avons utilisé une fonction de Lyapounov V qui est une forme linéaire :

$$V : (i_1, \dots, i_p) \in \mathbb{N}^p \mapsto \alpha_1 i_1 + \dots + \alpha_p i_p + 1,$$

où $\alpha_1, \dots, \alpha_p > 0$ dépendent des paramètres $a_1, \dots, a_p \in \mathbb{R}$ du modèle.

1.2.3.3 Description de quelques cas critiques

Le chapitre 5 de cette thèse est un travail en cours portant sur les cas critiques de nos études précédentes. Cette démarche a notamment été motivée par la diversité des comportements limites observés au sein de cet ensemble de cas limites. Il nous a paru pertinent d'examiner en détail ces cas critiques, non seulement en raison

de la richesse des outils probabilistes nécessaires, mais aussi pour répondre à nos questions restées sans réponse. Un premier objectif a été de traiter le cas critique où les paramètres du modèle satisfont, pour $p \in \mathbb{N}^*$:

$$a_1 + \cdots + a_p = 1.$$

Pour mieux comprendre l'évolution et le comportement asymptotique de ce processus dans le cas critique, motivés par le comportement erratique de ce dernier dans le cas décrit, nous nous sommes intéressés à une version continue de (\tilde{X}_n) , obtenu par passage à la limite après un changement approprié d'échelle d'espace et de temps.

Dans notre cas, nous avons utilisé les résultats d'approximations de diffusions de Ethier et Kurtz [23].

Le cas $p = 2$

Dans un premier temps, on se concentre sur le cas où la mémoire du processus vaut 2. C'est-à-dire :

Définition 1.2.23. Soit $a, b > 0$ tels que $a + b = 1$, $\lambda > 0$. On définit, avec $X_{-1} = X_0 = 0$:

$$\forall n \geq 1, \quad X_n \sim \mathcal{P}(aX_{n-1} + bX_{n-2} + \lambda).$$

De plus, on note pour tout $n \in \mathbb{N}$:

$$\xi_n := aX_n + bX_{n-1} + \lambda \in \mathcal{F}_n.$$

L'objectif est d'analyser le comportement asymptotique, lorsque $n \rightarrow +\infty$, de $\left(X_n(t) = \frac{X_{\lfloor nt \rfloor}}{n}, t \geq 0\right)_{n \in \mathbb{N}^*}$, et de démontrer sa convergence en loi vers un processus de diffusion, solution d'une équation différentielle stochastique (EDS) particulière. Pour cela, nous avons utilisé un processus auxiliaire, noté $(Y_n(t), t \geq 0)_{n \in \mathbb{N}^*}$, défini par :

$$Y_n(t) := \frac{X_{\lfloor tn \rfloor} + bX_{\lfloor tn \rfloor - 1}}{n}.$$

En effet, le processus $(Y_n(t))$ possède des propriétés intéressantes grâce auxquelles on peut lui appliquer le théorème d'approximation des diffusions de [23]. De ce fait, on montre le théorème suivant :

Théorème 1.2.24. Soit $T > 0$. La suite $(Y_n(t), t \geq 0)_{n \in \mathbb{N}^*}$ converge en loi sur $[0, T]$ vers l'unique solution de l'EDS :

$$Y(t) = \lambda t + \frac{1}{\sqrt{1+b}} \int_0^t \sqrt{Y(s)} dW(s), \quad (1.2.9)$$

où W est un mouvement brownien.

Par suite, on obtient la convergence en loi de $(X_n(t), t \in [0, T])_{n \in \mathbb{N}}$ vers un processus stochastique $(X(t), t \geq 0)$ en remarquant que :

$$\forall t \geq 0, \forall n \in \mathbb{N}^*, \quad X_n(t) = \frac{1}{1+b} \left(Y_n(t) + b \frac{X_{[tn]} - X_{[tn]-1}}{n} \right),$$

et en montrant que :

$$\forall T > 0, \quad \sup_{t \leq T} \frac{X_{[tn]} - X_{[tn]-1}}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Nous obtenons alors que X est solution de l'EDS :

$$X(t) = \frac{\lambda}{1+b} t + \frac{1}{1+b} \int_0^t \sqrt{X(s)} dW(s),$$

ce qui nous permet d'effectuer une comparaison avec le travail récent de Horst et Xu [45] portant notamment sur les processus de Hawkes linéaire dans un cas critique analogue à celui considéré dans ce chapitre.

1.2.4 Perspectives

Nous pensons que la compréhension complète autour de la stabilité du processus de Hawkes discret peut avoir tout son intérêt dans l'identification de candidats potentiels pour des critères suffisants d'existence pour le processus de Hawkes à temps continu. Notamment, on a vu au cours de ces travaux que dans le cas discret, l'ordre des valeurs positives et négatives des paramètres joue un rôle primordial dans la stabilité du processus : on peut alors imaginer que l'inhibition induite par les valeurs négatives du noyau intérieur h dans le cas continu ne jouera un rôle stabilisant que si ces dernières sont localisées à un endroit précis du support de h .

Pour enrichir et rendre plus pertinente la comparaison entre les versions discrète et continue du processus de Hawkes, la poursuite de l'étude de la stabilité du processus de Hawkes discret avec une mémoire arbitraire $p \in \mathbb{N}^*$ constitue bien sûr la suite de ce projet, bien que l'on anticipe que ce travail constitue un projet particulièrement difficile. Bien que certains résultats soient d'ores et déjà démontrés pour de tels processus, cela reste des cas simples dans lesquels le rôle de l'inhibition n'est pas pris en compte, ou bien avec une mémoire courte comme ceux présentés dans cette thèse. Cela souligne le saut de difficulté induit par la modélisation de l'inhibition dans un système stochastique, même dans un cadre discret et simplifié.

Ainsi, même en temps discret, capturer l'influence de l'inhibition semble être un problème d'une grande difficulté dans le cas général. Le travail effectué dans cette thèse vise précisément à insuffler une première dynamique en ce sens : nous espérons que les techniques qui ont fonctionné dans nos cas pourront être exportées pour des processus à mémoire plus longue. Notamment, il semble que le comportement asymptotique de ce processus soit étroitement lié aux racines du polynôme

$$P(X) = X^p - a_1 X^{p-1} - \dots - a_{p-1} X - a_p,$$

dans un sens probablement plus fin et qui échappe pour le moment à notre compréhension complète.

Explorations numériques

Contents

2.1	Des équations fonctionnelles intégrales	33
2.1.1	Équation intégrale de Volterra de seconde espèce	34
2.1.2	Équation fonctionnelle non-linéaire	35
2.1.2.1	Approximation numérique de la solution de (2.1.3)	38
2.2	Une classe particulière de fonctions h	40
2.2.1	Comparaison des solutions de l'équation de renouvellement linéaire et non-linéaire	40
2.2.2	Détermination de fonctions h ne satisfaisant pas le critère suffisant de stabilité	42

Ce chapitre est consacré à l'étude numérique menée au début de cette thèse, se focalisant principalement sur l'approximation numérique de certaines équations intégrales fortement liées à l'intensité conditionnelle des processus de Hawkes linéaires. Dans une démarche exploratoire, nous abordons également le cas des processus de Hawkes avec inhibition. Nous rappelons d'abord la définition du processus étudié.

Définition 2.0.1. Soit $\lambda > 0$, et $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ une fonction mesurable, à valeurs réelles. Un **processus de Hawkes avec inhibition** $\{N(t), t \geq 0\}$ est un processus de comptage satisfaisant :

- $N(0) = 0$.
- L'intensité conditionnelle du processus est notée λ^* , et peut s'écrire :

$$\forall t \geq 0, \quad \lambda^*(t) := \left(\lambda + \int_0^t h(t-s)N(ds) \right)_+.$$

2.1 Des équations fonctionnelles intégrales

Comme nous l'avons fait remarquer dans l'introduction, dans le cas linéaire (*i.e.* $h \geq 0$), pour $\lambda > 0$ et une fonction $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ donnés, l'espérance de l'intensité conditionnelle du processus de Hawkes satisfait l'**équation de renouvellement** suivante, où f est l'inconnue :

$$\forall t \geq 0, \quad f(t) = \lambda + \int_0^t h(t-s)f(s)ds. \quad (2.1.1)$$

Dans la littérature, l'équation (2.1.1) est un cas particulier d'**équation intégrale de Volterra de seconde espèce**.

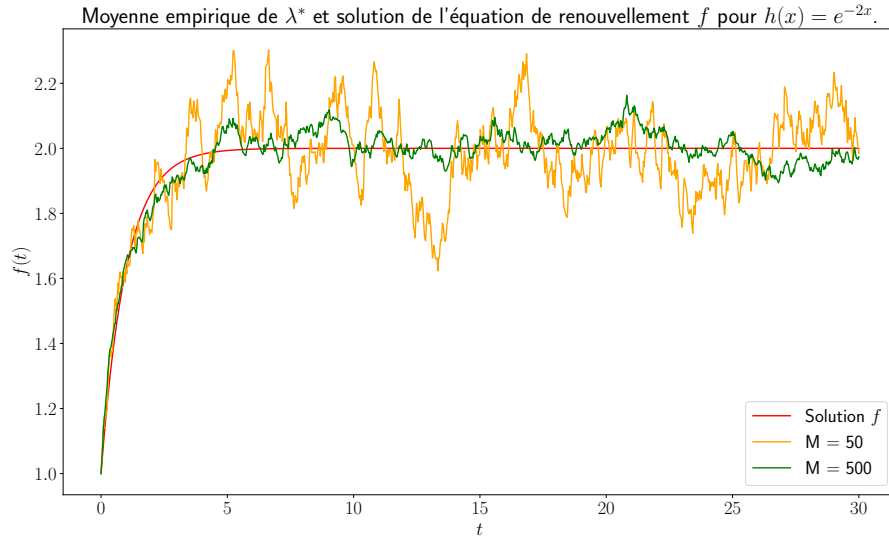


FIGURE 2.1 – Sur l'intervalle $[0, 30]$ et pour $\lambda = 1$: en orange et vert, la moyenne empirique de $M \in \{50, 500\}$ réalisations de l'intensité conditionnelle d'un processus de Hawkes linéaire où $h(x) = e^{-2x}$. En rouge, la solution de l'équation de renouvellement linéaire (2.1.1) associée.

2.1.1 Équation intégrale de Volterra de seconde espèce

Nous prenons les définitions du livre de Brunner [11], qui donne un panorama très large de résultats et d'applications des équations intégrales de Volterra.

Définition 2.1.1. Soient $T > 0$ et $I := [0, T]$, $D := \{(s, t) \in \mathbb{R}^2, 0 \leq s \leq t \leq T\}$. Soient $g : I \rightarrow \mathbb{R}$ et $K : D \rightarrow \mathbb{R}$ des fonctions données. L'équation :

$$\forall t \in I, \quad f(t) = g(t) + \int_0^t K(s, t) f(s) ds, \quad (2.1.2)$$

est appelée **équation intégrale de Volterra de seconde espèce linéaire** de noyau K , où f est l'inconnue. Si de plus, le noyau K est de la forme :

$$K(s, t) = h(t - s),$$

où h est une fonction positive, alors l'équation (2.1.2) est appelée **équation de renouvellement** (renewal equation).

Remarque 2.1.2. On peut montrer que dans le cas où les fonctions g et K sont continues, l'équation (2.1.2) possède une unique solution f . Pour cela, on considère l'opérateur intégral linéaire de Volterra, noté $\mathcal{V} : \mathcal{C}(I) \rightarrow \mathcal{C}(I)$, défini par :

$$(\mathcal{V}\phi)(t) := \int_0^t K(t, s)\phi(s)ds, \quad t \in I.$$

Une fonction $f \in \mathcal{C}(I)$ est alors solution de (2.1.2) si, et seulement si :

$$\forall t \in I, \quad f(t) = g(t) + (\mathcal{V}f)(t).$$

Revenons à l'équation (2.1.1). Dans le cas général (c'est-à-dire pour une fonction $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ quelconque), il n'est pas possible d'exprimer explicitement les solutions de cette équation. Cependant, leur approximation numérique est possible sous certaines hypothèses de régularité, comme détaillé, par exemple, dans le chapitre 11 du livre de Delves et Walsh [18], qui présente plusieurs méthodes différentes pouvant être utilisées. Nous détaillons ci-après l'approximation numérique de la solution f de l'équation (2.1.1) en utilisant la méthode des trapèzes. On suppose dans la suite que la fonction h est continue et dans $L^\infty(\mathbb{R}_+)$.

Approximation numérique de la solution de l'équation (2.1.1) On divise l'intervalle $[0, T]$ en N sous-intervalles de même taille $\varepsilon := \frac{T}{N}$. On obtient une suite de points

$$0 = t_0 < t_1 < \dots < t_N = T,$$

tels que pour tout $k \in \{0, \dots, N\}$, $t_k = k\varepsilon$. On note également f_k la valeur approchée de $f(t_k)$.

En tirant parti de (2.1.1), on peut définir le schéma suivant, dans lequel la valeur des f_k est définie de manière récursive. La solution f vérifie forcément $f(0) = \lambda$. On pose alors $f_0 := \lambda$ et en utilisant (2.1.1), pour $k \in \{1, \dots, N\}$, dès lors que f_0, \dots, f_{k-1} sont construits, on pose :

$$f_k := \lambda + \varepsilon \left(\frac{h(t_k)f_0}{2} + \sum_{i=1}^{k-1} h(t_k - t_i)f_i + \frac{h(t_0)f_k}{2} \right)$$

Ainsi, (f_1, \dots, f_N) est la solution d'un système de N équations linéaires qui se résout facilement, donnant pour tout $k \in \{1, \dots, N\}$:

$$f_k = \frac{2\lambda + \lambda\varepsilon h(k\varepsilon) + 2\varepsilon \sum_{i=1}^{k-1} h((k-i)\varepsilon)f_i}{2 - \varepsilon h(0)},$$

si l'on suppose avoir choisi ε assez petit (*i.e.* N assez grand) pour que $2 - \varepsilon h(0) \neq 0$. Le schéma d'approximation numérique ainsi défini est convergent, dans le sens où pour tout $k \in \{1, \dots, N\}$, $e_k := |f(t_k) - f_k| \xrightarrow{\varepsilon \rightarrow 0} 0$ (voir chapitre 11, partie 6.A de [18]).

2.1.2 Équation fonctionnelle non-linéaire

Dans le cas d'un processus de Hawkes tel qu'introduit dans la définition 2.0.1, l'espérance de l'intensité conditionnelle ne satisfait plus l'équation (2.1.1) en raison

de la non-linéarité induite par la fonction ReLU. Cependant, nous souhaitons tout de même comparer le comportement de la moyenne empirique de l'intensité d'un tel processus de Hawkes (notée $\mathbb{E}[\lambda^*]$) avec les solutions déterministes (notées f) de l'équation intégrale suivante, pour différentes fonctions h :

$$\forall t \geq 0, \quad f(t) = \left(\lambda + \int_0^t h(t-s)f(s)ds \right)_+ . \quad (2.1.3)$$

Le but de cette comparaison était de mesurer qualitativement l'écart de comportement entre $\mathbb{E}[\lambda^*]$ et les fonctions f satisfaisant (2.1.3).

Nous n'avons trouvé que très peu d'articles traitant de l'existence de solutions à ce genre d'équation. Dans un article datant de 1914, Galakjian [33] étudie l'existence de solutions à l'équation :

$$\forall t \geq 0, \quad u(t, t_0) = g \left(t, \int_{t_0}^t h_1(t, s, u(s, t_0))ds, \dots, \int_{t_0}^t h_m(t, s, u(s, t_0))ds \right),$$

où g, h_1, \dots, h_m sont des fonctions données. Plus récemment, Liu *et al.* [62] ont étudié l'équation fonctionnelle :

$$\forall t \geq 0, \quad u(t) = g \left(t, u(t), \int_0^t k(t, s, u(a(s)), u(b(s)))ds \right),$$

où g, u, a, b sont des fonctions données, donnant des conditions suffisantes d'existence et de stabilité de solutions à cette équation. Cependant, les hypothèses de cet article ne s'appliquant pas à notre cas, nous montrons dans un premier temps directement que l'équation (2.1.3) possède une unique solution.

Théorème 2.1.3. *On considère une fonction $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ une fonction continue de $L^\infty(\mathbb{R})$, et $\lambda > 0$. Alors l'équation (2.1.3) d'inconnue f admet une unique solution sur \mathbb{R}_+ .*

Démonstration. **Sur un intervalle compact** On commence par démontrer le résultat du théorème lorsque l'équation est posée pour $t \in [0, T]$ un intervalle compact, où $T > 0$. On note alors $E = \mathcal{C}(I)$ l'espace des fonctions réelles continues sur I , muni de la norme de la convergence uniforme, ce qui en fait un espace complet. On définit l'opérateur $\mathcal{W} : E \rightarrow E$ comme suit :

$$\forall t \in I, \quad \mathcal{W}(f)(t) = \left(\lambda + \int_0^t h(t-s)f(s)ds \right)_+ .$$

L'équation (2.1.3) s'écrit alors $f = \mathcal{W}f$. On va donc montrer que \mathcal{W} admet un unique point fixe dans E , en utilisant le théorème de point fixe de Picard.

Pour cela, montrons que \mathcal{W} admet une itérée contractante. Pour tout $f, g \in E$

et $t \in I$, on a en utilisant que la fonction $(\lambda + \cdot)_+$ est 1-lipschitzienne :

$$\begin{aligned} |\mathcal{W}(f)(t) - \mathcal{W}(g)(t)| &= \left| \left(\lambda + \int_0^t h(t-s)f(s)ds \right)_+ - \left(\lambda + \int_0^t h(t-s)g(s)ds \right)_+ \right| \\ &\leq \left| \int_0^t h(t-s)f(s) - \int_0^t h(t-s)g(s) \right| \\ &\leq \|h\|_\infty \int_0^t |f(s) - g(s)| ds \\ &\leq t \|h\|_\infty \|f - g\|_\infty \end{aligned}$$

On en déduit par récurrence que pour tout $k \in \mathbb{N}^*$,

$$|\mathcal{W}^k(f)(t) - \mathcal{W}^k(g)(t)| \leq \frac{t^k \|h\|_\infty^k}{k!} \|f - g\|_\infty.$$

En effet, pour $k \in \mathbb{N}^*$, on a d'après ce qui précède :

$$\begin{aligned} \left| \mathcal{W}^{k+1}(f)(t) - \mathcal{W}^{k+1}(g)(t) \right| &= \left| \mathcal{W}(\mathcal{W}^k(f))(t) - \mathcal{W}(\mathcal{W}^k(g))(t) \right| \\ &\leq \|h\|_\infty \int_0^t |\mathcal{W}^k(f)(s) - \mathcal{W}^k(g)(s)| ds \\ &\leq \frac{\|h\|_\infty^{k+1}}{k!} \|f - g\|_\infty \int_0^t s^k ds \\ &\leq \frac{t^{k+1} \|h\|_\infty^{k+1}}{(k+1)!} \|f - g\|_\infty. \end{aligned}$$

On a donc que pour tout $k \in \mathbb{N}^*$,

$$\|\mathcal{W}^k(f) - \mathcal{W}^k(g)\|_\infty \leq \frac{T^k \|h\|_\infty^k}{k!} \|f - g\|_\infty.$$

De plus, il existe $p \in \mathbb{N}^*$ suffisamment grand tel que $\frac{T^p \|h\|_\infty^p}{p!} < 1$, puisque c'est le terme général de la série exponentielle $e^{T\|h\|_\infty}$. Ainsi, l'application \mathcal{W}^p est contractante, et possède d'après le théorème du point fixe de Picard un unique point fixe $f \in E$. Ce point vérifie :

$$\mathcal{W}^p(\mathcal{W}(f)) = \mathcal{W}(\mathcal{W}^p(f)) = \mathcal{W}(f),$$

et donc $\mathcal{W}f$ est aussi un point fixe de \mathcal{W}^p .

Par unicité du point fixe de \mathcal{W}^p , on en déduit que $\mathcal{W}(f) = f$ et que f est un point fixe de \mathcal{W} . Inversement, tout point fixe de \mathcal{W} est également un point fixe de \mathcal{W}^p . Cela démontre le résultat pour tout intervalle compact.

Généralisation pour \mathbb{R}_+ L'intervalle \mathbb{R}_+ peut s'écrire $\mathbb{R}_+ = \cup_{j \in \mathbb{N}} I_j$ où $I_j = [0, j]$, réunion d'une suite croissante d'intervalles compacts $I_0 \subset I_1 \subset \dots \subset I_j \subset \dots$. On note f_j l'unique solution de l'équation (2.1.3) sur I_j .

Si f est une solution de l'équation (2.1.3) sur \mathbb{R} , alors la restriction de f à I_j coïncide nécessairement avec f_j , par unicité sur I_j .

Réciproquement, la fonction

$$f(t) = f_j(t), \quad \text{pour tout } j \in \mathbb{N} \text{ tel que } t \in I_j,$$

est bien définie par unicité et donne une solution à l'équation (2.1.3) sur \mathbb{R}_+ .

L'équation (2.1.3) admet donc une unique solution définie sur \mathbb{R}_+ . □

2.1.2.1 Approximation numérique de la solution de (2.1.3)

Nous proposons une méthode d'approximation numérique *via* la méthode des trapèzes comme dans le cas linéaire.

Théorème 2.1.4. *On se place sous les hypothèse du théorème 2.1.3. Soit $T > 0$ et $N \in \mathbb{N}^*$. On définit les points*

$$0 := t_0 < t_1 < \dots < t_N = T,$$

ainsi que $\varepsilon := \frac{T}{N}$, de telle manière à ce que $t_k = k\varepsilon$ pour $k \in \{0, \dots, N\}$.

On pose $f_0 := \lambda$ et, pour tout $k \in \{0, \dots, N\}$, on définit récursivement les nombres f_1, \dots, f_k de la façon suivante :

$$f_k = \left(\frac{2\lambda + \varepsilon h(t_k - t_0)f_0 + 2\varepsilon \sum_{j=1}^{k-1} h(t_k - t_j)f_j}{2 - \varepsilon h(0)} \right)_+. \quad (2.1.4)$$

Les nombres f_0, \dots, f_N ainsi définis forme un schéma d'approximation de la solution f de l'équation (2.1.3), convergent lorsque $N \rightarrow +\infty$ (i.e. $\varepsilon \rightarrow 0$).

Démonstration. Pour initialiser notre algorithme d'approximation, on note que la solution f de (2.1.3) satisfait encore $f(0) = \lambda$, on pose alors $f_0 := \lambda$.

En utilisant (2.1.3), on définit de façon récursive les nombres f_1, \dots, f_N de la façon suivante : si f_1, \dots, f_{k-1} sont construits, alors f_k satisfait l'équation

$$f_k = \left(\lambda + \varepsilon \left(\frac{h(t_k)f_0}{2} + \sum_{i=1}^{k-1} h(t_k - t_i)f_i + \frac{h(t_0)f_k}{2} \right) \right)_+, \quad (2.1.5)$$

ce qui définit implicitement f_k . Dans notre cas on peut résoudre l'équation linéaire puis appliquer la partie positive, ce qui donne :

$$f_k = \left(\frac{2\lambda + \varepsilon h(t_k - t_0)f_0 + 2\varepsilon \sum_{j=1}^{k-1} h(t_k - t_j)f_j}{2 - \varepsilon h(0)} \right)_+,$$

à condition que ε soit suffisamment petit pour que $2 - \varepsilon h(0) > 0$. En effet, si f_k est une solution de (2.1.5), alors :

$$f_k = 0 \iff \lambda + \varepsilon \left(\frac{h(t_k)f_0}{2} + \sum_{i=1}^{k-1} h(t_k - t_i)f_i + \frac{h(t_0)f_k}{2} \right) \leq 0.$$

et donc f_k satisfait (2.1.4). Réciproquement, on voit que f_k défini par (2.1.4) satisfait dans tous les cas l'équation (2.1.5).

Convergence du schéma d'approximation On souhaite estimer l'erreur d'approximation entre la vraie valeur $f(t_k)$ et la valeur approximée f_k , et montrer que le schéma est convergent, c'est-à-dire que pour tout $k \in \{0, \dots, N\}$, $e_k := f(t_k) - f_k \xrightarrow{\varepsilon \rightarrow 0} 0$.

Notons qu'on a pour tout k :

$$f(t_k) = \left(\lambda + \varepsilon \left(\frac{1}{2}h(t_k - t_0)f(t_0) + \sum_{j=1}^{k-1} h(t_k - t_j)f(t_j) + \frac{1}{2}h(0)f(t_k) \right) + E_\varepsilon(t_k) \right)^+, \quad (2.1.6)$$

où $E_\varepsilon(t_k)$ est le terme d'erreur dans l'approximation de l'intégrale $\int_0^{t_k} h(t_k - u)f(u)du$ par la méthode des trapèzes avec un pas de taille ε . Il est notoire que dans ce cas,

$$E_\varepsilon(t_k) = -\frac{\varepsilon^2 t_k}{12} \frac{\partial^2}{\partial s^2} (h(t_k - s)f(s)) \Big|_{s=\tau_k \in [0, t_k]} \quad (2.1.7)$$

En utilisant le fait que la fonction $(\lambda + \cdot)$ soit 1-lipschitzienne sur \mathbb{R} , on a pour tout $k \in \{1, \dots, N\}$,

$$\begin{aligned} |e_k| &= |f(t_k) - f_k| \\ &\leq \left| \varepsilon \left(\sum_{j=1}^{k-1} h(t_k - t_j)(f(t_k) - f_k) + \frac{1}{2}h(0)(f(t_k) - f_k) \right) + E_\varepsilon(t_k) \right| \\ &= \left| \varepsilon \left(\sum_{j=1}^{k-1} h(t_k - t_j) e_j + \frac{1}{2}h(0)e_k \right) + E_\varepsilon(t_k) \right| \\ &\leq \varepsilon \sum_{j=1}^{k-1} |h(t_k - t_j)| |e_j| + \frac{\varepsilon}{2} |h(0)| |e_k| + |E_\varepsilon(t_k)| \\ &\leq \varepsilon \|h\|_\infty \sum_{j=1}^{k-1} |e_j| + \varepsilon \|h\|_\infty |e_k| + |E_\varepsilon(t_k)| \end{aligned}$$

Cela nous mène à :

$$|e_k| \leq \frac{\varepsilon \|h\|_\infty}{1 - \varepsilon \|h\|_\infty} \sum_{j=1}^{k-1} |e_j| + \frac{1}{1 - \varepsilon \|h\|_\infty} |E_\varepsilon(t_k)|$$

et on en déduit d'après la version discrète du lemme de Grönwall que :

$$|e_k| \leq \frac{|E_\varepsilon(t_k)|}{1 - \varepsilon \|h\|_\infty} \exp\left(\frac{\|h\|_\infty}{1 - \|h\|_\infty} t_k\right).$$

Ainsi, pour tout $k \in \{0, \dots, N\}$, on a $|e_k| \xrightarrow{\varepsilon \rightarrow 0} 0$, ce qui conclut la preuve du théorème. \square

2.2 Une classe particulière de fonctions h

Les résultats de cette partie s'appuient sur un travail en cours de Gerold Alsmeyer et Pascal Maillard. Afin d'établir notre comparaison qualitative de comportement entre les solutions de l'équation linéaire (2.1.1) et de l'équation non-linéaire (2.1.3), nous nous sommes concentrés sur un choix particulier de fonction $h : \mathbb{R}_+ \rightarrow \mathbb{R}$, à savoir :

$$\forall x \in \mathbb{R}, \quad h(x) = e^{-bx} (\gamma C_a(x) + \delta S_a(x)), \quad (2.2.1)$$

où

$$C_a(x) := \begin{cases} \cos(\sqrt{ax}) & \text{si } a \geq 0, \\ \cosh(\sqrt{-ax}) & \text{si } a < 0, \end{cases}, \quad S_a(x) := \begin{cases} \frac{\sin(\sqrt{ax})}{\sqrt{a}} & \text{si } a > 0 \\ \frac{\sinh(\sqrt{-ax})}{\sqrt{-a}} & \text{si } a < 0 \\ 1 & \text{si } a = 0 \end{cases},$$

et $a, b, \gamma, \delta \in \mathbb{R}$.

2.2.1 Comparaison des solutions de l'équation de renouvellement linéaire et non-linéaire

Dans le cas de l'équation de renouvellement linéaire (2.1.1), pour un tel choix de fonctions h , les solutions f de (2.1.1) sont explicites. Dans ce cas, la résolution de cette équation peut s'effectuer à l'aide de méthodes utilisant la transformée de Laplace de h . En effet, ces fonctions ont la particularité d'avoir une transformée de Laplace simple à calculer, à savoir :

$$\forall \mu \in \mathbb{R}, \quad \hat{h}(\mu) = \frac{(b + \mu)\gamma + \delta}{a + (b + \mu)^2}. \quad (2.2.2)$$

En utilisant (2.2.2), et la structure de renouvellement de l'équation (2.1.1), on en déduit la proposition suivante :

Proposition 2.2.1. *Soit $a, b, \gamma, \delta \in \mathbb{R}$. On considère la fonction h définie par (2.2.1). On définit alors a', b', γ', δ' les quantités suivantes :*

$$\begin{aligned} a' &= a - \delta - \gamma^2/4 \\ b' &= b - \gamma/2 \\ \gamma' &= \gamma \\ \delta' &= \delta + \gamma^2/2 \end{aligned}$$

La solution f de l'équation de renouvellement (2.1.1) associée à h est alors donnée par :

$$f(t) = \lambda + \lambda \times \begin{cases} \frac{e^{-b't}}{(b')^2 + a'} \left[(-b'\gamma' - \delta') \cos(\sqrt{a'}t) + \left(\gamma'\sqrt{a'} - \frac{b'\delta'}{\sqrt{a'}} \right) \sin(\sqrt{a'}t) \right] + \frac{\gamma'b' + \delta'}{(b')^2 + a'} & \text{si } a' > 0 \\ e^{-b't} \left(\delta' - \frac{\gamma'}{b'} - \frac{\delta'}{(b')^2} \right) + \frac{\gamma'b' + \delta'}{(b')^2} & \text{si } a' = 0 \\ \frac{\gamma' + \delta'}{2(-b' + \sqrt{|a'|})} \left(e^{(-b' + \sqrt{|a'|})t} - 1 \right) + \frac{\gamma' - \delta'}{2(-b' - \sqrt{|a'|})} \left(e^{(-b' - \sqrt{|a'|})t} - 1 \right) & \text{si } a' < 0 \end{cases}$$

De ce fait, nous avons pu comparer les solutions des équations de renouvellement linéaire (2.1.1) et non-linéaire (2.1.3) avec ces mêmes fonctions h , en utilisant directement dans le cas linéaire les solutions explicites données par la proposition 2.2.1. De plus, les fonctions h définies par (2.2.1) sont signées, nous donnant donc un cadre pertinent pour interpréter nos observations quant à l'influence de la partie positive dans l'équation (2.1.5).

Nous nous sommes particulièrement penchés sur un cas qui a attiré notre attention. Pour des paramètres a, b, γ, δ bien choisis, la solution f de l'équation linéaire (2.1.1) a un comportement oscillatoire et exponentiellement divergent. Dans ce cas particulier, le rôle de la partie positive rajoutée dans l'équation (2.1.3) peut être bien visible : un exemple est donné dans la figure 2.2 avec $h(x) := \cos(x)$.

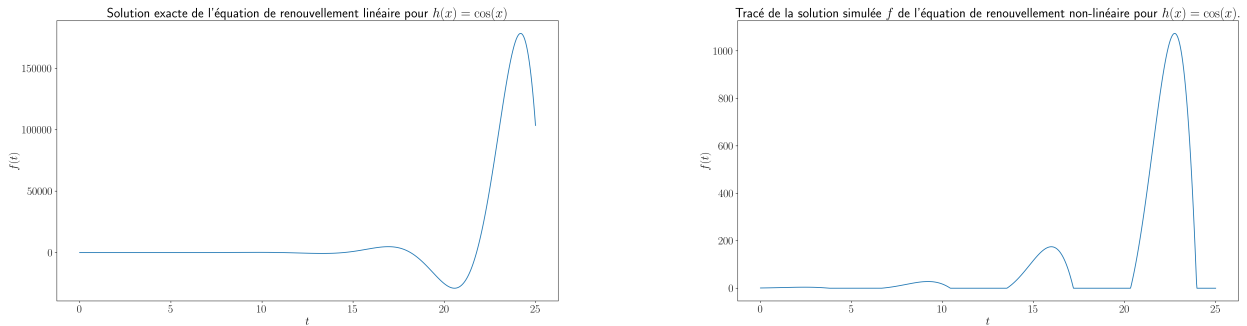


FIGURE 2.2 – Comparaison entre la solution exacte de l'équation de renouvellement linéaire (2.1.1) et la solution approchée numériquement de l'équation non-linéaire (2.1.3) sur $[0, 25]$ pour $h(x) = \cos(x)$ et $\lambda = 1$. Le comportement oscillatoire est conservé par la fonction ReLU mais les "pics" de la solution f sont fortement réduits.

Pour la solution approchée numériquement de l'équation non linéaire (2.1.3), on remarque le même comportement oscillatoire et croissant, mais avec des amplitudes réduites par rapport à celles observées dans le cas linéaire. Nous avons alors conjecturé que dans ce cas, la solution f de l'équation non-linéaire pouvait s'écrire $f = g^+$, où g est la solution d'une certaine équation de renouvellement linéaire. La solution f serait alors seulement une restriction de la fonction g à ses valeurs positives ou nulles.

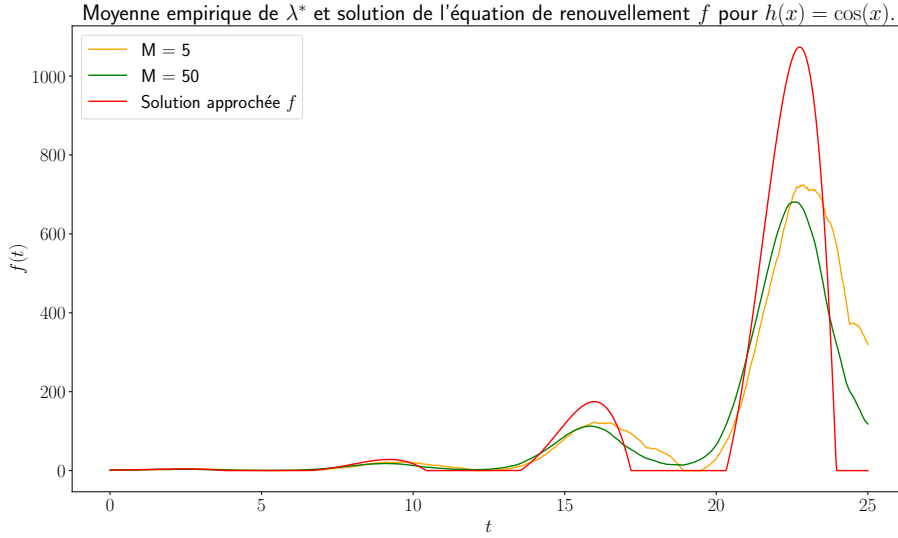


FIGURE 2.3 – Sur l'intervalle $[0, 25]$ et pour $\lambda = 1$: en orange et vert, la moyenne empirique de $M \in \{5, 50\}$ réalisations de l'intensité conditionnelle d'un processus de Hawkes non-linéaire où $h(x) = \cos(x)$. En rouge, la solution f de l'équation de renouvellement non-linéaire (2.1.3) associée. On mesure qualitativement un écart entre les valeurs de la moyenne empirique et de la solution de (2.1.3), mais leurs comportements au cours du temps est très proche.

2.2.2 Détermination de fonctions h ne satisfaisant pas le critère suffisant de stabilité

Dans un premier temps au cours de cette thèse, nous avons cherché à étendre le critère de stabilité donné par Costa *et al* dans [13] :

$$\|h^+\|_1 < 1, \quad (2.2.3)$$

ce critère ne prenant pas en compte les valeurs négatives de la fonction h , qui peuvent jouer un rôle crucial dans la modélisation de l'inhibition du système. N'ayant pas de pistes spécifiques à suivre, nous avons d'abord effectué une exploration numérique approfondie afin de dégager des intuitions et des conjectures sur les caractéristiques de la fonction h qui assureraient la stabilité du processus, c'est-à-dire pour lequel son intensité conditionnelle reste bornée presque sûrement au cours du temps.

En réutilisant une classe particulière de fonctions h précédemment introduite, et en tirant parti du fait que ces fonctions soient signées et possèdent une transformée de Laplace simple, nous avons pu identifier plusieurs exemples de fonctions $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ signées, telles que $\|h^+\|_1 > 1$, et dont le processus de Hawkes associé demeure stable. Par exemple, la fonction

$$h(x) := e^{-\frac{x}{4}} (\sin(x) - 10 \cos(x))$$

vérifie :

$$\|h^+\|_1 \approx 12, \quad \|h\|_1 \approx 25.5,$$

et le processus de Hawkes avec inhibition, défini selon la définition 2.0.1 et associé à cette fonction h , est stable comme l'illustre la figure 2.4.

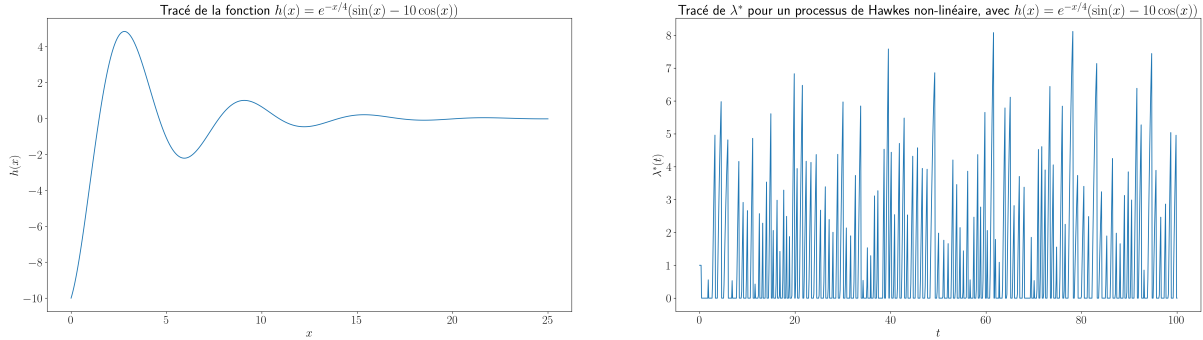


FIGURE 2.4 – Tracés de la fonction $h(x) := e^{-\frac{x}{4}} (\sin(x) - 10 \cos(x))$, et d'une réalisation de l'intensité conditionnelle λ^* du processus de Hawkes avec inhibition associé. Par rapport au processus de Hawkes associé, chaque nouveau point inhibe donc fortement le processus dans un premier temps, puis sa participation à l'intensité conditionnelle du processus alterne entre un effet positif et négatif tout en décroissant exponentiellement vite vers 0.

Plus généralement, en utilisant cette classe de fonctions h , nous pouvons déterminer des noyaux tels que $\|h^+\|_1$ soit arbitrairement grand et dont le processus de Hawkes associé soit stable. Cela nous conforte dans l'idée que le critère (2.2.3) est très restrictif dès lors que la fonction h peut prendre des valeurs négatives.

(Almost) complete characterization of stability of a discrete-time Hawkes process with inhibition and memory of length two

This self-contained chapter consists of the article [14], which is to appear in the *Applied Probability Journal*.

Abstract

We consider a Poisson autoregressive process whose parameters depend on the past of the trajectory. We allow these parameters to take negative values, modelling inhibition. More precisely, the model is the stochastic process $(X_n)_{n \geq 0}$ with parameters $a_1, \dots, a_p \in \mathbb{R}$, $p \in \mathbb{N}$ and $\lambda \geq 0$, such that for all $n \geq p$, conditioned on X_0, \dots, X_{n-1} , X_n is Poisson distributed with parameter

$$(a_1 X_{n-1} + \dots + a_p X_{n-p} + \lambda)_+$$

This process can be regarded as a discrete time Hawkes process with inhibition and a memory of length p . In this paper we initiate the study of necessary and sufficient conditions of stability for these processes, which seems to be a hard problem in general. We consider specifically the case $p = 2$, for which we are able to classify the asymptotic behavior of the process for the whole range of parameters, except for boundary cases. In particular, we show that the process remains stochastically bounded whenever the linear recurrence equation $x_n = a_1 x_{n-1} + a_2 x_{n-2} + \lambda$ remains bounded, but the converse is not true. Furthermore, the criterion for stochastic boundedness is not symmetric in a_1 and a_2 , in contrast to the case of non-negative parameters, illustrating the complex effects of inhibition.

Contents

3.1 Introduction	46
3.2 Notations, definitions and results	48
3.2.1 Definition and main result	48

3.2.2	The associated Markov chain	49
3.3	Proof of Theorem 3.2.2: recurrence	50
3.3.1	Foster-Lyapounov drift criteria	51
3.3.2	Case \mathcal{R}_1	52
3.3.3	Case \mathcal{R}_2	53
3.3.4	Case \mathcal{R}_3	55
3.4	Proof of Theorem 3.2.2: transience	57
3.4.1	Case T1	58
3.4.2	Case T2: $0 \leq a < 2$ and $a + b > 1$ or $a \geq 2$ and $a^2 + 4b > 0$.	59
3.4.2.1	Subcase T2a : $b \geq 0$	61
3.4.2.2	Subcase T2b : $b < 0$	61
3.5	Perspectives and open problems	64
3.A	Linear recurrence equations	65
3.B	Criteria for strong irreducibility	66

3.1 Introduction.

The motivation of this paper is to pave the way for obtaining sufficient and necessary conditions for the stability of non-linear Hawkes processes with inhibition. Hawkes processes are a class of point processes used to model events that have mutual influence over time. They were initially introduced by Hawkes (1971) ([41], [43]) and are now used in a variety of fields such as finance, biology, and neuroscience.

More precisely, a Hawkes process $(N_t^h)_{t \in \mathbb{R}} = (N^h([0, t]))_{t \in \mathbb{R}}$ is defined by its initial condition on $(-\infty, 0]$ and its stochastic conditional intensity denoted by Λ , characterized by :

$$\Lambda(t) = \phi \left(\lambda + \int_{-\infty}^t h(t-s) N^h(ds) \right), \quad (3.1.1)$$

where $\lambda > 0$, $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ are measurable, deterministic functions (see [16] for further details). The function h is called the *reproduction function*, and contains the information of the behaviour of the process throughout time. In the case where ϕ is non-decreasing, the sign of the function h encodes for the type of time dependence: when h is non-negative, the process is said to be *self-exciting*; when h is signed : the negative values of h can then be seen as self-inhibition (see [12] and [13]). The case where $h \geq 0$ and $\phi = id$ is called the linear case. Considering signed functions h requires to add non-linearity by the mean of a function ϕ which ensures that the intensity remains positive. In this paper, we will focus on the particular case where $\phi = (\cdot)_+$ is the ReLU function defined on \mathbb{R} by $(x)_+ = \max(0, x)$.

Several authors have established sufficient conditions on h ensuring the existence of a stable version of this process. For signed h , Brémaud and Massoulié [8] proved that a stable version of the process exists if $\|h\|_1 < 1$. In [13], Costa *et al.* proved that it is sufficient to have $\|h^+\|_1 < 1$, where $h^+(x) = \max(h(x), 0)$ using a coupling

argument. Unfortunately, this sufficient criterion does not take into account the effect of inhibition, captured by the negative part of h . Going further is difficult because non-linearity breaks the direct link between the function h and the probabilistic structure of the Hawkes process. Recent results have been obtained by [73] for a two dimensional non-linear Hawkes process with weighted exponential kernel, modeling the case of two populations of interacting neurons including both inhibition and excitation. The author provide a criterion on the weight function matrix for stability exploiting the markovian structure of the Hawkes process in that case.

In order to get an intuition on the results that one might obtain on Hawkes processes, we choose to consider a simplified, discrete analogue of those processes. Namely we will study an autoregressive process $(\tilde{X}_n)_{n \geq 1}$ with initial condition $(\tilde{X}_0, \dots, \tilde{X}_{-p+1})$ where $p \in \{1, 2, \dots\}$, and such that:

$$\forall n \geq 1, \quad \tilde{X}_n \sim \mathcal{P} \left(\phi \left(a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda \right) \right), \quad (3.1.2)$$

where $\mathcal{P}(\rho)$ denotes the Poisson distribution with parameter ρ , and a_1, \dots, a_p are real numbers.

In the linear case (a_1, \dots, a_p non-negative, and $\phi(x) = x$) these integer-valued processes are called INGARCH processes, and have already been studied in ([26], [28]), where a necessary condition for existence and stability of this class of processes has been derived and can be written $\sum_{i=1}^p a_i < 1$. Furthermore, the link between Hawkes processes and autoregressive Poisson processes has already been made for the linear case : M. Kirchner proved that the discretized autoregressive process (with $p = +\infty$) converges weakly to the associated Hawkes process (see [54] for details). An analogous discrete process has been proposed by Seol [77] using an autoregressive structure based on Bernoulli random variables.

In order to explore the effect of inhibition, we consider signed values for the parameters a_1, \dots, a_p . In this article, we focus on the specific case of $p = 2$, so that our model of interest can be written as :

$$\forall n \geq 2, \quad \tilde{X}_n \sim \mathcal{P} \left(\left(a \tilde{X}_{n-1} + b \tilde{X}_{n-2} + \lambda \right)_+ \right), \quad (3.1.3)$$

with $a, b \in \mathbb{R}$ and $\tilde{X}_0, \tilde{X}_1 \in \mathbb{N}$. (In this paper we will use the convention $0 \in \mathbb{N}$).

In this paper, the most important result is the classification of the process defined in (3.1.3). Note that the complete characterization of the behavior of this simple process is difficult due to the variety of behaviors observed. We will prove that the introduction of non-linearity through the ReLU function makes the process more stable relatively to its linear counterpart, in the sense that the parameter space $(a, b) \in \mathbb{R}^2$ for which the linear process $y_{n+1} = ay_n + by_{n-1} + \lambda$ admits a stationary version is strictly a subset of the parameter space for which the non-linear process admits a stationary version (see Appendix A.). Our results also enlighten the complex role of inhibition and in particular the asymmetric role of a and b associated with the range at which inhibition occurs. Our work suggests the existence of complex algebraic and geometric structures that are likely to play an

important role in the more general case of a memory of order p . In order to obtain our results, we used a wide range of probabilistic tools, corresponding to the variety of the behaviours of the trajectories of the process, depending on the parameters of the model.

3.2 Notations, definitions and results

3.2.1 Definition and main result

Let $a, b \in \mathbb{R}$ and $\lambda > 0$. We consider a discrete time process $(\tilde{X}_n)_{n \geq 1}$ with initial condition $(\tilde{X}_0, \tilde{X}_{-1})$ such that the following holds for all $n \geq 1$:

$$\text{conditioned on } \tilde{X}_{-1}, \dots, \tilde{X}_{n-1}: \tilde{X}_n \sim \mathcal{P} \left(\left(a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + \lambda \right)_+ \right),$$

where $(\cdot)_+$ is the ReLU function defined on \mathbb{R} by $(x)_+ := \max(0, x)$.

As we said previously, some papers have already been dealing with the linear version of this process : if a and b are non-negative, the parameter of the Poisson random variable in (3.1.3) is also non-negative, and the ReLU function vanishes. In this case, Proposition 1 in [26] states that the process is a second-order stationary process if $a + b < 1$. This weak stationarity ensures that the mean, variance and autocovariance are constant with time.

Let us define the function

$$b_c(a) = \begin{cases} 1 & a \leq 0 \\ 1 - a & a \in (0, 2) \\ -\frac{a^2}{4} & a \geq 2 \end{cases}$$

and define the following sets (see Figure 3.1):

$$\mathcal{R} = \{(a, b) \in \mathbb{R}^2 : b < b_c(a)\} \tag{3.2.1}$$

$$\mathcal{T} = \{(a, b) \in \mathbb{R}^2 : b > b_c(a)\}. \tag{3.2.2}$$

Our main result is the following

Theorem 3.2.1.

- If $(a, b) \in \mathcal{R}$, then the sequence $(\tilde{X}_n)_{n \geq 0}$ converges in law as $n \rightarrow \infty$.
- if $(a, b) \in \mathcal{T}$, then the sequence $(\tilde{X}_n)_{n \geq 0}$ satisfies that almost surely

$$\tilde{X}_n + \tilde{X}_{n+1} \xrightarrow[n \rightarrow \infty]{} +\infty.$$

This result derives from the study of the natural Markov chain associated with \tilde{X}_n that is defined by:

$$X_n := (\tilde{X}_n, \tilde{X}_{n-1}), \quad n \geq 0. \tag{3.2.3}$$

Before giving more details about the behaviour of $(X_n)_{n \geq 0}$, let us comment on Theorem 3.2.1. In particular, we stress that the condition for convergence in law is not symmetrical in a and b . More precisely, for any $a \in \mathbb{R}$, the sequence (\tilde{X}_n) can be tight provided that b is chosen small enough, but the converse is not true as soon as $b > 1$. This induces that inhibition has a stronger regulating effect when occurs after an excitation, rather than before.

The question of the critical behavior of the process on the boundary $\{b = b_c(a)\}$ remains open and represent to a difficult question for further work.

3.2.2 The associated Markov chain

As mentioned above, the main part of the article is devoted to the study of a Markov chain (X_n) which encodes the time dependency of (\tilde{X}_n) . We will rely on the recent treatment by Douc, Moulines, Priouret and Soulier [19] for results about Markov chains. In particular, we use their notion of irreducibility, which is weaker than the usual notion of irreducibility typically found in textbooks on Markov chains (on discrete state space). Thus, a Markov chain is called **irreducible** if there exists an **accessible state**, i.e., a state that can be reached with positive probability from any other state. Following [19], we refer to the usual notion of irreducibility (i.e., every state is accessible) as **strong irreducibility**.

The transition matrix of the Markov chain $(X_n)_{n \geq 0}$ defined in (3.2.3) is thus given for $(i, j, k, l) \in \mathbb{N}^4$ by:

$$P((i, j), (k, l)) = \delta_{il} \frac{e^{-s_{ij}} s_{ij}^k}{k!}, \quad (3.2.4)$$

where :

$$s_{ij} := (ai + bj + \lambda)_+ \text{ and } \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}.$$

In other words, starting from a state (i, j) , the next step of the Markov chain will be (k, i) where $k \in \mathbb{N}$ is the realization of a Poisson random variable with parameter s_{ij} . In particular, if $s_{ij} = 0$, then the next step of the Markov chain is $(0, i)$.

Since the probability that a Poisson random variable is zero is strictly positive, it is possible to reach the state $(0, 0)$ with positive probability from any state in two steps. In particular, the state $(0, 0)$ is accessible and the Markov chain is irreducible. Furthermore, the Markov chain is aperiodic [19, Section 7.4], since $P((0, 0), (0, 0)) = e^{-\lambda} > 0$. Note that strong irreducibility may not hold (see Proposition 3.B.1 in the appendix).

Recall the definition of the sets \mathcal{R} and \mathcal{T} in (3.2.1) and (3.2.2).

Theorem 3.2.2.

- Let $(a, b) \in \mathcal{R}$. Then the Markov chain $(X_n)_{n \geq 0}$ is **geometrically ergodic**, i.e., it admits an invariant probability measure π and there exists $\beta > 1$, such that for every initial state,

$$\beta^n d_{TV}(\text{Law}(X_n), \pi) \rightarrow 0, \quad n \rightarrow \infty,$$

where d_{TV} denotes total variation distance.

- Let $(a, b) \in \mathcal{T}$. Then the Markov chain is transient, i.e., every state is visited a finite number of times almost surely, for every initial state.

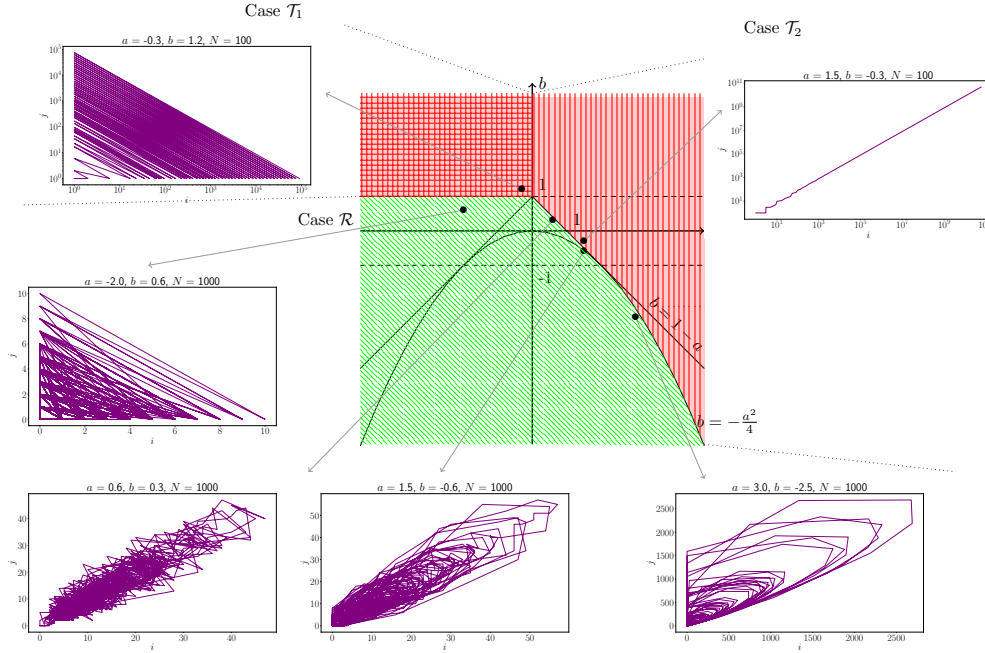


Figure 3.1 – The partition of the parameter space described in 3.2.2. The green region corresponds to \mathcal{R} , while the red region corresponds to \mathcal{T} . The smaller figures are typical trajectories of the Markov chain $(X_n)_{n \geq 0}$ for each region of the parameter space. In the all the simulations, we chose $\lambda = 1$. The region delineated by a dashed triangular line corresponds to the region of parameter space for which the linear recurrence equation $y_n = ay_{n-1} + by_{n-2} + \lambda$ is bounded for all $n \in \mathbb{N}$, for any given $y_0, y_1 \in \mathbb{R}$. See Appendix A for more details.

Theorem 3.2.1 is a simple consequence of this result. Indeed, in the case of $(a, b) \in \mathcal{R}$, the convergence in law of \tilde{X}_n simply derives from the convergence in law of X_n since \tilde{X}_n is the first coordinate of X_n . In the transient case, $(a, b) \in \mathcal{T}$ the result in Theorem 3.2.1 simply derives from the fact that $\|X_n\|_1 \rightarrow_{n \rightarrow \infty} \infty$ almost surely.

The rest of the article is devoted to the proof of Theorem 3.2.2. We first focus on the recurrent case in Section 3.3, then on the transient case in Section 3.4.

3.3 Proof of Theorem 3.2.2: recurrence

In this section, we prove the recurrence part of Theorem 3.2.2. The proof goes by exhibiting three functions satisfying Foster-Lyapounov drift conditions for different ranges of the parameters (a, b) covering the whole recurrent regime \mathcal{R} .

3.3.1 Foster-Lyapounov drift criteria

Drift criteria are powerful tools that have been introduced by Foster [31], and deeply studied, and popularized, by Meyn and Tweedie [67], among others. These drift criteria allow to prove convergence to the invariant measure of Markov chains and yield explicit rates of convergence. We use here the treatment from Douc, Moulines, Priouret and Soulier [19], which is influenced by Meyn and Tweedie [67], but is more suitable for Markov chains which are irreducible but not strongly irreducible.

A set of states $C \subset \mathbb{N}^2$ is called **petite** [19, Definition 9.4.1], if there exists a state $x_0 \in \mathbb{N}^2$ and a probability distribution $(p_n)_{n \in \mathbb{N}}$ on \mathbb{N} such that

$$\inf_{x \in C} \sum_{n \in \mathbb{N}} p_n P^n(x, x_0) > 0,$$

where we recall that $P^n(x, x_0)$ is the n -step transition probability from x to x_0 . Since the Markov chain $(X_n)_{n \geq 0}$ is irreducible, any finite set is petite (take x_0 to be the accessible state) and any finite union of petite sets is petite [19, Proposition 9.4.5].

Let $V : \mathbb{N}^2 \rightarrow [1, \infty)$ be a function, $\varepsilon \in (0, 1]$, $K < \infty$ and $C \subset \mathbb{N}^2$ a set of states. We say that the **drift condition** $D(V, \varepsilon, K, C)$ is satisfied if

$$\Delta V(x) := \mathbb{E}_x[V(X_1) - V(X_0)] \leq -\varepsilon V(x) + K \mathbf{1}_C$$

It is easy to see that this condition implies condition $D_g(V, \lambda, b, C)$ from [19, Definition 14.1.5], with $\lambda = 1 - \varepsilon$ and $b = K$.

Proposition 3.3.1. *Assume that the drift condition $D(V, \varepsilon, K, C)$ is verified for some V , ε , K and C as above and assume that C is petite. Then there exists $\beta > 1$ and a probability measure π on \mathbb{N}^2 , such that, for every initial state $x \in \mathbb{N}^2$,*

$$\beta^n \times \sum_{y \in \mathbb{N}^2} V(y) |\mathbb{P}_x(X_n = y) - \pi(y)| \rightarrow 0, \quad n \rightarrow \infty.$$

In particular, for every initial state $x \in \mathbb{N}^2$,

$$\beta^n d_{TV}(\text{Law}(X_n), \pi) \rightarrow 0, \quad n \rightarrow \infty$$

and π is an invariant probability measure for the Markov chain $(X_n)_{n \geq 0}$.

Proof. As mentioned in Section 3.2.2, the Markov chain is irreducible and aperiodic. The first statement then follows by combining parts (ii) and (a) of Theorem 15.1.3 in [19] with the remark preceding Corollary 14.1.6 in [19]. The second statement immediately follows, noting that $V \geq 1$. \square

We will consider separately the following ranges of the parameters:

$$\begin{aligned} \mathcal{R}_1 &= \{(a, b) \in \mathbb{R}^2 : a, b < 1 \text{ and } a + b < 1\}, \\ \mathcal{R}_2 &= \{a > 0 \text{ and } a^2 + 4b < 0\}, \\ \mathcal{R}_3 &= \{1 \leq a < 2 \text{ and } -1 < b < 1 - a\}. \end{aligned}$$

We then have $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$, see Figure 3.2.

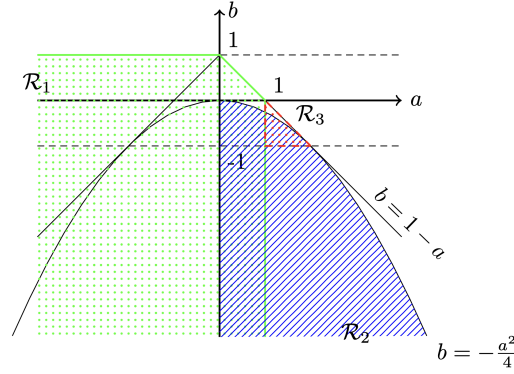


Figure 3.2 – Illustration of the three zones of parameters on which the proof of ergodicity will be carried.

3.3.2 Case \mathcal{R}_1

This case is the natural extension of the results that have been already proved for the linear process (see Proposition 1 in [26]).

Let $V : \mathbb{N}^2 \rightarrow \mathbb{R}_+$ be the function defined by

$$V(i, j) := \alpha i + \beta j + 1,$$

where $\alpha, \beta > 0$ are parameters to be chosen later.

We then have that $V(i, j) \geq 1$ for all $(i, j) \in \mathbb{N}^2$. We look for $\varepsilon > 0$ such that $\Delta V(x) + \varepsilon V(x) \leq 0$ except for a finite number of $x \in \mathbb{N}^2$.

Let $\varepsilon > 0$ to be properly chosen later. Then,

$$\begin{aligned} \Delta V(i, j) + \varepsilon V(i, j) &= \sum_{k \in \mathbb{N}} \frac{e^{-s_{ij}} s_{ij}^k}{k!} (\alpha k + \beta i + 1) - (\alpha i + \beta j + 1) + \varepsilon (\alpha i + \beta j + 1) \\ &= \alpha s_{ij} + i(\beta - \alpha + \alpha \varepsilon) + j(\beta \varepsilon - \beta) + \varepsilon, \end{aligned}$$

Note that $s_{ij} = 0$ or $s_{ij} = ai + bj + \lambda > 0$. In both cases, $\Delta V + \varepsilon V$ is a linear function of (i, j) . We will thus choose α, β such that the coefficients of $\Delta V + \varepsilon V$ are negative, so there will be only a finite number of (i, j) that satisfies $\Delta V(i, j) + \varepsilon V(i, j) \geq 0$.

Let us first consider couples (i, j) such that $s_{ij} = 0$. According to the above, it is sufficient to have :

$$\begin{cases} \beta - \alpha + \alpha \varepsilon < 0 \\ \beta \varepsilon - \beta < 0 \end{cases} \iff \begin{cases} \beta < \alpha(1 - \varepsilon) \\ \varepsilon < 1 \end{cases}.$$

In the sequel, we impose $\varepsilon < 1$.

If $s_{ij} = ai + bj + \lambda > 0$, then :

$$\Delta V(i, j) + \varepsilon V(i, j) = i(\alpha a - \alpha + \beta + \alpha \varepsilon) + j(\alpha b + \beta \varepsilon - \beta) + \lambda \alpha + \varepsilon.$$

For the same reasons as before, it is sufficient to have $\alpha, \beta > 0$ such that :

$$\begin{cases} \alpha a - \alpha + \beta + \alpha \varepsilon < 0 \\ \alpha b + \beta \varepsilon - \beta < 0 \end{cases} \iff \begin{cases} \beta < \alpha(1 - a - \varepsilon) \\ \beta > \frac{\alpha b}{1 - \varepsilon} \end{cases} \quad (\text{since } \varepsilon < 1) .$$

Let $\alpha := 1$. With the above statements we thus want to choose $\beta, \varepsilon > 0$ such that :

$$\begin{cases} \frac{b}{1 - \varepsilon} < \beta < 1 - a - \varepsilon \\ \beta < 1 - \varepsilon \end{cases} \quad \text{i.e.} \quad \frac{b}{1 - \varepsilon} < \beta < \min\{1 - a - \varepsilon, 1 - \varepsilon\}.$$

Recall that $a + b < 1$, so it is possible to find $\varepsilon_0 \in (0, 1)$ small enough so that :

$$\forall \tilde{\varepsilon} \leq \varepsilon_0, \quad \frac{b}{1 - \tilde{\varepsilon}} < 1 - a - \tilde{\varepsilon}.$$

- If $a \geq 0$, then $\min\{1 - a - \varepsilon, 1 - \varepsilon\} = 1 - a - \varepsilon$ and since $a < 1$, we can choose $\varepsilon \leq \varepsilon_0$ so small that $\frac{b}{1 - \varepsilon} < \beta < \min\{1 - a - \varepsilon, 1 - \varepsilon\}$ on one hand, and $1 - a - \varepsilon > 0$ on the other hand. It is thus possible to choose $\beta > 0$ such that :

$$\frac{b}{1 - \varepsilon} < \beta < \min\{1 - a - \varepsilon, 1 - \varepsilon\}.$$

- If $a < 0$, then $\min\{1 - a - \varepsilon, 1 - \varepsilon\} = 1 - \varepsilon$. Since $b < 1$, it is possible to set $\varepsilon \leq \varepsilon_0$ so small that $b < (1 - \varepsilon)^2$. Hence we have $\frac{b}{1 - \varepsilon} < 1 - \varepsilon$, so that it is possible to choose $\beta > 0$ that satisfies our constraints.

Note that $\Delta V(0, 0) = \lambda > 0$. Hence, with $\alpha, \beta, \varepsilon > 0$ chosen as above, we have that :

$$\Delta V(i, j) \leq -\varepsilon V(i, j) \text{ except for a finite number of states } (i, j) \in \mathbb{N}^2.$$

This proves that a drift condition $D(V, \varepsilon, C)$ holds for finite set C , which yields the result.

3.3.3 Case \mathcal{R}_2

In this section, we assume that $a > 0$ and $a^2 + 4b < 0$.

The Lyapounov function we will consider is the following one:

$$\forall (i, j) \in \mathbb{N}^2, \quad V(i, j) = \frac{i}{j+1} + 1.$$

Before getting into the details, a remark about this function. While we initially discovered it by trial and error, it has an interesting geometric interpretation. As seen in Figure 3.3, in case \mathcal{R}_2 , the macroscopic trajectories of the Markov chain tend to turn counterclockwise until they hit the j -axis and eventually get pulled back to $(0, 0)$. This provides a heuristic understanding of why V should be a Lyapounov

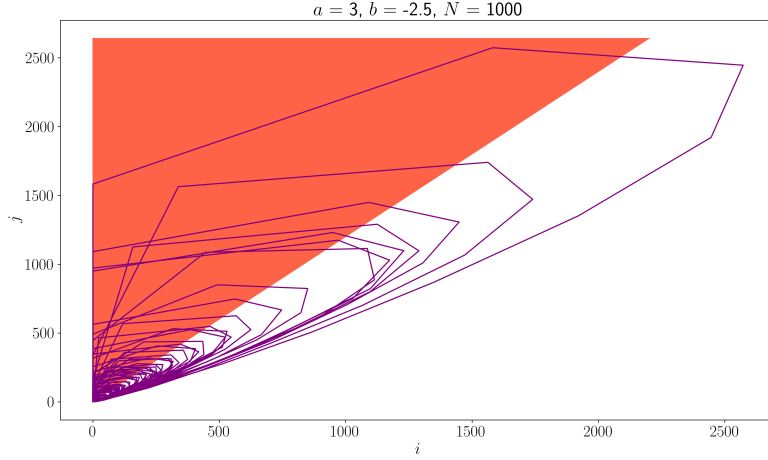


Figure 3.3 – An illustration of Case \mathcal{R}_2 . Here, the parameters are $a = 3, b = -2.5$ and $N = 1000$. In red, the set A of couples (i, j) such that $s_{ij} = s_{0i} = 0$.

function. Indeed, it is an increasing function of the angle between the vector (i, j) and the j -axis, and therefore $V(X_n)$ should have a tendency to decrease whenever X_n is far away from the j -axis.

We now turn to the details. We will need to distinguish the region A of the states (i, j) where $s_{ij} = 0$ (shown in red in Figure 3.3):

$$A := \{(i, j) \in \mathbb{N}^2 : s_{ij} = 0\} = \{(i, j) \in \mathbb{N}^2 : ai + bj + \lambda \leq 0\}. \quad (3.3.1)$$

We have the following lemma:

Lemma 3.3.2. *The set A is petite.*

Proof. By definition of A , we have $s_{ij} = 0$ for all $(i, j) \in A$, hence $P((i, j), (0, i)) = 1$. Furthermore, for every $i \in \mathbb{N}$, since $b < -a^2/4 < 0$, we have

$$P((0, i), (0, 0)) = e^{-s_{0i}} = e^{-(\lambda+bi)_+} \geq e^{-\lambda}.$$

It follows that

$$\inf_{(i,j) \in A} P^2((i, j), (0, 0)) \geq e^{-\lambda} > 0,$$

which shows that A is petite. □

Lemma 3.3.3. *There exists a finite set $C \subset \mathbb{N}^2$ and $\varepsilon \in (0, 1)$, such that the drift condition $D(V, \varepsilon, K, A \cup C)$ is satisfied for some $K < \infty$.*

Proof. Since $\frac{a^2}{4} + b < 0$, there exists $\varepsilon \in (0, 1)$ small enough such that $\frac{a^2}{4} + b < b\varepsilon < 0$.

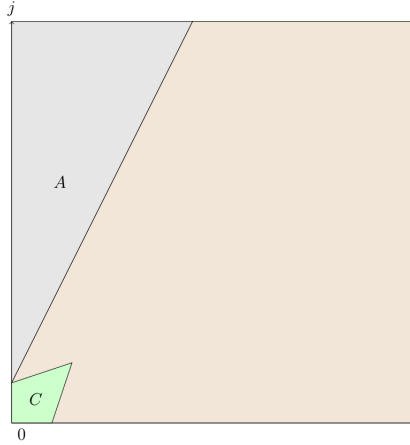


Figure 3.4 – Graphical representation of the sets A, C described above.

Consider $(i, j) \notin A$, and compute :

$$\Delta V(i, j) + \varepsilon V(i, j) = \frac{(\varepsilon - 1)i^2 + bj^2 + aij + L(i, j)}{(i + 1)(j + 1)},$$

where $L(i, j)$ is a polynomial of degree 1.

On the numerator we recognize a quadratic form, and as $\frac{a^2}{4} + b(1 - \varepsilon) < 0$, we have that this quadratic form is negative-definite. Thus, there is only a finite number of $(i, j) \notin A$ such that $\Delta V(i, j) + \varepsilon V(i, j) > 0$. We define $C \subset \mathbb{N}^2 \setminus A$ to be the finite set of such (i, j) .

Note that for every $(i, j) \in A$, we have

$$\Delta V(i, j) + \varepsilon V(i, j) \leq \mathbb{E}_{(i, j)}[V(X_1)] = V(0, i) = 1.$$

Hence, setting $K = 1 \vee \max_{x \in C} \mathbb{E}_x[V(X_1)] \in [1, \infty)$, the finiteness of K following from the fact that C is finite, we have that the drift condition $D(V, \varepsilon, K, A \cup C)$ is satisfied.

Figure 3.4 illustrates the cutting of the state space that we just described. \square

In the case \mathcal{R}_2 , by Lemma 3.3.2 and Lemma 3.3.3, we can now apply Proposition 4.3.3. Note that $A \cup C$ is petite because A is petite (Lemma 3.3.2), C is finite, hence petite, and the union of two petite sets is again petite. This yields the proof of the case R2 of Theorem 3.2.2.

3.3.4 Case \mathcal{R}_3

To finish the proof of Theorem 3.2.2, it suffices to consider parameters a and b such that $1 \leq a < 2$ and $-\frac{a^2}{4} < b < 1 - a$. However, for the sake of conciseness, we will prove the ergodicity of the Markov chain on a larger space, namely \mathcal{R}_3 . As a consequence, this case will cover some parameters sets which have already been

56 Chapter 3. Discrete-time Hawkes process with length two memory

considered in case \mathcal{R}_2 . Note that this does not represent any issue in our strategy of proof. The choice of \mathcal{R}_3 will become clearer later on.

We will thus assume here that $1 \leq a < 2$ and $-1 < b < 1 - a$. Let us denote by V the following function :

$$\forall (i, j) \in \mathbb{N}^2, \quad V(i, j) := 1 + \left(i^2 - aij + \frac{b^2 + 1}{2} j^2 \right) \mathbb{1}_{A^c}(i, j).$$

First notice that the quadratic form in V is positive-definite.

Indeed, if $1 \leq a < 2$, then $b^2 > (1 - a)^2$ and :

$$4 \times \frac{b^2 + 1}{2} - a^2 > 2(1 - a)^2 + 2 - a^2 = (a - 2)^2 > 0.$$

Thus, the function V satisfies $V \geq 1$.

Compute, for $(i, j) \notin A$ and $\varepsilon \in (0, 1)$ to be properly chosen later :

$$\begin{aligned} \Delta V(i, j) + \varepsilon V(i, j) &= \sum_{k=0}^{\infty} \frac{e^{-s_{ij}} s_{ij}^k}{k!} V(k, i) + (\varepsilon - 1)V(i, j) \\ &\leq \sum_{k=0}^{\infty} \frac{e^{-s_{ij}} s_{ij}^k}{k!} \left(1 + \left(k^2 - aki + \frac{b^2 + 1}{2} i^2 \right) \right) + (\varepsilon - 1)V(i, j) \\ &= s_{ij}(s_{ij} + 1) - ais_{ij} + \frac{b^2 + 1}{2} i^2 + (\varepsilon - 1)V(i, j) + 1 \\ &= \left(\frac{b^2 - 1}{2} + \varepsilon \right) i^2 + a(b + 1 - \varepsilon)ij + \left(\frac{b^2(1 + \varepsilon) + \varepsilon - 1}{2} \right) j^2 + L(i, j), \end{aligned}$$

where $L(i, j)$ is a polynomial of degree 1.

We want to choose $\varepsilon \in (0, 1)$ such that the above quadratic form is negative-definite, that is, such that :

$$\frac{b^2 - 1}{2} + \varepsilon < 0, \quad \text{and} \quad \left(\frac{b^2 - 1}{2} + \varepsilon \right) \left(\frac{b^2(1 + \varepsilon) + \varepsilon - 1}{2} \right) - \frac{a^2}{4}(b + 1 - \varepsilon)^2 > 0. \quad (3.3.2)$$

On the one hand, we have $b^2 - 1 < 0$. On the other hand, the second inequality in (3.3.2) can be written as follows :

$$(b^2 - 1)^2 - a^2(b + 1)^2 + k_{\varepsilon, a, b} > 0,$$

where $k_{\varepsilon, a, b} \in \mathbb{R}$ satisfies $k_{\varepsilon, a, b} \xrightarrow{\varepsilon \rightarrow 0} 0$.

In addition, note that :

$$(a, b) \in \mathcal{R}_3 \implies (b^2 - 1)^2 - a^2(b + 1)^2 > 0.$$

From the foregoing, we deduce that there exists $\varepsilon \in (0, 1)$ small enough such that both conditions of (3.3.2) are satisfied. Thus, there is only a finite number of $(i, j) \notin A$ such that $\Delta V(i, j) + \varepsilon V(i, j) > 0$. We define $C \subset \mathbb{N}^2 \setminus A$ to be the finite set of such (i, j) .

Finally, similarly as in Lemma 3.3.2, the set A is petite, because $b < 1 - a \leq 0$. Furthermore similarly as in the case \mathcal{R}_2 , for all $(i, j) \in A$, $\mathbb{E}_{(i,j)}(V(X_1)) = V(0, i)$ is bounded, since $(0, i) \in A$ except for a finite number of i . Since the set C is finite, we have that $A \cup C$ is a petite set and up to an adequate choice of K the drift condition $D(V, \varepsilon, K, A \cup C)$ is satisfied.

3.4 Proof of Theorem 3.2.2: transience

In this section, we show that the Markov chain $(X_n)_{n \geq 0}$ is transient in the regime \mathcal{T} of the parameters. We will distinguish between the following two cases:

- Case T1 : $a < 0, b > 1$ (section 3.4.1).
- Case T2 : $0 \leq a < 2$ and $a + b > 1$ or $a \geq 2$ and $a^2 + 4b > 0$ (section 3.4.2).

In both cases, we will apply the following lemma:

Lemma 3.4.1. *Let S_1, S_2, \dots be a sequence of subsets of \mathbb{N}^2 and $0 < m_1 < m_2 < \dots$ an increasing sequence of integers. Suppose that*

1. *On the event $\bigcap_{n \geq 1} \{X_{m_n} \in S_n\}$, we have $X_n \neq (0, 0)$ for all $n \geq 1$,*
2. *$\mathbb{P}_{(0,0)}(X_{m_1} \in S_1) > 0$ and for all $n \geq 1$ and every $x \in S_n$, we have $\mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) > 0$.*
3. *There exist $(p_n)_{n \geq 1}$ taking values in $[0, 1]$ and such that $\sum_{n \geq 1} (1 - p_n) < \infty$, such that*

$$\forall n \geq 1 : \forall x \in S_n : \mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) \geq p_n.$$

Then the Markov chain $(X_n)_{n \geq 0}$ is transient.

Proof. Since $(0, 0)$ is an accessible state, it is enough to show that

$$\mathbb{P}_{(0,0)}(X_n \neq (0, 0), \forall n \geq 1) > 0.$$

Using assumption 1, it is sufficient to prove that

$$\mathbb{P}_{(0,0)}(X_{m_n} \in S_n, \forall n \geq 1) > 0. \quad (3.4.1)$$

By assumption 3, there exists $n_0 \geq 1$ such that $\prod_{n \geq n_0} p_n > 0$. It follows that for every $x \in S_{n_0}$,

$$\mathbb{P}_x(X_{m_n - m_{n_0}} \in S_n, \forall n > n_0) \geq \prod_{n \geq n_0} p_n > 0.$$

Furthermore, by assumption 2, we have that

$$\mathbb{P}_{(0,0)}(\forall n \leq n_0, X_{m_n} \in S_n) > 0.$$

Combining the last two inequalities yields (3.4.1) and finishes the proof. \square

3.4.1 Case T1

In this region of parameters, the Markov chain eventually reaches the i and j axes. Indeed, since $a < 0$, if (X_n) hits a state $(i, 0)$ with $i \geq -\frac{\lambda}{a}$, as $s_{i0} = (ai + \lambda)_+ = 0$, the next step of the Markov chain will be $(0, i)$. Afterwards, the Markov chain will hit the state $(\mathcal{P}(bi + \lambda), 0)$, with $bi + \lambda > i$. Consequently, to follow the example, if we focus on the i axis, starting from $(k, 0)$ with k big enough, the Markov chain will return in two steps to a state $(k', 0)$ belonging to the i -axis, satisfying $k' > k$ with high probability.

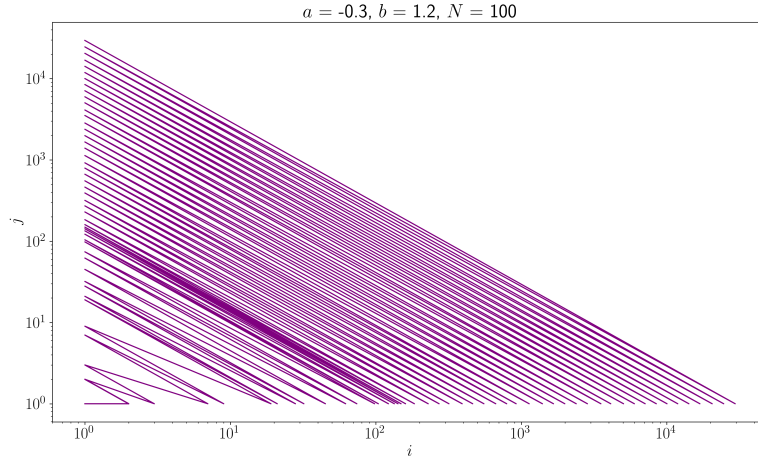


Figure 3.5 – Log-log plot of a typical trajectory of (X_n) , to make the erratic behaviour of the first points of the Markov chain more visible. Here, the parameters are $a = -0.3, b = 1.2$ and $N = 100$.

In order to formalize these observations, it is very natural to consider the Markov chain induced by the transition matrix P^2 , namely $(X_{2n+1})_{n \geq 0}$. For $i \geq -\frac{\lambda}{a}$, $s_{i0} = 0$ and thus :

$$\begin{aligned} \mathbb{P}\left(X_{2n+1} = (k, 0) \mid X_{2n-1} = (i, 0), i \geq \frac{-\lambda}{a}\right) &= \frac{e^{-s_{0i}} s_{0i}^k}{k!} \\ &= \frac{e^{-(bi+\lambda)} (bi + \lambda)^k}{k!}. \end{aligned} \quad (3.4.2)$$

Note that if $a \leq -\lambda$, this results holds for $i \in \mathbb{N}$.

Equation (4.4.2) means that if $\tilde{X}_{2n-1} \geq -\frac{\lambda}{a}$, and $\tilde{X}_{2n-2} = 0$, then $\tilde{X}_{2n} = 0$, and \tilde{X}_{2n+1} is a Poisson random variable with parameter $b\tilde{X}_{2n-1} + \lambda$.

Let us now prove our statement.

Proof of the transience of (X_n) when $a < 0$ and $b > 1$. Fix $r \in (1, b)$. We wish to apply Lemma 3.4.1 with

$$m_n = 2n - 1, \quad n \geq 1$$

and

$$S_n = \{(i, 0) \in \mathbb{N} : i \geq r^n\}.$$

We verify that the assumptions (1)-(3) from Lemma 3.4.1 hold. For the first assumption, note that if $X_{2n-1} = (i, 0) \in S_n$, then $X_{2n} = (j, i)$ for some j , hence $X_{2n-1} \neq (0, 0)$ and $X_{2n} \neq (0, 0)$ since $i \geq 1$. In particular, assumption (1) holds.

We now verify that the second assumption holds. For states $x, y \in \mathbb{N}^2$, write $x \rightarrow_1 y$ if $\mathbb{P}_x(X_1 = y) > 0$. Furthermore, for $S \subset \mathbb{N}^2$, write $x \rightarrow_1 S$ if $x \rightarrow_1 y$ for some $y \in S$. Note that $(0, 0) \rightarrow_1 (i, 0)$ for every $i \in \mathbb{N}$, so that $(0, 0) \rightarrow_1 S_1$. Now, for every $i \in \mathbb{N}$, we have $(i, 0) \rightarrow_1 (0, i)$, and then, because $b > 0$, $(0, i) \rightarrow_1 (j, 0)$ for every $j \in \mathbb{N}$. In particular, from every $x \in S_n$, we can indeed reach S_{n+1} in two steps. Hence, the second assumption is verified as well.

We now prove the third assumption. We claim that there exists $n_0 \in \mathbb{N}$, such that the following holds:

$$\forall n \geq n_0, \forall x \in S_n : \mathbb{P}_x(X_2 \in S_{n+1}) \geq 1 - \frac{b}{(b-r)^2 r^n}. \quad (3.4.3)$$

To prove (4.4.3), first note that according to the earlier remark on (4.4.2), if n_0 is chosen such that $r^{n_0} \geq -\lambda/a$, then starting from a state $(i, 0)$ with $i \geq r^{n_0}$, we have $\tilde{X}_1 = 0$ almost surely and $\tilde{X}_2 \sim \mathcal{P}(bi + \lambda)$. Therefore, if $n \geq n_0$ and $i \geq r^n \geq r^{n_0}$,

$$\begin{aligned} 1 - \mathbb{P}_{(i,0)}(\tilde{X}_2 \geq r^{n+1}, \tilde{X}_1 = 0) &= \mathbb{P}_{(i,0)}(\tilde{X}_2 < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}(bi + \lambda) < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}(br^n) < r^{n+1}) \\ &= \mathbb{P}(\mathcal{P}(br^n) - br^n < r^n(r - b)) \\ &\leq \mathbb{P}(|\mathcal{P}(br^n) - br^n| > r^n(b - r)) \\ &\leq \frac{b}{(b-r)^2 r^n}, \end{aligned}$$

by the Bienaymé-Chebychev inequality. This proves (4.4.3). Now, (4.4.3) implies,

$$\forall x \in S_n : \mathbb{P}_x(X_2 \in S_{n+1}) \geq p_n := \left(1 - \frac{b}{(b-r)^2 r^n}\right)_+,$$

and

$$\sum_{n \geq 1} (1 - p_n) \leq \sum_{n \geq 1} \frac{b}{(b-r)^2 r^n} < \infty.$$

This proves that the third assumption of Lemma 3.4.1 holds. The lemma then shows that the Markov chain is transient. \square

3.4.2 Case T2: $0 \leq a < 2$ and $a + b > 1$ or $a \geq 2$ and $a^2 + 4b > 0$

For this case, we will take benefit of the comparison between the stochastic process (\tilde{X}_n) and its linear deterministic version. Namely, let us consider the linear recurrence relation defined by $y_0, y_1 \in \mathbb{N}$ and :

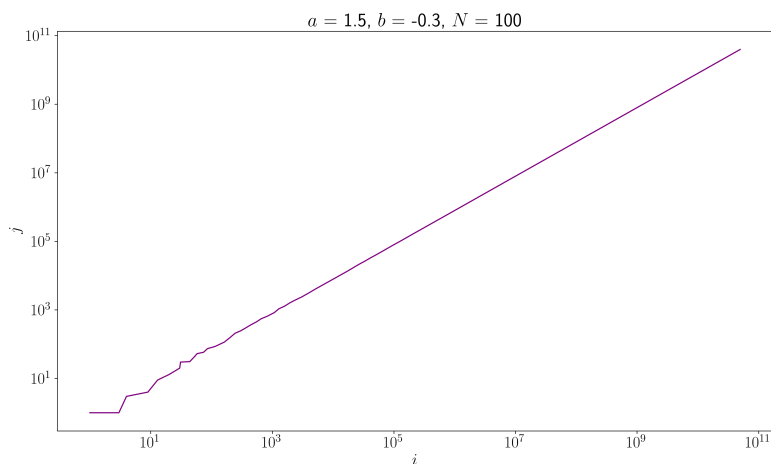


Figure 3.6 – Log-log plot of a typical trajectory of (X_n) , with $a = 1.5, b = -0.3$ and $N = 100$

$$\forall n \geq 0, \quad y_{n+2} = ay_{n+1} + by_n + \lambda. \quad (3.4.4)$$

The solutions to this equation are determined by the eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 0 & b \\ 1 & a \end{pmatrix}$, which is the companion matrix of the polynomial $X^2 - aX - b$ (see Appendix A. for more details).

An easy calculation shows that in case T2, we have $a^2 + 4b > 0$, hence the eigenvalues are simple and real-valued. We denote the largest eigenvalue by:

$$\theta := \frac{a + \sqrt{a^2 + 4b}}{2}.$$

The following holds in case T2, as can be easily verified:

$$\theta > 1 \quad (3.4.5)$$

$$\theta^2 + b > 0. \quad (3.4.6)$$

In fact, one can check that case T2 exactly corresponds to the region in the space of parameters a, b where $\theta > 1$, meaning that the sequence $(y_{n+1}, y_n)_{n \geq 0}$, with $(y_n)_{n \geq 0}$ the solution to (3.4.4), grows exponentially inside the positive quadrant, along the direction of the eigenvector $(\theta, 1)$.

In what follows, we fix $1 < r < \theta$, such that

$$r^2 - ar - b < 0, \quad (3.4.7)$$

where we use the fact that $\theta > 1$ is the largest root of the polynomial $X^2 - aX - b$.

We split our study into two different sub-cases depending on the sign of b .

3.4.2.1 Subcase T2a : $b \geq 0$

In this case, we have $a\tilde{X}_n + b\tilde{X}_{n-1} + \lambda > 0$ for all $n \in \mathbb{N}$, and so $\tilde{X}_{n+1} \sim \mathcal{P}(a\tilde{X}_n + b\tilde{X}_{n-1} + \lambda)$, i.e. no truncation is necessary. It is classical that in this case, \tilde{X}_n grows exponentially in n almost surely, but we provide a simple proof for completeness.

We therefore apply Lemma 3.4.1 with the sequence $m_n = n$ and

$$S_n = \{(i, j) \in \mathbb{N}^2, i \geq r^n, j \geq r^{n-1}\}.$$

With these notations, assumption 1 is automatically satisfied. Assumption 2 is also satisfied, because $(i, j) \rightarrow_1 (k, i)$ for every $i, j, k \in \mathbb{N}$, since $ai + bj + \lambda > 0$ for every $i, j \in \mathbb{N}$ as explained above.

In order to prove assumption 3, let us consider $n \in \mathbb{N}$ and let $(i, j) \in S_n$. By definition, starting from (i, j) , $\tilde{X}_1 \sim \mathcal{P}(ai + bj + \lambda)$. Thus,

$$\begin{aligned} \mathbb{P}_{(i,j)}(\tilde{X}_1 < r^{n+1}) &= \mathbb{P}(\mathcal{P}(ai + bj + \lambda) < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}(ar^n + br^{n-1}) < r^{n+1}) \\ &= \mathbb{P}(\mathcal{P}(ar^n + br^{n-1}) - (ar^n + br^{n-1}) < r^{n-1}(r^2 - ar - b)). \end{aligned}$$

Recall that $r^2 - ar - b < 0$ by (3.4.7), which implies :

$$\begin{aligned} \mathbb{P}_{(i,j)}(\tilde{X}_1 < r^{n+1}) &\leq \mathbb{P}(|\mathcal{P}(ar^n + br^{n-1}) - (ar^n + br^{n-1})| > -r^{n-1}(r^2 - ar - b)) \\ &\leq \frac{(a+b)r^2}{r^n(r^2 - ar - b)^2}, \end{aligned}$$

where we used again the Bienaymé-Chebychev inequality. Thus

$$\mathbb{P}_{(i,j)}(X_1 \in S_{n+1}) \geq \left(1 - \frac{(a+b)r^2}{r^n(r^2 - ar - b)^2}\right)_+ =: p_n.$$

This allows to conclude the proof with Lemma 3.4.1, as in the previous case.

3.4.2.2 Subcase T2b : $b < 0$

In this case, because of the negativity of b it is more difficult to find an adequate lower-bound of $a\tilde{X}_n + b\tilde{X}_{n-1}$. We will thus prove a stronger result, which is illustrated on Figure 3.6 : asymptotically, the process (\tilde{X}_n) grows exponentially and the ratio $\tilde{X}_{n+1}/\tilde{X}_n$ is close to θ .

From (3.4.7) and (3.4.6), we can choose $\varepsilon > 0$ small enough such that :

$$r^2 - a(r - \varepsilon) - b < 0 \tag{3.4.8}$$

$$\theta^2 - \theta\varepsilon + b > 0. \tag{3.4.9}$$

We will use Lemma 3.4.1 using $m_n = n$ and for $n \in \mathbb{N}^*$:

$$S_n = \left\{ (i, j) \in \mathbb{N}^2, i \geq r^n, j \geq r^{n-1}, \left| \frac{i}{j} - \theta \right| \leq \varepsilon \right\}.$$

62 Chapter 3. Discrete-time Hawkes process with length two memory

Note that Assumption 1 from Lemma 3.4.1 is again automatically verified. Assumption 2 is also verified, since for $(i, j) \in S_n$, we have

$$ai + bj + \lambda > (a(\theta - \varepsilon) + b)j \geq (a(r - \varepsilon) + b)j > 0, \quad (3.4.10)$$

by (3.4.8), and so $(i, j) \rightarrow_1 (k, i)$ for every $k \in \mathbb{N}$.

We now show that Assumption 3 from Lemma 3.4.1 is verified. Let $n \in \mathbb{N}$ and $(i, j) \in S_n$. Then :

$$\mathbb{P}_{(i,j)}(X_1 \notin S_{n+1}) \leq \mathbb{P}_{(i,j)}(\tilde{X}_1 < r^{n+1}) + \mathbb{P}_{(i,j)}\left(\left|\frac{\tilde{X}_1}{i} - \theta\right| > \varepsilon\right) \quad (3.4.11)$$

We first bound the first term on the right-hand side of (3.4.11). By (3.4.10), we have

$$\begin{aligned} & \mathbb{P}_{(i,j)}(\tilde{X}_1 < r^{n+1}) \\ &= \mathbb{P}(\mathcal{P}(ai + bj + \lambda) < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}([a(r - \varepsilon) + b]r^{n-1}) < r^{n+1}) \\ &= \mathbb{P}(\mathcal{P}([a(r - \varepsilon) + b]r^{n-1}) - [a(r - \varepsilon) + b]r^{n-1} < r^{n-1}[r^2 - a(r - \varepsilon) - b]). \end{aligned}$$

Furthermore, using (3.4.8) and applying the Bienaymé-Chebychev inequality we obtain :

$$\begin{aligned} & \mathbb{P}_{(i,j)}(\tilde{X}_1 < r^{n+1}) \\ &\leq \mathbb{P}\left(\left|\mathcal{P}([a(r - \varepsilon) + b]r^{n-1}) - [a(r - \varepsilon) + b]r^{n-1}\right| > -r^{n-1}[r^2 - a(r - \varepsilon) - b]\right) \\ &\leq \frac{[a(r - \varepsilon) + b]r^{n-1}}{\left[r^{n-1}[r^2 - a(r - \varepsilon) - b]\right]^2} \\ &= \frac{[a(r - \varepsilon) + b]r}{r^n[r^2 - a(r - \varepsilon) - b]^2} = \frac{C_1}{r^n}, \end{aligned} \quad (3.4.12)$$

where C_1 is a constant that does not depend on n .

We now bound the second term on the right-hand side of (3.4.11). Let us write :

$$\left|\frac{\tilde{X}_1}{i} - \theta\right| = \left|\frac{\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]}{i}\right| + \left|\frac{\mathbb{E}_{(i,j)}[\tilde{X}_1]}{i} - \theta\right|$$

First notice that, for any $(i, j) \in S_n$:

$$\begin{aligned} \left| \frac{\mathbb{E}_{(i,j)}[\tilde{X}_1]}{i} - \theta \right| &= \left| \frac{ai + bj + \lambda}{i} - \theta \right| = \left| a + b\frac{j}{i} + \frac{\lambda}{i} - \theta \right| \\ &\leq \underbrace{\left| a + \frac{b}{\theta} - \theta \right|}_{=0} + |b| \left| \frac{j}{i} - \frac{1}{\theta} \right| + \frac{\lambda}{i} \\ &< \frac{|b|}{\theta(\theta - \varepsilon)} \varepsilon + \frac{\lambda}{i}, \end{aligned} \tag{3.4.13}$$

where we used that if $|x - \theta| < \varepsilon$ and $\varepsilon < \theta$, then

$$\left| \frac{1}{x} - \frac{1}{\theta} \right| = \frac{|\theta - x|}{x\theta} < \frac{\varepsilon}{\theta(\theta - \varepsilon)}.$$

To prove that $\mathbb{P}_{(i,j)} \left(\left| \frac{\tilde{X}_1}{i} - \theta \right| > \varepsilon \right) \leq \frac{C_2}{r^n}$, where C_2 is a constant that does not depend on n , we deduce from (3.4.13) that it is sufficient to show that :

$$\mathbb{P}_{(i,j)} \left(\frac{|\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]| + \lambda}{i} > \delta\varepsilon \right) \leq \frac{C_2}{r^n},$$

where, by (3.4.9), we have

$$\delta := 1 - \frac{|b|}{\theta(\theta - \varepsilon)} > 0.$$

Furthermore, since $b < 0$ and $(i, j) \in S_n$, we have that

$$ai + bj + \lambda \leq ai + \lambda \leq (a + \lambda)i.$$

We finally have, using the Bienaymé-Chebyshev inequality :

$$\begin{aligned} \mathbb{P}_{(i,j)} \left(\frac{|\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]| + \lambda}{i} > \delta\varepsilon \right) &= \mathbb{P}_{(i,j)} \left(|\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]| > \delta\varepsilon i - \lambda \right) \\ &\leq \frac{(a + \lambda)i}{(\delta\varepsilon i - \lambda)^2} \\ &= \frac{a + \lambda}{(\delta\varepsilon\sqrt{i} - \lambda/\sqrt{i})^2} \\ &\leq \frac{a + \lambda}{(\delta\varepsilon r^{n/2} - \lambda r^{-n/2})^2} \end{aligned}$$

Last inequality holds for a sufficiently large n . Indeed, since $i \geq r^n$, we always have

$$\delta\varepsilon\sqrt{i} - \lambda/\sqrt{i} > \delta\varepsilon r^{n/2} - \lambda r^{-n/2},$$

and for n large enough we have

$$\delta\varepsilon\sqrt{i} - \frac{\lambda}{\sqrt{i}} > 0.$$

This yields, for some constant $C_2 < \infty$,

$$\mathbb{P}_{(i,j)} \left(\frac{|\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]| + \lambda}{i} > \left(1 - \frac{|b|}{\theta(\theta - \varepsilon)}\right) \varepsilon \right) \leq \frac{C_2}{r^n}. \quad (3.4.14)$$

Combining (3.4.12) and (3.4.14) we have that :

$$\mathbb{P}_{(i,j)}(X_1 \in S_{n+1}) \geq \left(1 - \frac{C_1 + C_2}{r^n}\right)_+ =: p_n.$$

Which will finally lead us to the result, by using Lemma 3.4.1 as before.

3.5 Perspectives and open problems

Critical behavior

In the case of linear Hawkes processes, it is well known that, at criticality, the process achieves fractal-like, i.e. heavy-tail behavior, related to critical branching processes. It is tempting to believe that this should remain true on the whole boundary between the phases \mathcal{R} and \mathcal{T} , but the fractal exponents might differ.

Numerical experiments indicates different behavior depending whether $a < 2$ or $a > 2$. On Figure 3.7, we observe a linear growth of the discrete time process \tilde{X}_n , with oscillations to 0 when $a < 0$ and $b = 1$ (left panel) and without oscillations in the case $a + b = 1$. The situation seems to be different for $a \geq 2$ and $b = -a^2/4$.

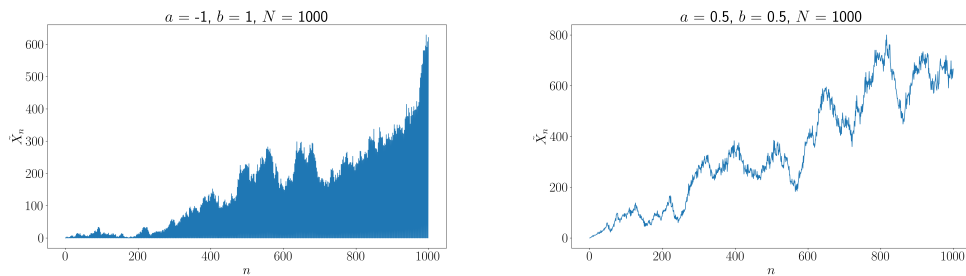


Figure 3.7 – Trajectories of $(\tilde{X}_n)_{0 \leq n \leq 1000}$ for critical parameters $(a, b) = (-1, 1)$ on the left and $(a, b) = (0.5, 0.5)$ on the right. We observe a linear growth of the process as it could be expected in critical cases.

When $a > 2$ we observe an exponential growth (Fig. 3.8 (left)) similarly to the transient regime, while the case $a = 2$ presents large excursions away from 0 but deciphering transient or recurrent behavior is difficult. These simulations show that the study of these critical cases is an interesting research topic for the future.

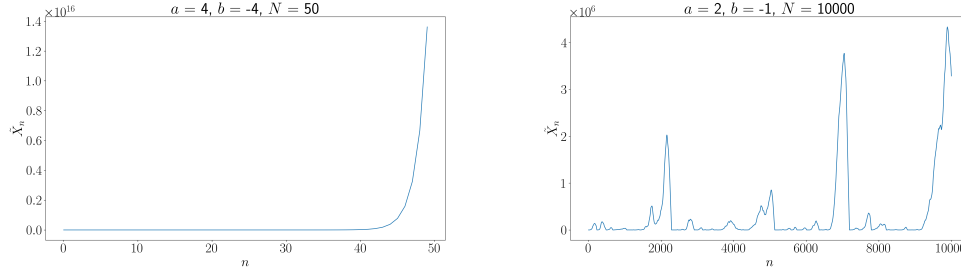


Figure 3.8 – Trajectories of $(\tilde{X}_n)_{0 \leq n \leq 10000}$ for critical parameters $(a, b) = (4, -4)$ on the left and $(a, b) = (2, -1)$ on the right.

Generalization of the model

As explained at the beginning of the article, the results obtained here should be seen as a starting point for the search for necessary and sufficient conditions for the stability of Hawkes processes with inhibition, in discrete or continuous time.

We believe that obtaining a similar classification in the cases $p > 2$ or $p = \infty$ is a very difficult problem. It should be closely related to the study of the asymptotic behavior of certain deterministic equations, such as the non-linear recurrence equation

$$x_n = (a_1 x_{n-1} + \cdots + a_p x_{n-p})_+.$$

It seems that the algebraic structures underlying these equations are intricate and, to this date, unknown. The understanding of these structures seem crucial for the study of the asymptotic behavior of the solutions to these equations. In a work in progress, we achieve partial results in this direction.

3.A Linear recurrence equations

Let $\alpha \in \mathbb{R}$, $p \in \mathbb{N}$ and $a_1, \dots, a_p \in \mathbb{R}$. Consider the linear recurrence equation

$$x_n = a_1 x_{n-1} + \cdots + a_p x_{n-p} + \alpha, \quad n \geq 1,$$

with given initial data $x_0, \dots, x_{-p+1} \in \mathbb{R}$. Define the matrix

$$A = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{pmatrix},$$

where vanishing entries are meant to be zero.

Then, setting

$$\bar{x}_n = \begin{pmatrix} x_n \\ \vdots \\ x_{n-p+1} \end{pmatrix} \quad \text{and} \quad \bar{\alpha} = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

the sequence $(\bar{x}_n)_{n \geq 1}$ solves the system of linear recurrences

$$\bar{x}_n = A\bar{x}_{n-1} + \bar{\alpha}, \quad n \geq 1.$$

Recall that the spectral radius $\rho(A)$ of the matrix A is defined by

$$\rho(A) = \max(|\theta_1|, \dots, |\theta_p|),$$

where $\theta_1, \dots, \theta_p \in \mathbb{C}$ are the complex eigenvalues of A , counted with algebraic multiplicity. Equivalently, $\theta_1, \dots, \theta_p$ are the roots, counted with multiplicity, of the characteristic polynomial

$$P(z) = \det(zI - A) = z^p - a_1 z^{p-1} - \dots - a_p.$$

We recall the following classical fact:

Theorem 3.A.1 ([34] Chapter 9, Thm 9.1). *The following are equivalent:*

1. \bar{x}_n converges as $n \rightarrow \infty$, for every initial data x_0, \dots, x_{-p+1}
2. $\rho(A) < 1$

In the case $p = 2$, setting $a = a_1$ and $b = a_2$, we have $\mathbb{P}(z) = z^2 - az - b$. Its roots are

$$\theta_{\pm} = \frac{a}{2} \pm \sqrt{\frac{a^2}{4} + b}.$$

In particular,

$$\rho(A) = \begin{cases} \frac{|a|}{2} + \sqrt{\frac{a^2}{4} + b}, & \text{if } \frac{a^2}{4} + b \geq 0 \\ \sqrt{-b}, & \text{if } \frac{a^2}{4} + b < 0. \end{cases}$$

A quick calculation shows that

$$\rho(A) < 1 \iff |a| + b < 1 \text{ and } b > -1. \quad (3.A.1)$$

This corresponds to the triangular dashed region of parameters in Figure 3.1 of section 2.

3.B Criteria for strong irreducibility

The Markov chain considered in this article is irreducible in the (weak) sense of Douc, Moulines, Priouret and Soulier [19], but not necessarily strongly irreducible, i.e. irreducible in the classical sense. In this section, we study the decomposition of the state space into communicating classes. We recall the basic definitions. Let $x, y \in \mathbb{N}^2$. We say that x **leads to** y , or, in symbols, $x \rightarrow y$, if there exists $n \geq 0$ such that $\mathbb{P}(X_n = y \mid X_0 = x) > 0$. We say that x **communicates with** y if $x \rightarrow y$ and $y \rightarrow x$. This is an equivalence relation which partitions the state space \mathbb{N}^2 into classes called **communicating classes**.

Recall that the Markov chain is called **strongly irreducible** if all states are accessible, equivalently, if \mathbb{N}^2 is a communicating class. A communicating class $C \subset \mathbb{N}^2$ is called **closed** if there does not exist $x \in C$ and $y \in C^c$, such that $x \rightarrow y$.

Proposition 3.B.1.

- The Markov chain (X_n) is strongly irreducible on \mathbb{N}^2 if and only if $a \geq 0$, or if $a > -\lambda$ and $a + b \geq 0$.
- The communicating class of $(0, 0)$ contains

$$\mathcal{S} = \{(0, 0)\} \cup \{(0, k), k \in \mathbb{N}^*\} \cup \{(k, 0), k \in \mathbb{N}^*\}. \quad (3.B.1)$$

and is actually equal to \mathcal{S} iff $a \leq -\lambda$.

We will use the following :

Lemma 3.B.2. Let $i, j, k, \ell \in \mathbb{N}$. The transition matrix P of the Markov chain (X_k) satisfies :

$$P^2((i, j), (k, \ell)) = \frac{e^{-(s_{ij} + s_{\ell i})} s_{ij}^\ell s_{\ell i}^k}{\ell! k!},$$

and for all $n \geq 3$,

$$P^n((i, j), (k, \ell)) = \sum_{m_1, \dots, m_{n-2} \in \mathbb{N}} \frac{\exp \left\{ - \sum_{q=1}^n s_{\sigma_{q+1}^n \sigma_{q+2}^n} \right\} \prod_{q=1}^n s_{\sigma_{q+1}^n \sigma_{q+2}^n}}{m_1! \dots m_{n-2}! k! \ell!}, \quad (3.B.2)$$

with $\sigma^n := (\sigma_1^n, \sigma_2^n, \dots, \sigma_{n+2}^n) = (k, \ell, m_{n-2}, \dots, m_1, i, j)$.

Proof of proposition 3.B.1. As mentioned above:

$$(i, j) \rightarrow (0, i) \rightarrow (0, 0),$$

for any $(i, j) \in \mathbb{N}^2$ since it only requires that 2 successive 0 are drawn from the Poisson random variable. Therefore, to prove strong irreducibility, it is thus sufficient to prove that $(0, 0) \rightsquigarrow (i, j)$, for all $(i, j) \in \mathbb{N}^2$. Let us consider different cases, depending on the values of the parameters a and b .

– **If $a \geq 0$:** Since $\lambda > 0$, $s_{00} > 0$ thus $(j, 0)$ is accessible from $(0, 0)$, for all $j \in \mathbb{N}$. Moreover, when $a \geq 0$, $s_{j0} = (aj + \lambda)_+ > 0$ and then $(j, 0) \rightarrow (i, j)$, yielding the result.

– **If $-\lambda < a < 0$ and $a + b \geq 0$:** Let $k \in \mathbb{N}$. Since $a + b \geq 0$ and $a + \lambda > 0$, we have :

$$s_{k+1, k} = (a(k+1) + bk + \lambda)_+ = ((a+b)k + a + \lambda)_+ > 0,$$

Let $(i, j) \in \mathbb{N}^2$. Since $s_{k+1, k} > 0$ for all k , we deduce that any $(\ell, k+1)$ is accessible from $(k+1, k)$. Thus, in order to reach (i, j) from $(0, 0)$, we move from small steps to $(j, j-1)$, and then reach (i, j) :

$$(0, 0) \rightarrow (1, 0) \rightarrow (2, 1) \rightarrow \dots \rightarrow (j, j-1) \rightarrow (i, j).$$

which concludes the proof of this case.

– **If $a \leq -\lambda$:** We will prove that the communicating class of $(0, 0)$ is given by (3.B.1).

Let $k \in \mathbb{N}^*$, then as previously, we have that $(0, 0) \rightarrow (k, 0)$ since $s_{00} > 0$, however since $a \leq -\lambda$, $s_{k0} = (ak + \lambda)_+ = 0$, and the next step of the Markov chain will be $(0, k)$. Depending on the value of the parameter b , the next step of the Markov chain will either be $(0, 0)$ if $s_{0k} = 0$, or $(k', 0)$ with $k' \geq 0$ if $s_{0k} > 0$ and so on. This proves that the class $\mathcal{cl}(0, 0)$ is closed and given by (3.B.1).

– **If $-\lambda < a < 0$ and $a + b < 0$** In this case we can only prove that the Markov chain is not strongly irreducible on \mathbb{N}^2 but we do not identify the communicating class of $(0, 0)$.

- **Case 1 : $b \leq 0$.**

Since $a < 0$, we can choose k_* such that

$$ak_* + \lambda \leq 0.$$

We will show that it is not possible to reach the state $(1, k_*)$. Assuming the opposite leads to the existence of $\ell \in \mathbb{N}$ such that $(k_*, \ell) \rightarrow (1, k_*)$ which implies that $s_{k_*, \ell} > 0$. If $b < 0$, we deduce that necessarily

$$ak_* + b\ell + \lambda > 0 \implies \ell < \frac{-ak_* - \lambda}{b} \leq 0,$$

so $\ell < 0$ which is contradictory. We then deduce that the Markov chain is reducible.

If $b = 0$, $s_{k_*, \ell} > 0$ would imply that $ak_* + \lambda > 0$ which contradicts the definition of k_* .

- **Case 2 : $b > 0$.**

Since $a + b < 0$, it is possible to choose $k_* \in \mathbb{N}$ large enough so that :

$$(a + b)k_* + \lambda \leq 0.$$

In particular, $0 \geq ak_* + bk_* + \lambda \geq ak_* + \lambda$, so that :

$$\frac{-ak_* - \lambda}{b} \geq k_* \geq \frac{-\lambda}{a}.$$

Notice that $k_* \geq 2$ since $k_* \geq \frac{-\lambda}{a} > 1$.

We will show that it is not possible to reach $(1, k_*)$ starting from $(0, 0)$. Assuming the opposite leads us to the existence of $n \in \mathbb{N}$ such that

$$P^n((0, 0), (1, k_*)) > 0.$$

Using (3.B.2) in the Lemma 3.B.2 implies that it exists $m_1, \dots, m_{n-2} \in \mathbb{N}$ such that :

$$\begin{cases} s_{k_*, m_{n-2}} > 0 \\ s_{m_{n-2} m_{n-3}}^{k_*} > 0 \\ \dots \\ s_{m_2 m_1}^{m_3} > 0 \\ s_{m_1 0}^{m_2} > 0 \end{cases}$$

First, we thus have that :

$$ak_* + bm_{n-2} + \lambda > 0 \implies m_{n-2} > \frac{-ak_* - \lambda}{b} \geq k_*,$$

then, since $k_* > 0$, we necessarily have $s_{m_{n-2} m_{n-3}} > 0$. It yields :

$$am_{n-2} + bm_{n-3} + \lambda > 0 \implies m_{n-3} > \frac{-am_{n-2} - \lambda}{b} \geq \frac{-ak_* - \lambda}{b} \geq k_*.$$

We thus have by immediate induction that :

$$\forall i \in \{1, \dots, n-2\}, m_i \geq k_* \geq \frac{-\lambda}{a}.$$

Finally, $s_{m_1, 0} > 0$ implies $am_1 + \lambda > 0$, which is contradictory.

We conclude that there is no finite path between $(0, 0)$ and $(1, k_*)$, so the Markov chain $(X_k)_{k \geq 0}$ is reducible on \mathbb{N}^2 .

□

Sufficient conditions of stability of a discrete-time Hawkes process with inhibition and memory of length three

This self-contained chapter consists in an article that will soon be submitted. Notice that the paper [14] cited several times is the Chapter 3.

Abstract

In this paper, we study a discrete-time analogue of a Hawkes process, modelled as a Poisson autoregressive process whose parameters depend on the past of the trajectory. The model is characterized to allow these parameters to take negative values, modelling inhibitory dynamics. More precisely, the model is the stochastic process $(\tilde{X}_n)_{n \geq 0}$ with parameters $a_1, \dots, a_p \in \mathbb{R}$, $p \in \mathbb{N}$ and $\lambda > 0$, such that for all $n \geq p$, conditioned on $\tilde{X}_0, \dots, \tilde{X}_{n-1}$, \tilde{X}_n is Poisson distributed with parameter

$$\left(a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda\right)_+.$$

This process can be seen as a discrete time Hawkes process with inhibition with a memory of length p . This work is an extension of a prior work where we studied the specific case $p = 2$, for which we were able to classify the asymptotic behaviour of the process for the whole range of parameters, except for boundary cases. After giving partial results in the general case, we focus on the case $p = 3$, motivated by the dual objectives of extending the results derived for the $p = 2$ case in a previous work and enhancing, as well as broadening, our methodological approaches.

Contents

4.1	Introduction.	72
4.2	Main results	73
4.2.1	The general setting	73
4.2.2	Three-parameter discrete-time Hawkes process	74
4.2.2.1	Existing results in the case $c = 0$	75
4.2.2.2	Linear recurrence, polynomials and discriminant	76
4.2.3	Transience results	77
4.2.4	Recurrence results	78

4.2.5	Conjecture and numerical illustrations	78
4.3	Proofs of Recurrent cases	83
4.3.1	Lyapunov function and Drift criterion	83
4.3.2	Proof of Theorem 4.2.1	85
4.3.3	Proof of Theorem 4.2.3	87
4.4	Proofs of Transience cases	92
4.4.1	Case $a, b < 0, c > 1$	93
4.4.2	Case $b > 1, ab + c < 0$	94
4.A	Proof of transience when $a_1, \dots, a_p \geq 0$ and $a_1 + \dots + a_p > 1$.	97
4.B	Proofs of technical lemmas	98
4.B.1	Proof of Lemma 4.2.4	98
4.B.2	Proof of Lemma 4.3.7	99
4.B.3	Proof of Lemma 4.B.1	99
4.C	Perspectives	100

4.1 Introduction.

Hawkes processes are a class of point processes initially introduced by Hawkes in 1971 ([41], [43]). These processes became well-known within the scientific community for their applications in modelling seismic events, specifically earthquakes. The fundamental aim in the formulation of this stochastic process was to formalize the concept of *excitation* over time, which describes the phenomenon that the occurrence of an event increases the probability of another event, and so on. In a Hawkes process, the excitation’s intensity depends on the entire history of the process, and influences its behaviour over time. An essential characteristic of the Hawkes process is its non-Markovian nature (in most cases), primarily because it depends on the complete past trajectory of the process.

In recent years, there has been a rise of interest in Hawkes processes (see for example [8], [55]), particularly in generalized versions that incorporate the notion of *inhibition* ([13],[73], [12]). Inhibition denotes the possibility of existence of events that decrease the probability of future events to occur. Mathematically, adding inhibition into the intensity causes new mathematical challenges. In particular obtaining necessary and sufficient conditions for the existence of stationary version remains an open question. To address this issue, we proposed in a previous paper ([14]) an analogous, discrete, and simplified process. This approach was devised to yield insights into the simplified process that could, in turn, offer intuition for potential generalizations or interpretations concerning the continuous Hawkes process.

Our model is the following, for $p \in \mathbb{N}^*$ and initial condition $(\tilde{X}_0, \dots, \tilde{X}_{-p+1})$, the discrete-time Hawkes process with memory of length p satisfies that conditionally on X_{n-p}, \dots, X_{n-1}

$$\tilde{X}_n \sim \mathcal{P} \left(\left(a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda \right)_+ \right) \quad \forall n \geq 1, \quad (4.1.1)$$

where $\mathcal{P}(\rho)$ denotes the Poisson distribution with parameter ρ . The parameters a_1, \dots, a_p are real numbers, and $\lambda > 0$. The notation $(\cdot)_+$ refers to the ReLU function defined on \mathbb{R} by :

$$(x)_+ := \max(0, x).$$

The process \tilde{X}_n defined in (4.1.1) is called *Poisson-autoregressive* processes, and has been deeply studied in the "linear case" that is, the same definition with every parameters being positive, so that the positive part $(\cdot)_+$ vanishes in (4.1.1) (for more details see [26], [28]). In [55], Kirchner proved the weak convergence of an infinite memory version ($p = +\infty$) to the continuous-time Hawkes process in the linear case for further details. As far as we know, there are very few papers dealing with a non-linear version of such a process (for example [30]). In [14], we gave an almost complete characterization of stability of this process (excluding boundary cases) in the particular case $p = 2$.

In this paper, we investigate first investigate the general case, of p fixed to obtain partial and expected results, then, we focus on the case $p = 3$. Our intention is to take profit of the groundwork laid in our prior research project, both in terms of results and methodological approaches as expounded in [14]. More specifically, we establish some highly encouraging statements for the understanding of the processes defined by (4.1.1): indeed, results we were able to derive for the specific cases of $p = 2$ and $p = 3$ have led us to hypothesize that the behaviour of these discrete-time processes is intricately encoded within a polynomial of degree p , namely :

$$X^p - a_1 X^{p-1} - \dots - a_{p-1} X - a_p.$$

Nonetheless, we underscore the challenges encountered even in simple cases. As illustrated in this paper, where we restrict our focus to a three-parameter process, the complexity of this study is far from trivial. We were actually unable to fully classify the behaviour of (\tilde{X}_n) depending on the model parameters, but we shall provide numerical studies, accompanied by geometric interpretations and conjectural insights in some difficult cases.

4.2 Main results

4.2.1 The general setting

Let us start by considering the general case, of a Poisson auto-regressive process with memory p defined as in (4.1.1) : for $p \in \mathbb{N}^*$ and initial condition $(\tilde{X}_0, \dots, \tilde{X}_{-p+1})$, the discrete-time Hawkes process with memory of length p satisfies that conditionally on $\tilde{X}_{n-p}, \dots, \tilde{X}_{n-1}$,

$$\tilde{X}_n \sim \mathcal{P} \left(\left(a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda \right)_+ \right) \quad \forall n \geq 1,$$

Existing results for the linear case

The case where all parameters a_1, \dots, a_p are nonnegative is well understood. The process (\tilde{X}_n) is an *Integer-Valued GARCH process* (INGARCH(0, p) process, see [26] and [61]), also known in the literature as an *Auto-regressive Conditional Poisson process* (ACP(p) see [44]) and its long time behaviour depends on $\sum_{i=1}^p a_i$. Namely,

- If $\sum_{i=1}^p a_i < 1$, the process admits a stationary version. This result can be well understood when considering the branching interpretation of the INGARCH processes [55], in which $\sum_{i=1}^p a_i$ is the mean number of offspring generated by a living individual.
- If $\sum_{i=1}^p a_i > 1$, then \tilde{X}_n grows exponentially in n almost surely. For completeness we will provide a simple proof of this general result, postponed to the Appendix 4.A at the end of this paper.

A first result for the non-linear version

Our results will be stated in terms of the Markov chain defined by

$$X_n := (\tilde{X}_n, \tilde{X}_{n-1}, \dots, \tilde{X}_{n-p+1}) \in \mathbb{N}^p, \tag{4.2.1}$$

which can easily be interpreted as results concerning the long-time behaviour of the stochastic process $(\tilde{X}_n)_{n \geq 0}$:

Theorem 4.2.1. *If $(a_1)_+ + \dots + (a_p)_+ < 1$, then $(X_n)_{n \geq 0}$ is a geometrically ergodic Markov chain.*

We draw the reader’s attention to the fact that the condition in Theorem 4.2.1 is a discrete analogue of the sufficient condition for the existence of a continuous-time Hawkes process given in the article [13].

As announced in the previous section, this result naturally extends the known condition for all positive parameters. However, this sufficient condition is quite restrictive because it only consider the positive part of coefficients and therefore forget the inhibition. It is natural to inquire whether, in the case where one of the parameters is greater than 1, one can offset the other parameters sufficiently negatively to obtain stability. This question appears to be quite challenging to resolve, as the non-linearity comes into play in such instances. A generalization of the study of this process to an arbitrary number p of parameters seems beyond reach at the moment. Therefore, we confine ourselves in the following to the specific case of $p = 3$.

4.2.2 Three-parameter discrete-time Hawkes process

In the case where $p = 3$, we choose for sake of clarity to relabel the parameters as $a_1 = a$, $a_2 = b$ and $a_3 = c$. The discrete time process $(\tilde{X}_n)_{n \geq 0}$ is now defined

with initial condition $(\tilde{X}_0, \tilde{X}_{-1}, \tilde{X}_{-2})$ as $n \geq 1$:

$$\text{conditioned on } \tilde{X}_{-2}, \dots, \tilde{X}_{n-1}: \tilde{X}_n \sim \mathcal{P} \left(\left(a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + c\tilde{X}_{n-3} + \lambda \right)_+ \right). \quad (4.2.2)$$

Our objective is to complete the results of Theorem 4.2.1 by considering cases where the three parameters have different signs and one of them is larger than 1.

The naturally associated Markov chain $(X_n)_{n \geq 0} \in \mathbb{N}^3$ writes

$$\forall n \geq 0, X_n := \left(\tilde{X}_n, \tilde{X}_{n-1}, \tilde{X}_{n-2} \right) \in \mathbb{N}^3. \quad (4.2.3)$$

More formally, for an initial condition and parameter $a, b, c \in \mathbb{R}, \lambda > 0$, knowing $X_n = (i, j, k) \in \mathbb{N}^3$ the next step X_{n+1} of the Markov chain will be :

$$X_{n+1} = (\ell, i, j),$$

where ℓ is a realization of a Poisson random variable of parameter

$$s_{ijk} := (ai + bj + ck + \lambda)_+ \in [0, +\infty).$$

Note that by convention, if $s_{ijk} = 0$ the next step of the Markov chain is $(0, i, j)$. The transition matrix of the Markov chain $(X_n)_{n \geq 0}$ is thus given for $(i, j, k, \ell, p, q) \in \mathbb{N}^6$ by:

$$P((i, j, k), (\ell, p, q)) = \delta_{ip} \delta_{jq} \frac{e^{-s_{ijk}} s_{ijk}^\ell}{\ell!}, \text{ where } \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}. \quad (4.2.4)$$

4.2.2.1 Existing results in the case $c = 0$

When $c = 0$, the process $(\tilde{X}_n)_{n \geq 0}$ is a discrete Hawkes process with memory of length two, which is studied in [14]. The main result of that article is the following :

Theorem (Theorem 1 in [14]). *Assume $c = 0$. Define the function*

$$b_*(a) = \begin{cases} 1 & a \leq 0 \\ 1 - a & a \in (0, 2) \\ -\frac{a^2}{4} & a \geq 2 \end{cases}$$

and define the following sets :

$$\mathcal{R} = \{(a, b) \in \mathbb{R}^2 : b < b_*(a)\} \quad (4.2.5)$$

$$\mathcal{T} = \{(a, b) \in \mathbb{R}^2 : b > b_*(a)\}. \quad (4.2.6)$$

- If $(a, b) \in \mathcal{R}$, then the sequence $(\tilde{X}_n)_{n \geq 0}$ converges in law as $n \rightarrow \infty$.

- If $(a, b) \in \mathcal{T}$, then the sequence $(\tilde{X}_n)_{n \geq 0}$ satisfies that almost surely

$$\tilde{X}_n + \tilde{X}_{n+1} \xrightarrow[n \rightarrow \infty]{} +\infty.$$

Let us highlight that both parameters a and b have asymmetric roles : for any values of a , if b is chosen small enough, then the Markov chain $(X_n)_{n \geq 0}$ will be recurrent, while as soon as $b > 1$, it will be transient.

The rest of the article is devoted to the case where $c \neq 0$. We classified our results into two classes : recurrence results, and transience one. Since we did not managed to classify entirely the space of parameters $(a, b, c) \in \mathbb{R}^3$ we completed our theoretical results with numerical results.

Before stating our results, let us introduce important notions that will be necessary in the proofs.

4.2.2.2 Linear recurrence, polynomials and discriminant

Due to the concentration property of the Poisson distribution, as long as the parameter of the Poisson random variable in the definition of $(\tilde{X}_n)_{n \geq 0}$ remains positive, we expect that the process behaves similarly to the deterministic sequence $(\tilde{u}_n)_{n \geq 0}$, which is defined as follows:

$$\tilde{u}_n = a\tilde{u}_{n-1} + b\tilde{u}_{n-2} + c\tilde{u}_{n-3} + \lambda. \quad (4.2.7)$$

It is a well-known fact that the roots of the polynomial

$$P(X) := X^3 - aX^2 - bX - c, \quad (4.2.8)$$

play a significant role in the identification of sequences that satisfy the recurrence (4.2.7). More precisely, the stability of the linear recurrence $(\tilde{u}_n)_{n \geq 0}$ (in the sense that sequences satisfying recurrence (4.2.7) remain bounded for all $n \in \mathbb{N}$) is characterized by the condition :

$$\max\{|\zeta|, P(\zeta) = 0\} < 1. \quad (4.2.9)$$

The results we will state on the recurrence of the Markov chain $(X_n)_{n \geq 0}$ use another quantity associated with the polynomial P , namely its *discriminant*. Concerning the polynomial P we are interested in, its discriminant can be easily computed, yielding :

$$\text{Disc}(P) := a^2b^2 + 4b^3 - 4a^3c - 18abc - 27c^2 \quad (4.2.10)$$

The sign of $\text{Disc}(P)$ determines the number of roots of P that are complex. For a degree 3 polynomial, on the one hand, if $\text{Disc}(P) < 0$, the polynomial P has an unique simple real root, and two complex conjugate roots. On the other hand, if $\text{Disc}(P) \geq 0$, all the roots of P are real numbers, and they are simple when $\text{Disc}(P) > 0$. For more details about resultant and discriminant, we refer the reader to chapter 12 of [36].

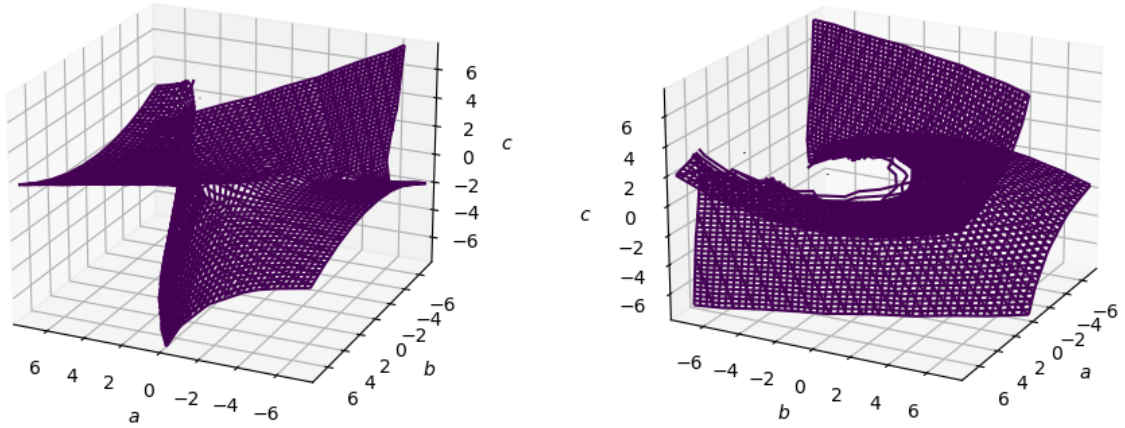


Figure 4.1 – Graphical representation of the surface of \mathbb{R}^3 delimited by $\text{Disc}(P) = 0$. The axis represents the different values of parameters $a, b, c \in \mathbb{R}$. Note that $\text{Disc}(P)$ is negative outside the two surfaces.

In Figure 4.1, we propose a graphical representation of the surface of \mathbb{R}^3 delimited by $\text{Disc}(P) = 0$.

Returning to our comparison involving deterministic sequences $(\tilde{u}_n)_{n \geq 0}$ satisfying (4.2.7), when $\text{Disc}(P) < 0$, since the polynomial P has 2 complex roots, we know that the solutions $(\tilde{u}_n)_{n \geq 0}$ will exhibit oscillatory behaviour, potentially diverging under stronger conditions. Consequently, $(\tilde{u}_n)_{n \geq 0}$ will eventually become negative. This is where the positive part in the definition of $(\tilde{X}_n)_{n \geq 0}$ comes into play : by truncating the parameter value of the Poisson random variable to 0, we can expect a stabilizing effect on the asymptotic behaviour of this process.

4.2.3 Transience results

We start by considering cases where b or c are larger than 1. We derive sufficient conditions which guarantee that the process $(\tilde{X}_n)_{n \geq 0}$ will diverge in the sense that almost surely,

$$\tilde{X}_n + \tilde{X}_{n+1} + \tilde{X}_{n+2} \xrightarrow[n \rightarrow +\infty]{} +\infty.$$

In terms of the Markov chain $(X_n)_{n \geq 0}$, this is the object of the following :

Theorem 4.2.2. *If $a, b, c \in \mathbb{R}$ satisfy one of the following condition*

1. $a, b < 0$ and $c > 1$, or
2. $b > 1$ and $ab + c < 0$

then the Markov chain $(X_n)_{n \geq 0}$ is transient.

Theorem 4.2.2 states that when b or c is greater than 1, and the other parameters are negative, $(X_n)_{n \geq 0}$ is transient. To rephrase, there is no inhibition, no matter how strong, that yields a stable process as long as b or c is greater than 1.

78 Chapter 4. Discrete-time Hawkes process with length three memory

The proof of Theorem 4.2.2 can be found in sections 4.4.1 for case *i*) and 4.4.2 for case *ii*). Let us remark that that second case, the condition $ab + c < 0$ is actually related to $\text{Disc}(P)$ being negative as it is precised in Lemma 4.B.1

4.2.4 Recurrence results

In this section, we prove that for $a > 1$, it is possible to choose, b and c sufficiently negative, for the process X_n to be ergodic.

Theorem 4.2.3. *Let $a > 1$. If $b < 0$ and $c < 0$ are such that $\text{Disc}(P) < 0$, then $(X_n)_{n \geq 0}$ is a geometrically ergodic Markov chain.*

The proof of Theorem 4.2.3 is deferred to section 4.3.3. As a remark, note that in Theorem 4.2.3, when $c = 0$ we have $\text{Disc}(P) = (a^2 + 4b)b^2$, making the condition $\text{Disc}(P) < 0$ equivalent to the condition $a^2 + 4b < 0$ mentioned in section 4.2.2.1.

Based on the expression of $\text{Disc}(P)$, we obtain the following technical lemma precising that for given $a > 0$ and $b \in \mathbb{R}$, if c is chosen negative enough, then $\text{Disc}(P) < 0$.

Lemma 4.2.4. *Consider $a > 0$ and $b \in \mathbb{R}$ then*

- if $a^2 + 3b < 0$, then $\text{Disc}(P) < 0$ for all values of $c \in \mathbb{R}$.
- if $a^2 + 3b \geq 0$ then $\text{Disc}(P) < 0 \Leftrightarrow c \in] -\infty, c_- [\cup] c_+, \infty [$ where

$$c_- = \frac{-1}{27}(2a^3 + 9ab + 2(a^2 + 3b)^{3/2}) \quad c_+ = \frac{-1}{27}(2a^3 + 9ab - 2(a^2 + 3b)^{3/2})$$

Note that furthermore if $a^2 + 4b \geq 0$, then $c_- < 0 < c_+$.

The proof of this Lemma is postponed in Appendix 4.B.

Theorem 4.2.3 states that for any given excitation $a > 1$, there exist sufficiently strong inhibitions $b, c < 0$ such that the parameters of the process $(\tilde{X}_n)_{n \geq 0}$ satisfy the assumptions of Theorem 4.2.3. This completes the results of Theorem 4.2.2 and implies that the order of parameters is important, introducing an asymmetry in their roles, which is a non trivial property of the process $(\tilde{X}_n)_{n \geq 0}$ in view of its initial definition.

Figure 4.2 shows for different values of a the region of parameters $(b, c) \in \mathbb{R}^2$ covered by the recurrence and transient results enumerated previously.

4.2.5 Conjecture and numerical illustrations

The proof of Theorem 4.2.3 is based on a Lyapunov argument following the theory developed by Meyn and Tweedie in the 80's. Under assumptions $a > 1, b < 0, c < 0$ such that $\text{Disc}(P) < 0$, we proved that there exists a Lyapunov function for the Markov chain (X_n) , and in a second time, we identified a *small* set (see section 4.3.1).

The construction of the Lyapunov function requires $c < 0$ while $b < 0$ is used to construct an appropriate small set. Based on the numerical experiments conducted,

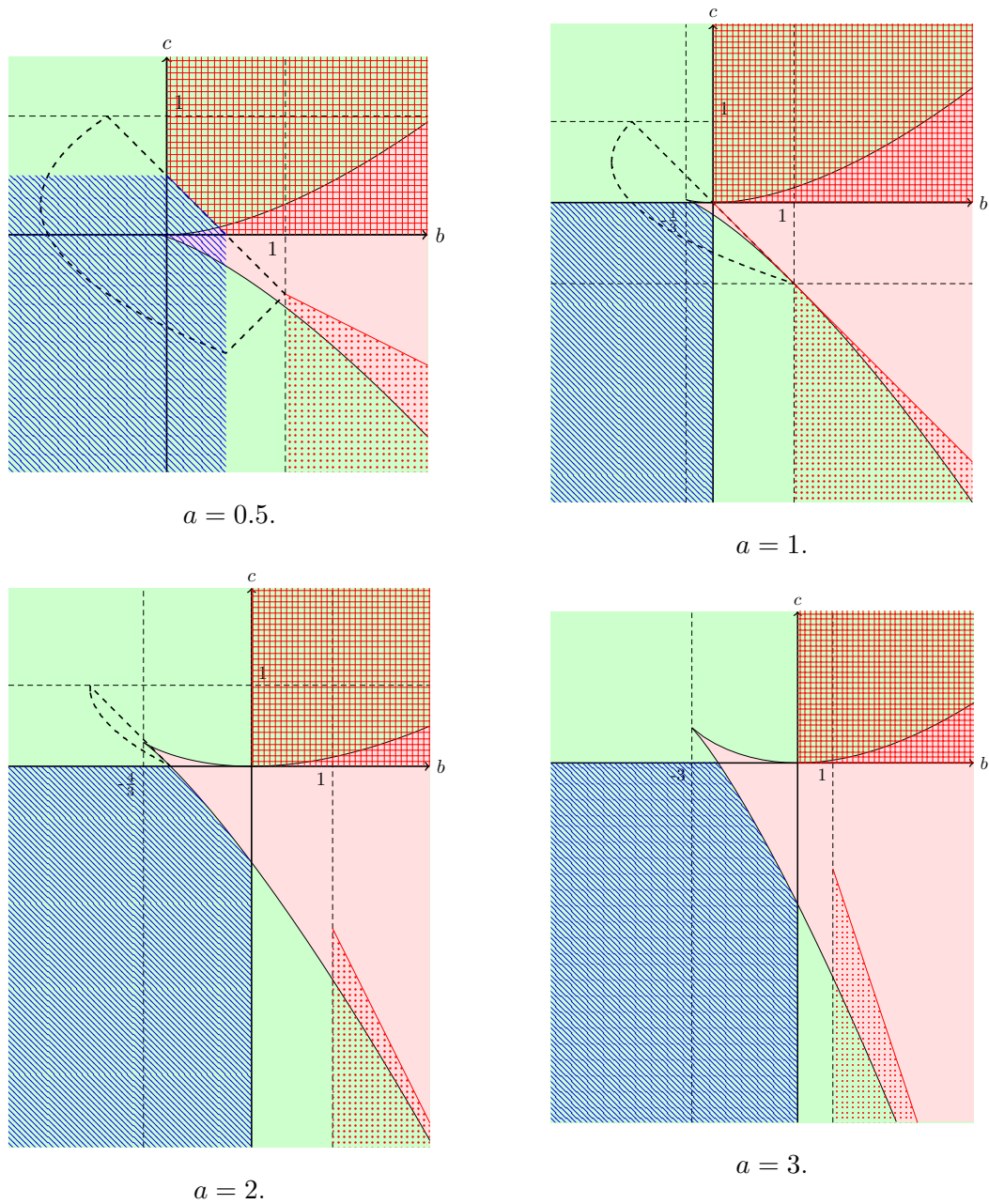


Figure 4.2 – Some representations of the areas $(a, b, c) \in \mathbb{R}^3$ covered by recurrence and transience results, for fixed $a \in \{0.5, 1, 2, 3\}$. In each figure, the green region corresponds to $\text{Disc}(P) < 0$, and the red region to $\text{Disc}(P) > 0$. The black dashed region represents the area where the linear recurrence (4.2.7) is stable (note that this never happens when $a = 3$). The blue lined region corresponds to the area covered by Theorem 4.2.3, and by Theorem 4.2.1 for the case $a = 0.5$. The red dotted region corresponds to Theorem 4.2.2. The red grid region corresponds to the known result about transience which is discussed in Section 4.2.1 and proven in Appendix 4.A, for the case $p = 3$.

80 Chapter 4. Discrete-time Hawkes process with length three memory

we believe that it is possible to slightly relax the assumptions of Theorem 4.2.3 as follows:

Conjecture 4.2.5. *Let $a > 1$. If $b \leq 1$ and $c < 0$ are such that $\text{Disc}(P) < 0$, then (X_n) is an ergodic Markov chain.*

Note that furthermore, if $a > 1, b > 1$ and $c < 0$ such that $\text{Disc}(P) < 0$, then using properties of the discriminant, it is possible to prove that $ab + c < 0$ (see Lemma 4.B.1 in Appendix 4.B. As a consequence we can apply Theorem 4.2.2 *ii*) and deduce that the Markov chain (X_n) is transient.

Let us now present our numerical simulations to handle the case $a > 1, b \in [0, 1)$ and $c < 0$ such that $\text{Disc}(P) = 0$. In all the following simulations we chose the initial condition $(X_{-2}, X_{-1}, X_0) = (0, 0, 0)$ and $\lambda = 1$. The state $(0, 0, 0)$ is of notable interest since it is an accessible state: for any values of a, b, c , it is always possible to reach this state from any other state in 3 steps. We illustrate the fact that the following stopping time :

$$\tau_0 := \inf\{n \geq 1 \mid X_n = (0, 0, 0)\},$$

satisfies

$$\begin{cases} \mathbb{P}_{(0,0,0)}(\tau_0 = +\infty) = 0 & \text{if } b \leq 1, \\ \mathbb{P}_{(0,0,0)}(\tau_0 = +\infty) > 0 & \text{if } b > 1. \end{cases}$$

To approach as closely as possible the simulation of τ_0 , we instead consider the following random variable, where n is chosen to be large (in the following illustration, $n = 10000$) :

$$\tilde{\tau}_0 := \tau_0 \wedge n.$$

However, due to numerical reasons, we cannot directly simulate this random variable. This is because, since we expect the Markov chain to be transient when $b > 1$, such a direct simulation would yield arbitrarily large values when simulating (X_n) , potentially leading to computational errors. To avoid this, we fix a very large threshold denoted by M and simulate :

$$\hat{\tau}_0 := \tilde{\tau}_0 \mathbb{1}_{(X_n) \text{ did not explode before reaching } (0,0,0)} + (n+1) \mathbb{1}_{(X_n) \text{ exploded}},$$

where we considered that the Markov chain (X_n) exploded if there exists $m \leq n$ such that $\tilde{X}_m > M$. In the case of an explosion, this amounts to considering $\tau_0 = +\infty$. We then set an arbitrary value, here $n+1$, for $\hat{\tau}_0$.

We consider a setting where b varies in $[0, 4]$ while maintaining a negative discriminant, in order to observe the phase transition at $b = 1$. We fixed $a = 3$ and $c = -15$: for this choice of parameters, for any $b \in [0, 4]$, $\text{Disc}(P) < 0$. For the sake of conciseness, unless stated otherwise, we will present our numerical study for these given fixed values; however, we have observed consistent behaviours for all tested values of a and c such that $\text{Disc}(P) < 0$ for all b considered.

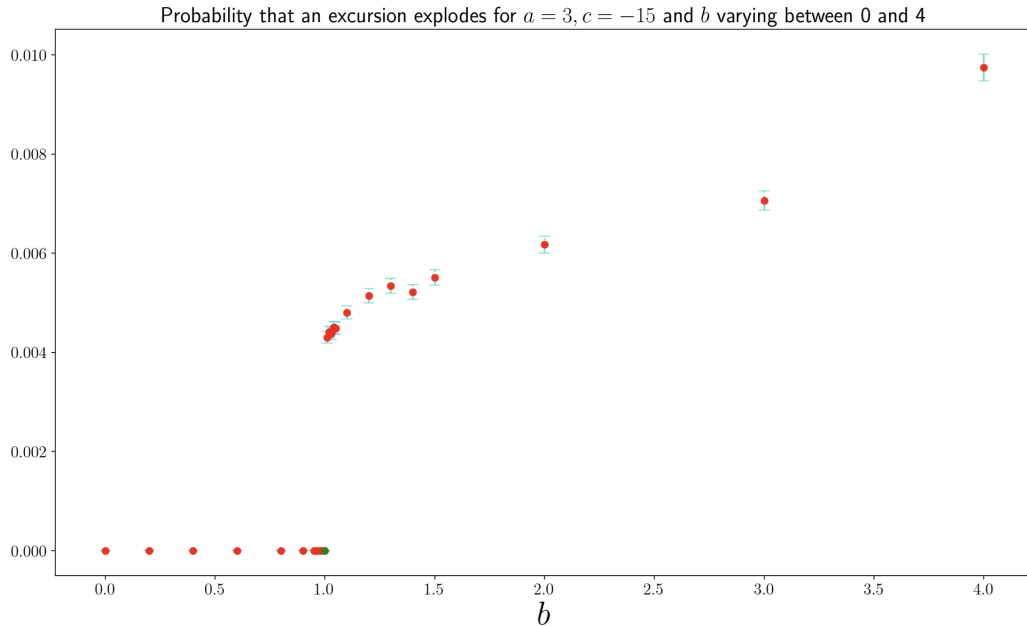


Figure 4.3 – Plot (in red) of the proportion of excursions that have exploded relative to the total number of excursions, for some $b \in [0, 4]$. For clarity, we put a green dot when $b = 1$. Clopper-Pearson type confidence intervals (with $\alpha = 0.01$) around the calculated values have also been plotted in blue. It is noteworthy that the positive proportions exhibit remarkably low values : even within the context where $b > 1$, the number of excursions that have exploded remains very small compared to the overall number of excursions.

We first estimate the probability that an excursion starting from $(0, 0, 0)$ explodes. In Figure 4.3 we draw the Monte Carlo estimator based on 10^6 repetitions.

We provide confidence intervals using Clopper-Pearson intervals which are *exact* confidence intervals for binomial proportion p , having a coverage at least equal to $1 - \alpha$ for all $p \in (0, 1)$. The formula of this interval is given by (see Section 2.1 of [78]) :

$$\left(B(\alpha/2, X, N - X + 1), \quad B(1 - \alpha/2, X + 1, N - X) \right),$$

where $B(\gamma, x, y)$ is the γ -quantile of the $Beta(x, y)$ distribution, and $0 \leq X \leq N$ is the number of success. In our case, for each b we have $N \approx 10^6$. It is also suitable in cases where no success is observed (i.e. $X = 0$) which is the situation here for all values of $b \leq 1$. In this case, closed-form expression for the interval bound is available, and given by :

$$\left(0, 1 - (\alpha/2)^{1/N} \right).$$

On Figure 4.3, we observe the transition occurring at $b = 1$: for $b \leq 1$, none of the simulated excursions has exploded, whereas there is a strictly positive number

of them for $b > 1$. Additionally, it is noteworthy that it seems to be a growing trend in the probability of exploding while b is growing.

We complete our numerical study by drawing the empirical distribution function of $\hat{\tau}_0$ in Figure 4.4.

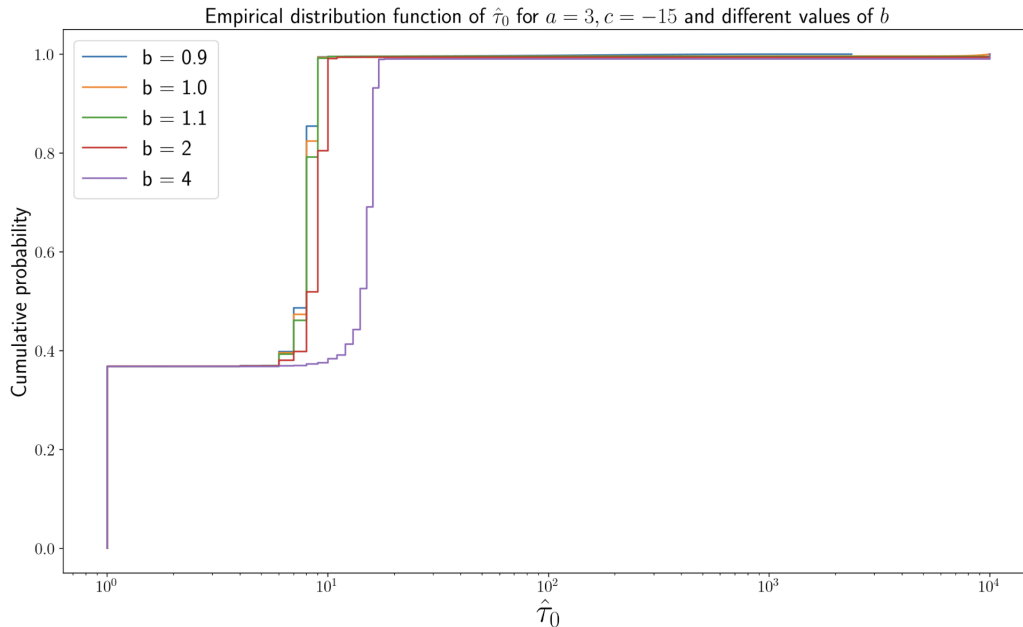


Figure 4.4 – Empirical distribution function of the random variable $\hat{\tau}_0$, for some $b \in [0.9, 4]$ (the $\hat{\tau}_0$ -axis is log scaled). For values of b strictly greater than 1, an atom appears at $n + 1$: this is a rephrasing of the phenomenon observed in Figure 4.3. In the case of $b = 1$, some excursions may attain significant lengths without having exploded.

This confirms that as the parameter b increases, the length of excursions $\hat{\tau}_0$ also increases. For example, for $b = 0.9$, the length of excursions never reaches 10^4 . This growth of $\hat{\tau}_0$ with b appears to become more pronounced as b increases. Nevertheless, we observe that the transition from $b \leq 1$ to $b > 1$ does not drastically change the length of the majority of excursions. However, it enables the observation of excursions of arbitrarily long lengths which happens, as discussed before, in a proportion significantly smaller compared to the total number of excursions considered.

We now put ourselves within a different context in an attempt to explain this very low proportion of excursions that have exploded, even with large values of b . We believe that this proportion is strongly dependent on the parameter a . Therefore, in Figure 4.5 we conducted a series of simulations where the parameter a varies between 0.1 and 8, while maintaining $\text{Disc}(P) < 0$ in order to be consistent with the conjectural framework. To achieve this, we set $b = 8$ and $c = -121$. The construction of Figure 4.5 closely follows the procedure outlined for Figure 4.3.

To conclude this section of numerical illustrations, we end with a final figure

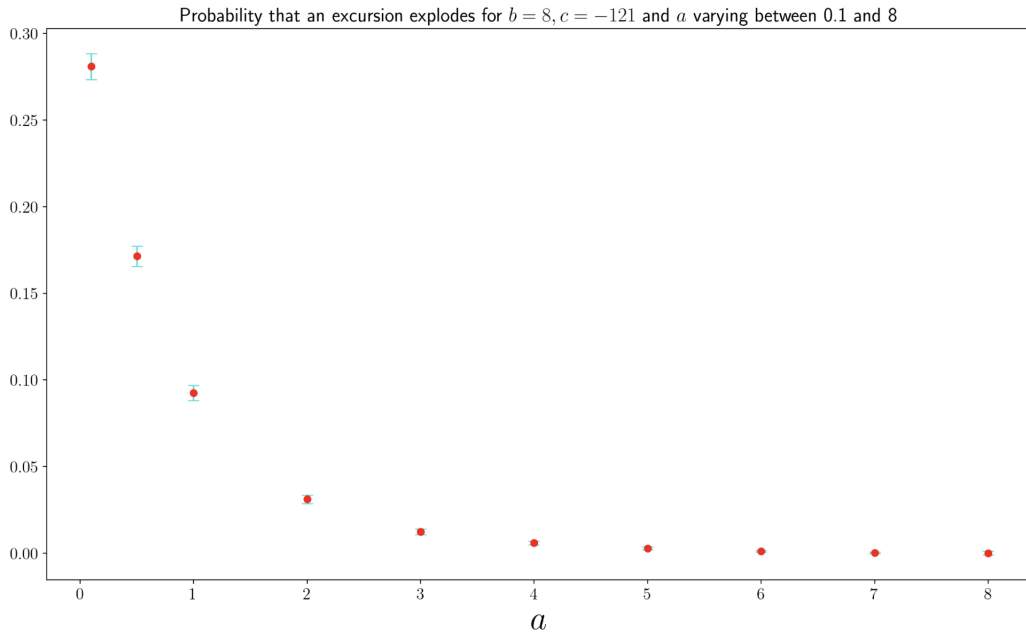


Figure 4.5 – Plotting the proportions (and confidence intervals) of excursions that have exploded relative to the total number of excursions. The graph is constructed similarly to Figure 4.3. A strong decrease to 0 is observed, which is counter-intuitive at first glance.

where we use parameters $a = 3, b = 1.1, c = -15$. We went a step further in analysing exploding excursions for this setting. It seemed that they all followed a similar pattern, alternating between 0 and positive numbers after a certain time. Since $b > 1$, the positive numbers tended to increase on average. In the next graph, we randomly picked some excursions of (\tilde{X}_n) that exploded with $b = 1.1$ as an illustration of our statement. Each color in the graph corresponds to one excursion, traced over the initial thirty values.

On Figure 4.6, we observe that after $n = 6$, each excursion behaves as described: at all even time steps, $\tilde{X}_n = 0$ while at all odd time steps, $\tilde{X}_n > 0$. We draw the reader's attention to the fact that this type of state $(0, x, 0)$ where $x > 0$, precisely forms the sets S_n in the proof of Theorem 4.2.2 for the case $b > 1, ab + c < 0$ (see section 4.4.2). This partially addresses the remark at the end of this section : in practice, we do observe the Markov chain (X_n) reaching these states, although, as seen in Figure 4.3, this rarely occurs.

4.3 Proofs of Recurrent cases

4.3.1 Lyapunov function and Drift criterion

Drift criteria, initially introduced by Foster [31], and extensively studied and popularized by Meyn and Tweedie [67], and others, constitute powerful analytical

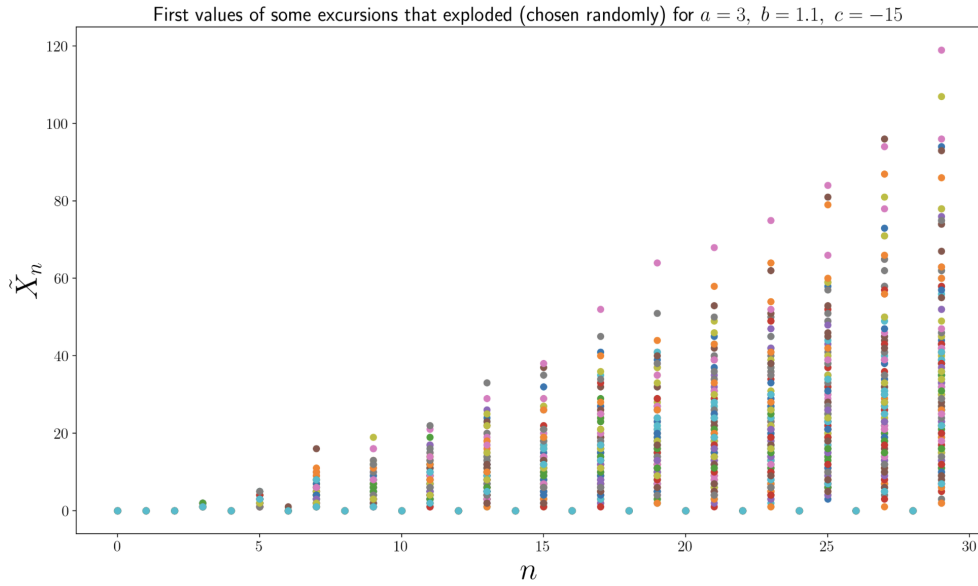


Figure 4.6 – Plotting the initial values of select trajectories that have exploded, in the specific case where $b = 1.1$. It is noteworthy that this phenomenon is observable for any value of $b > 1$.

tools. These criteria play a pivotal role in establishing convergence to the invariant measure of Markov chains and determining explicit rates of convergence. We adopt the approach outlined by Douc, Moulines, Priouret, and Soulier [19], which draws inspiration from Meyn and Tweedie’s work but is particularly suitable for Markov chains that are irreducible but not strongly irreducible. We begin with the following definitions :

Definition 4.3.1. A subset $C \subset \mathbb{N}^3$ is called **petite** [19, Definition 9.4.1], if there exists a state $x_0 \in \mathbb{N}^3$ and a probability distribution $(p_n)_{n \in \mathbb{N}}$ on \mathbb{N} such that

$$\inf_{x \in C} \sum_{n \in \mathbb{N}} p_n P^n(x, x_0) > 0,$$

where $P^n(x, x_0)$ represents the n -step transition probability from state x to x_0 .

It is noteworthy that, due to the irreducibility of the Markov chain $(X_n)_{n \geq 0}$, any finite set is petite (with x_0 being the accessible state) and the union of finite petite sets remains petite [19, Proposition 9.4.5].

Definition 4.3.2. A function $V : \mathbb{N}^3 \rightarrow [1, \infty)$ is a **Lyapunov function** if there exists a parameter $\varepsilon \in (0, 1]$, a finite constant $K < \infty$ and a set of states $C \subset \mathbb{N}^3$ for which V satisfies the **drift condition** $D(V, \varepsilon, K, C)$

$$\forall x \in \mathbb{N}^3, \quad \Delta V(x) := \mathbb{E}_x[V(X_1) - V(X_0)] \leq -\varepsilon V(x) + K \mathbf{1}_C(x). \quad (4.3.1)$$

It is easy to see that (4.3.1) implies condition $D_g(V, \lambda, b, C)$ from [19, Definition 14.1.5], with $\lambda = 1 - \varepsilon$ and $b = K$.

Note that we are in a setting where the Markov chain (X_n) is aperiodic and irreducible. Indeed, since the probability that a Poisson random variable is zero is strictly positive, it is possible to reach the state $(0, 0, 0)$ with positive probability from any state in three steps. In particular, the state $(0, 0, 0)$ is accessible and the Markov chain is irreducible. Furthermore, the Markov chain is aperiodic [19, Section 7.4], since $P((0, 0, 0), (0, 0, 0)) = e^{-\lambda} > 0$. For this setting, in [19, Theorem 9.4.10] it is proven that the notions of *petite* and *small* sets are equivalent. We will thus use the simpler definition of *small* sets, that is, a set $C \subset \mathbb{N}^3$ such that it exists $x_0 \in \mathbb{N}^3$ and $n \in \mathbb{N}$ satisfying :

$$\inf_{x \in C} P^n(x, x_0) > 0.$$

Proposition 4.3.3. *Assume that the drift condition $D(V, \varepsilon, K, C)$ is verified for some V, ε, K and C as above and assume that C is petite. Then there exists $\beta > 1$ and a probability measure π on \mathbb{N}^3 , such that, for every initial state $x \in \mathbb{N}^3$,*

$$\beta^n \times \sum_{y \in \mathbb{N}^3} V(y) |\mathbb{P}_x(X_n = y) - \pi(y)| \rightarrow 0, \quad n \rightarrow \infty.$$

In particular, for every initial state $x \in \mathbb{N}^3$,

$$\beta^n d_{TV}(\text{Law}(X_n), \pi) \rightarrow 0, \quad n \rightarrow \infty$$

and π is an invariant probability measure for the Markov chain $(X_n)_{n \geq 0}$.

4.3.2 Proof of Theorem 4.2.1

Recall that $p \in \mathbb{N}$ and that Markov chain (X_n) is defined by

$$X_n = (\tilde{X}_n, \tilde{X}_{n-1}, \dots, \tilde{X}_{n-p+1}),$$

where, knowing $\{\tilde{X}_n = x_1, \dots, \tilde{X}_{n-p+1} = x_p\}$, \tilde{X}_{n+1} is thus distributed as a Poisson random variable with parameter

$$\mathbf{s} := (a_1 x_1 + \dots + a_p x_p + \lambda)_+. \quad (4.3.2)$$

Let us denote by A the subset of \mathbb{N}^p defined by :

$$A = \{(x_1, \dots, x_p) \in \mathbb{N}^p \mid a_1 x_1 + \dots + a_p x_p + \lambda \leq 0\}.$$

We will first prove the following :

Lemma 4.3.4. *If $(a_1)_+ + \dots + (a_p)_+ < 1$, then there exists $\alpha_1, \dots, \alpha_p > 0$ such that the function $V : \mathbb{N}^p \rightarrow [1; +\infty)$ defined by*

$$V(x_1, \dots, x_p) := \alpha_1 x_1 + \dots + \alpha_p x_p + 1,$$

is a Lyapunov function.

86 Chapter 4. Discrete-time Hawkes process with length three memory

Proof. To begin with, note that the assumptions allow us to choose $\eta > 0$ such that

$$(a_1)_+ + \cdots + (a_p)_+ + \eta < 1.$$

Let $\varepsilon > 0$ to be properly chosen later. Then, for all $x = (x_1, \dots, x_p) \in \mathbb{N}^p$:

$$\begin{aligned} \Delta V(x) + \varepsilon V(x) &= \mathbb{E}_x[V(X_1) - V(X_0)] + \varepsilon V(x) & (4.3.3) \\ &= \sum_{\ell=0}^{+\infty} \frac{e^{-\mathbf{s}\ell}}{\ell!} V(\ell, x_1, \dots, x_{p-1}) + (\varepsilon - 1)V(x_1, \dots, x_p) \\ &= \alpha_1 \mathbf{s} + (\alpha_2 + \alpha_1 \varepsilon - \alpha_1)x_1 + (\alpha_3 + \alpha_2 \varepsilon - \alpha_2)x_2 \\ &+ \dots & (4.3.4) \\ &+ (\alpha_p + \alpha_{p-1} \varepsilon - \alpha_{p-1})x_{p-1} + (\varepsilon \alpha_p - \alpha_p)x_p + \varepsilon \end{aligned}$$

Recall that \mathbf{s} is either equal to 0, or equal to $a_1 x_1 + \cdots + a_p x_p + \lambda > 0$, which means that in any case $\Delta V + \varepsilon V$ is a multi-linear function of (x_1, \dots, x_p) . We will thus choose $\alpha_1, \dots, \alpha_p > 0$ and $\varepsilon > 0$ such that each coefficient in front of the x_i terms in (4.3.3) is negative, so there will be only a finite number of (x_1, \dots, x_p) that satisfies $\Delta V(x_1, \dots, x_p) + \varepsilon V(x_1, \dots, x_p) \geq 0$.

- Let us first consider $(x_1, \dots, x_p) \notin A$, that is, such that $\mathbf{s} = a_1 x_1 + \cdots + a_p x_p + \lambda > 0$. According to (4.3.3), we have :

$$\begin{aligned} \Delta V(x_1, \dots, x_p) + \varepsilon V(x_1, \dots, x_p) &= (\alpha_1 a_1 + \alpha_2 + \alpha_1 \varepsilon - \alpha_1)x_1 \\ &+ (\alpha_1 a_2 + \alpha_3 + \alpha_2 \varepsilon - \alpha_2)x_2 \\ &+ \dots & (4.3.5) \\ &+ (\alpha_1 a_{p-1} + \alpha_p + \alpha_{p-1} \varepsilon - \alpha_{p-1})x_{p-1} \\ &+ (\alpha_1 a_p + \varepsilon \alpha_p - \alpha_p)x_p \\ &+ \alpha_1 \lambda + \varepsilon \end{aligned}$$

First, let

$$\alpha_p := (a_p)_+ + \frac{\eta}{p} > 0,$$

and, for $i \in \{1, \dots, p-1\}$, let

$$\alpha_i := \alpha_{i+1} + (a_i)_+ + \frac{\eta}{p} > 0.$$

We thus have that

$$\alpha_1 = (a_1)_+ + \cdots + (a_p)_+ + \eta < 1.$$

This implies that :

$$\alpha_1 a_p - \alpha_p = \alpha_1 a_p - \left[(a_p)_+ + \frac{\eta}{p} \right] < a_p - (a_p)_+ - \frac{\eta}{p} < 0 \quad (4.3.6)$$

For the second-to-last line of (4.3.5), this implies that for a sufficiently small $\varepsilon > 0$,

$$\alpha_1 a_p + \varepsilon \alpha_p - \alpha_p < 0.$$

We can easily see that the same argument applies to all coefficients in front of the x_i terms in (4.3.5) if ε is chosen to be sufficiently small, making the coefficients all negative.

- Now, if $(x_1, \dots, x_p) \in A$, from (4.3.3) we have that :

$$\begin{aligned} \Delta V(x_1, \dots, x_p) + \varepsilon V(x_1, \dots, x_p) &= (\alpha_2 + \alpha_1 \varepsilon - \alpha_1)x_1 \\ &+ (\alpha_3 + \alpha_2 \varepsilon - \alpha_2)x_2 \\ &+ \dots \\ &+ (\alpha_p + \alpha_{p-1} \varepsilon - \alpha_{p-1})x_{p-1} \\ &+ (\varepsilon \alpha_p - \alpha_p)x_p \\ &+ \varepsilon \end{aligned} \tag{4.3.7}$$

According to the definition of $\alpha_1, \dots, \alpha_p$, it is straightforward that

$$\forall i \in \{1, \dots, p-1\}, \quad \alpha_{i+1} - \alpha_i < 0.$$

Here again, for ε sufficiently small, every coefficient preceding the x_i terms in (4.3.7) is negative.

In summary, we have shown that with a proper choice of parameters $\alpha_1, \dots, \alpha_p, \varepsilon$, there is only a finite number of states (x_1, \dots, x_p) for which $\Delta V(x_1, \dots, x_p) + \varepsilon V(x_1, \dots, x_p) > 0$. We will denote the set of such states as $C \subset \mathbb{N}^p$. This proves that V is a Lyapunov function for the Markov chain (X_n) . □

Since C is a finite subset of \mathbb{N}^p , it is a *small* set. Note that for every $x \in C$, we have

$$\Delta V(x) + \varepsilon V(x) \leq \max_{x \in C} \mathbb{E}_x[V(X_1) - V(X_0)] =: K < +\infty,$$

where K is finite since the set C is finite. This ensures that drift condition $D(V, \varepsilon, K, C)$ is satisfied.

We can thus apply Proposition 4.3.3 adapted for a Markov chain that takes values in \mathbb{N}^p , which concludes the proof.

4.3.3 Proof of Theorem 4.2.3

To begin with we will provide a brief review of fundamental definitions within the theory of quadratic forms.

A quadratic form is an application q from \mathbb{R}^n to \mathbb{R} satisfying :

- $\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, q(\lambda x) = \lambda^2 q(x)$
- $B : (x, y) \mapsto \frac{1}{2} [q(x+y) - q(x) - q(y)]$ is a symmetric bilinear application.

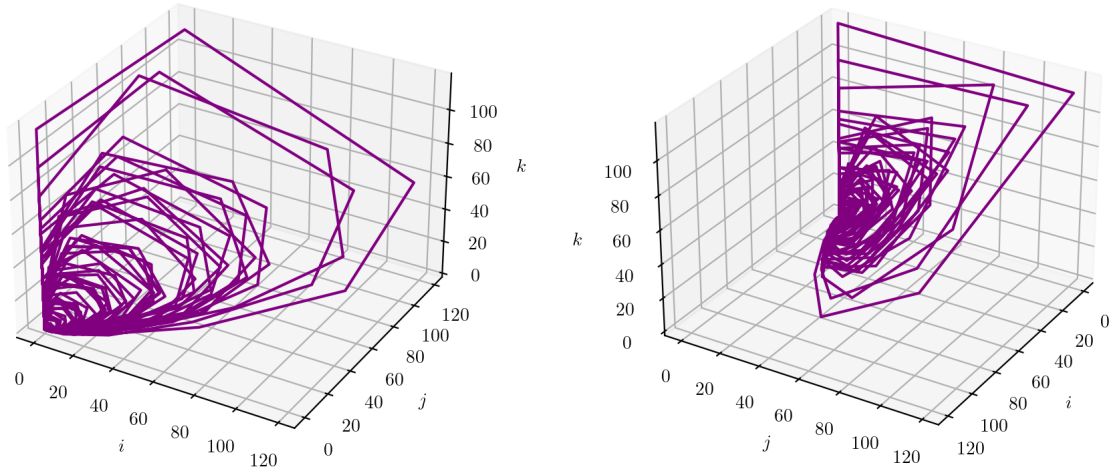


Figure 4.7 – Typical trajectory of (X_n) , for parameters satisfying assumptions of Theorem 4.2.3. Here, the parameters are $a = 2.5, b = -1, c = -3$. For these parameters, we have $\text{Disc}(P) = -188.25$.

The application B described above is called the *polar form* of q (note that for all $x \in \mathbb{R}^n, B(x, x) = q(x)$). We define the *kernel* of q , which is :

$$\text{Ker } q := \{x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, B(x, y) = 0\}.$$

One set of interest is the isotropic cone of q , defined as :

$$\text{Co}(q) := \{x \in \mathbb{R}^n, q(x) = 0\} \supset \text{Ker } q.$$

Note that computing the isotropic cone is difficult in general, especially when q is a signed application. However, if q is either positive or negative, $\text{Co}(q)$ can be easily computed by solving a system of linear equations. For more details about quadratic forms, we refer to chapter 4 of [69].

In order to use Proposition 4.3.3 we have to identify a small set, and prove we can build an appropriate Lyapunov function. We first begin to show :

Lemma 4.3.5. *Let us define the following set of states :*

$$A := \{(i, j, k) \in \mathbb{N}^3, ai + bj + ck + \lambda \leq 0\}.$$

Assuming that $b, c < 0$, then A is a small set.

Proof. By definition of A , we have $s_{ijk} = 0$ for all $(i, j, k) \in A$, hence $P((i, j, k), (0, i, j)) = 1$. Furthermore, for every $i \in \mathbb{N}$, since $b, c < 0$, we have

$$P((0, i, j), (0, 0, i)) = e^{-s_{0ij}} = e^{-(\lambda+bi+cj)_+} \geq e^{-\lambda}.$$

And since $c < 0$,

$$P((0, 0, i), (0, 0, 0)) = e^{-s_{00i}} = e^{-(\lambda+ci)_+} \geq e^{-\lambda}.$$

It follows that

$$\inf_{(i,j,k) \in A} P^3((i, j, k), (0, 0, 0)) \geq e^{-2\lambda} > 0,$$

which shows that A is *small*. \square

For $\alpha \geq 0$, define $V_\alpha : \mathbb{N}^3 \rightarrow [1; +\infty)$ as follows:

$$V_\alpha(i, j, k) := \frac{i + \alpha j}{j + \alpha k + 1} + 1. \quad (4.3.8)$$

We will prove the following :

Lemma 4.3.6. *Under the hypotheses of Theorem 4.2.3, there exists $\alpha > 0$, $\varepsilon > 0$, $K < \infty$ and a finite set $B \subset \mathbb{N}^3$, such that the function V_α defined by (4.3.8) satisfies the drift condition $D(V_\alpha, \varepsilon, K, C)$, with $C = A \cup B$.*

Proof. Let $(i, j, k) \notin A$, so that $s_{ijk} = ai + bj + ck + \lambda > 0$. We compute :

$$\begin{aligned} \Delta V_\alpha(i, j, k) &= \sum_{\ell=0}^{+\infty} \frac{s_{ijk}^\ell e^{-s_{ijk}}}{\ell!} V_\alpha(\ell, i, j) - V_\alpha(i, j, k) \\ &= \sum_{\ell=0}^{+\infty} \frac{s_{ijk}^\ell e^{-s_{ijk}}}{\ell!} \frac{\ell + \alpha i}{i + \alpha j + 1} - \frac{i + \alpha j}{j + \alpha k + 1} \\ &= \frac{(s_{ijk} + \alpha i)(j + \alpha k + 1) - (i + \alpha j)(i + \alpha j + 1)}{(i + \alpha j + 1)(j + \alpha k + 1)} \\ &= \frac{q(i, j, k) + L(i, j, k)}{(i + \alpha j + 1)(j + \alpha k + 1)}, \end{aligned}$$

where

$$q(i, j, k) = -i^2 + (b - \alpha^2)j^2 + c\alpha k^2 + (a - \alpha)ij + \alpha(a + \alpha)ik + (c + b\alpha)jk,$$

and L is a linear form.

Note that if we consider (i, j, k) to be elements of \mathbb{R}^3 , then q is a real quadratic form in the sense of the definitions given in the first section of this paper. We will thus use the theoretical framework as if we worked on \mathbb{R}^3 , but note that our space of interest, which is \mathbb{N}^3 , is far smaller.

Step 1: Choice of an appropriate α

Note that q is a quadratic form of degree 3, having the following representation matrix in the canonical base :

$$M_\alpha := \begin{pmatrix} -1 & \frac{a - \alpha}{2} & \frac{\alpha(a + \alpha)}{2} \\ \frac{a - \alpha}{2} & b - \alpha^2 & \frac{c + b\alpha}{2} \\ \frac{\alpha(a + \alpha)}{2} & \frac{c + b\alpha}{2} & c\alpha \end{pmatrix}.$$

90 Chapter 4. Discrete-time Hawkes process with length three memory

Denote by $R(\alpha)$ the first minor of order 2 of M_α , that is,

$$R(\alpha) = \begin{vmatrix} -1 & \frac{a-\alpha}{2} \\ \frac{a-\alpha}{2} & b-\alpha^2 \end{vmatrix} = -a^2 - 4b + 2a\alpha + 3\alpha^2.$$

The Gauss reduction algorithm for q yields :

$$q(i, j, k) = - \left(i + \frac{\alpha-a}{2}j - \frac{\alpha(a+\alpha)}{2}k \right)^2 - R(\alpha) \left(j - \frac{K(\alpha)}{2R(\alpha)}k \right)^2 + \frac{\det M_\alpha}{R(\alpha)}k^2, \quad (4.3.9)$$

where

$$K(\alpha) := c + \alpha b - \frac{\alpha(\alpha+a)(\alpha-a)}{2}. \quad (4.3.10)$$

A simple calculation gives :

$$\det M_\alpha = \frac{1}{4}(\alpha^3 + a\alpha^2 - b\alpha + c)^2 = \frac{1}{4}Q(\alpha)^2 \geq 0,$$

where the polynomial $Q(X)$ is defined by

$$Q(X) := X^3 + aX^2 - bX + c = -P(-X).$$

Recall that we assumed that $\text{Disc}(P) < 0$, so P has a unique real root, hence Q as well. Denote by α_Q this root. In particular, this gives that $\det M_{\alpha_Q} = 0$. Since the dominant coefficient of Q is positive, and $Q(0) = c < 0$, we have that $\alpha_Q > 0$.

Let us now prove that with the choice of $\alpha = \alpha_Q$ the quadratic form q is negative, which is equivalent to $R(\alpha_Q) > 0$ thanks to the decomposition (4.3.9).

Lemma 4.3.7. *Under the assumptions of Theorem 4.2.3, if α_Q is the unique real root of Q , then $R(\alpha_Q) > 0$.*

The very computational proof of this lemma is postponed to Appendix 4.B.

Step 2 : Cone of the quadratic form.

Let us now prove that $\text{Co}(q) \cap (\mathbb{R}_+)^3 = \{(0, 0, 0)\}$. Thanks to (4.3.9), we can easily compute :

$$\text{Co}(q) = \text{Vect} \left\{ \left(\frac{\alpha_Q(a+\alpha_Q)}{2} + \frac{K(\alpha_Q)(a-\alpha)}{4R(\alpha_Q)}, \frac{K(\alpha_Q)}{2R(\alpha_Q)}, 1 \right) \right\}.$$

To prove that $\text{Co}(q) \cap (\mathbb{R}_+)^3 = \{(0, 0, 0)\}$, it's sufficient to prove that two of the directional vector coordinates of $\text{Co}(q)$ have opposite signs. We will thus show that $\frac{K(\alpha_Q)}{2R(\alpha_Q)} < 0$. Since α_Q satisfies $R(\alpha_Q) > 0$, we have

$$3\alpha_Q^2 + 2a\alpha_Q - a^2 - 4b > 0 \iff a^2 + 2b < 3\alpha_Q^2 + 2a\alpha_Q - 2b,$$

we deduce that

$$\begin{aligned} K(\alpha_Q) &= -\alpha_Q^3 + (a^2 + 2b)\alpha_Q + 2c < -\alpha_Q^3 + (3\alpha_Q^2 + 2a\alpha_Q - 2b)\alpha_Q + 2c \\ &< 2(\alpha_Q^3 + a\alpha_Q^2 - b\alpha_Q + c) \\ &< 2Q(\alpha_Q) = 0. \end{aligned}$$

This yields $\frac{K(\alpha_Q)}{2R(\alpha_Q)} < 0$ and thus $\text{Co}(q) \cap (\mathbb{R}_+)^3 = \{(0, 0, 0)\}$.

Step 3 : Continuity argument Let us now prove that there is only a finite number of $(i, j, k) \in \mathbb{N}^3 \setminus A$ such that $\Delta V_{\alpha_Q}(i, j, k) + \varepsilon V_{\alpha_Q}(i, j, k) > 0$.

We have

$$\Delta V_{\alpha_Q}(x, y, z) + \varepsilon V_{\alpha_Q}(x, y, z) = \frac{q(x, y, z) + \varepsilon \tilde{q}(x, y, z) + L(x, y, z) + \varepsilon \tilde{L}(x, y, z)}{(x + \alpha_Q y + 1)(y + \alpha_Q z + 1)},$$

where

$$\tilde{q}(x, y, z) = x^2 + \alpha_Q(\alpha_Q + 1)y^2 + (2\alpha_Q + 1)xy + \alpha_Q xz + \alpha_Q^2 yz$$

and \tilde{L} is a linear functional in (x, y, z) .

Let us consider $(x, y, z) \in (\mathbb{R}_+)^3$ with $\|(x, y, z)\| = 1$, from Step 2 we deduce that there exists a positive constant C such

$$q(x, y, z) \leq -C$$

We deduce that for a second constant \tilde{C}

$$q^\varepsilon(x, y, z) = q(x, y, z) + \varepsilon \tilde{q}(x, y, z) \leq -C + \varepsilon \tilde{C}.$$

Therefore, we can choose ε small enough such that $q^\varepsilon(x, y, z) < 0$ for all $(x, y, z) \in (\mathbb{R}_+)^3$. In order to conclude, note that for any $(x, y, z) \in (\mathbb{R}_+)^3$,

$$\Delta V_{\alpha_Q}(x, y, z) + \varepsilon V_{\alpha_Q}(x, y, z) \leq q^\varepsilon(x, y, z) + \frac{L(x, y, z) + \varepsilon \tilde{L}(x, y, z)}{(x + \alpha_Q y + 1)(y + \alpha_Q z + 1)}$$

where we used that $\alpha_Q > 0$. Using the fact that the second term of the r.h.s. is bounded in \mathbb{R}^3 , this proves that for a sufficiently small $\varepsilon \in (0, 1)$, there is only a finite number of (i, j, k) satisfying $(i, j, k) \in \mathbb{N}^3 \setminus A$ such that $\Delta V(i, j, k) + \varepsilon V(i, j, k) > 0$. Let us note B this set.

Now for every $(i, j, k) \in A$, we have

$$\begin{aligned} \Delta V(i, j, k) + \varepsilon V(i, j, k) &\leq \mathbb{E}_{(0, i, j)}[V(X_1)] \\ &= V(0, i, j) \\ &= \frac{\alpha i}{i + \alpha j + 1} + 1 \\ &\leq 2 \vee (\alpha + 1) =: \tilde{K} < +\infty \end{aligned}$$

We can now conclude that the drift condition $D(V, \varepsilon, K, A \cup B)$ is satisfied. \square

92 Chapter 4. Discrete-time Hawkes process with length three memory

By Lemma 4.3.5, the set A is a *small* set. Since C is a finite set, we have that $A \cup C$ is a *small* set too. This concludes the proof of Theorem 4.2.3.

We would like to point out to the reader that in the case where $b \in (0, 1)$ and $\text{Disc}(P) < 0$, the function V defined by (4.3.8) remains a Lyapunov function. However, in this situations, the set A is no longer small. For a deeper discussion of this case, please refer to Section 4.2.5.

4.4 Proofs of Transience cases

In this section, we prove Proposition 4.2.2. To accomplish this, we will systematically employ a methodology analogous to our previous approach (see section 4 in [14]), using the following lemma adapted for the present three-parameter process :

Lemma 4.4.1. *Let S_1, S_2, \dots be a sequence of subsets of \mathbb{N}^3 and $0 < m_1 < m_2 < \dots$ an increasing sequence of integers. Suppose that*

1. *On the event $\bigcap_{n \geq 1} \{X_{m_n} \in S_n\}$, we have $X_n \neq (0, 0, 0)$ for all $n \geq 1$,*
2. *$\mathbb{P}_{(0,0,0)}(X_{m_1} \in S_1) > 0$ and for all $n \geq 1$ and every $x \in S_n$, we have $\mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) > 0$.*
3. *There exist $(p_n)_{n \geq 1}$ taking values in $[0, 1]$ and such that $\sum_{n \geq 1} (1 - p_n) < \infty$, such that*

$$\forall n \geq 1 : \forall x \in S_n : \mathbb{P}_x(X_{m_{n+1}-m_n} \in S_{n+1}) \geq p_n.$$

Then the Markov chain $(X_n)_{n \geq 0}$ is transient.

Proof. Since $(0, 0, 0)$ is an accessible state, it is enough to show that

$$\mathbb{P}_{(0,0,0)}(X_n \neq (0, 0, 0), \forall n \geq 1) > 0.$$

Using assumption 1, it is sufficient to prove that

$$\mathbb{P}_{(0,0,0)}(X_{m_n} \in S_n, \forall n \geq 1) > 0. \quad (4.4.1)$$

By assumption 3, there exists $n_0 \geq 1$ such that $\prod_{n \geq n_0} p_n > 0$. It follows that for every $x \in S_{n_0}$,

$$\mathbb{P}_x(X_{m_n-m_{n_0}} \in S_n, \forall n > n_0) \geq \prod_{n \geq n_0} p_n > 0.$$

Furthermore, by assumption 2, we have that

$$\mathbb{P}_{(0,0,0)}(\forall n \leq n_0, X_{m_n} \in S_n) > 0.$$

Combining the last two inequalities yields (4.4.1) and finishes the proof. \square

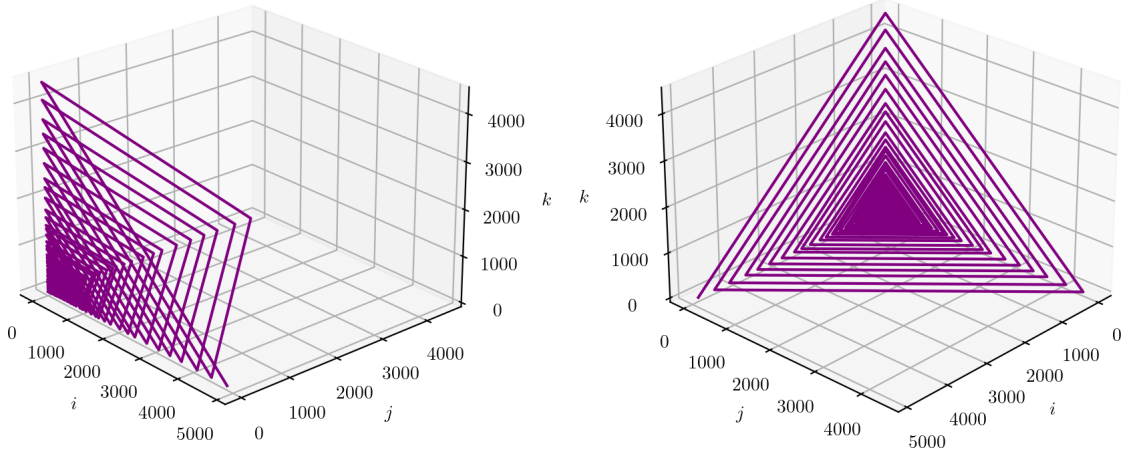


Figure 4.8 – Typical trajectory for the case $a, b < 0, c > 1$. Here, the parameters are $a = -1, b = -1, c = 1.1$.

4.4.1 Case $a, b < 0, c > 1$

In this region of parameters, the Markov chain eventually reaches one of the axes (see Figure 4.8).

Indeed, given that both a and b are negative, when the Markov chain (X_n) reaches a state $(i, 0, 0)$ with $i \geq \max\{-\frac{\lambda}{a}, -\frac{\lambda}{b}\}$, the fact that $s_{i00} = (ai + \lambda)_+ = 0$ and $s_{0i0} = (bi + \lambda)_+ = 0$, implies that the subsequent two steps of the Markov chain will lead to states $(0, i, 0)$ and $(0, 0, i)$. Afterwards, the Markov chain will hit the state $(\mathcal{P}(ci + \lambda), 0, 0)$, where $ci + \lambda > i$.

Consequently, focusing on the i axis as an illustration, starting from $(k, 0, 0)$ with a sufficiently large k , the Markov chain (X_n) is highly likely to return to a state $(k', 0, 0)$ where $k' > k$, along the i axis, within three steps.

To formalize these observations, it is natural to consider the Markov chain generated by the third power of the transition matrix, denoted as P^3 , namely $(X_{3n+1})_{n \geq 0}$.

For $i \geq \max\{-\frac{\lambda}{a}, -\frac{\lambda}{b}\}$, we have $s_{i00} = s_{0i0} = 0$ and thus :

$$\begin{aligned} \mathbb{P}\left(X_{3n+3} = (k, 0, 0) \mid X_{3n} = (i, 0, 0), i \geq \max\left\{-\frac{\lambda}{a}, -\frac{\lambda}{b}\right\}\right) &= \frac{e^{-s_{00i}} s_{00i}^k}{k!} \\ &= \frac{e^{-(ci+\lambda)} (ci + \lambda)^k}{k!}. \end{aligned} \tag{4.4.2}$$

Equation (4.4.2) means that if $\tilde{X}_{3n} \geq \max\{-\frac{\lambda}{a}, -\frac{\lambda}{b}\}$, and $\tilde{X}_{3n-1} = \tilde{X}_{3n-2} = 0$, then $\tilde{X}_{3n+1} = \tilde{X}_{3n+2} = 0$, and \tilde{X}_{3n+3} is a Poisson random variable with parameter $c\tilde{X}_{3n} + \lambda$.

We will now prove our statement.

Proof of the transience of (X_n) when $a, b < 0$ and $c > 1$. Fix $r \in (1, c)$. We wish to apply lemma 4.4.1 with

$$m_n = 3n - 1, \quad n \geq 1$$

and

$$S_n = \{(i, 0, 0) \in \mathbb{N}^3 : i \geq r^n\}.$$

Assumption (1) holds immediately since if $X_{3n-1} = (i, 0, 0) \in S_n$, then $X_{3n} = (j, i, 0)$ and $X_{3n+1} = (k, j, i)$ for some $j, k \in \mathbb{N}$, hence $(X_{3n-1}, X_{3n}, X_{3n+1}) \neq (0, 0, 0)$.

We now verify that the second assumption holds. For states $x, y \in \mathbb{N}^3$, write $x \rightarrow_1 y$ if $\mathbb{P}_x(X_1 = y) > 0$. Furthermore, for $S \subset \mathbb{N}^3$, write $x \rightarrow_1 S$ if $x \rightarrow_1 y$ for some $y \in S$. Note that $(0, 0, 0) \rightarrow_1 (i, 0, 0)$ for every $i \in \mathbb{N}$, so that $(0, 0, 0) \rightarrow_1 S_1$. Now, for every $i \in \mathbb{N}$, we have $(i, 0, 0) \rightarrow_1 (0, i, 0)$, then, because $b < 0$, $(0, i, 0) \rightarrow_1 (0, 0, i)$ and then since $c > 0$, $(0, 0, i) \rightarrow_1 (j, 0, 0)$ for every $j \in \mathbb{N}$. In particular, from every $x \in S_n$, we can indeed reach S_{n+1} in three steps. Hence, the second assumption is verified as well.

We now prove the third assumption. We claim that there exists $n_0 \in \mathbb{N}$, such that the following holds:

$$\forall n \geq n_0, \forall x \in S_n : \mathbb{P}_x(X_3 \in S_{n+1}) \geq 1 - \frac{c}{(c-r)^2 r^n}. \quad (4.4.3)$$

To prove (4.4.3), first note that according to the earlier remark on (4.4.2), if n_0 is chosen such that $r^{n_0} \geq \max\{-\frac{\lambda}{a}, -\frac{\lambda}{b}\}$, then starting from a state $(x_0, 0, 0)$ with $x_0 \geq r^{n_0}$, we have $\tilde{X}_1 = \tilde{X}_2 = 0$ almost surely and $\tilde{X}_3 \sim \mathcal{P}(c\tilde{X}_0 + \lambda)$. Therefore, if $n \geq n_0$ and $x_0 \geq r^n \geq r^{n_0}$,

$$\begin{aligned} 1 - \mathbb{P}_{(x_0, 0, 0)}(\tilde{X}_3 \geq r^{n+1}, \tilde{X}_1 = \tilde{X}_2 = 0) &= \mathbb{P}_{(x_0, 0, 0)}(\tilde{X}_3 < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}(cx_0 + \lambda) < r^{n+1}) \\ &\leq \mathbb{P}(\mathcal{P}(cr^n) < r^{n+1}) \\ &= \mathbb{P}(\mathcal{P}(cr^n) - cr^n < r^n(r-c)) \\ &= \mathbb{P}(|\mathcal{P}(cr^n) - cr^n| > r^n(c-r)) \\ &\leq \frac{c}{(c-r)^2 r^n}, \end{aligned}$$

by the Bienaymé-Chebychev inequality. This proves (4.4.3), which implies,

$$\forall x \in S_n : \mathbb{P}_x(X_3 \in S_{n+1}) \geq p_n := \left(1 - \frac{c}{(c-r)^2 r^n}\right)_+,$$

and

$$\sum_{n \geq 1} (1 - p_n) \leq \sum_{n \geq 1} \frac{c}{(c-r)^2 r^n} < \infty.$$

This proves that the third assumption of Lemma 4.4.1 holds which shows that the Markov chain is transient. \square

4.4.2 Case $b > 1, ab + c < 0$

Proof of the transience of (X_n) when $b > 1$ and $ab + c < 0$. We will prove the result using exactly the same method as in the previous case, that is, we will employ

Lemma 4, with modifications to the definitions of m_n and S_n to accommodate the current case of interest. Before going into the proof details, let us provide some intuition regarding the behaviour of (X_n) in this case.

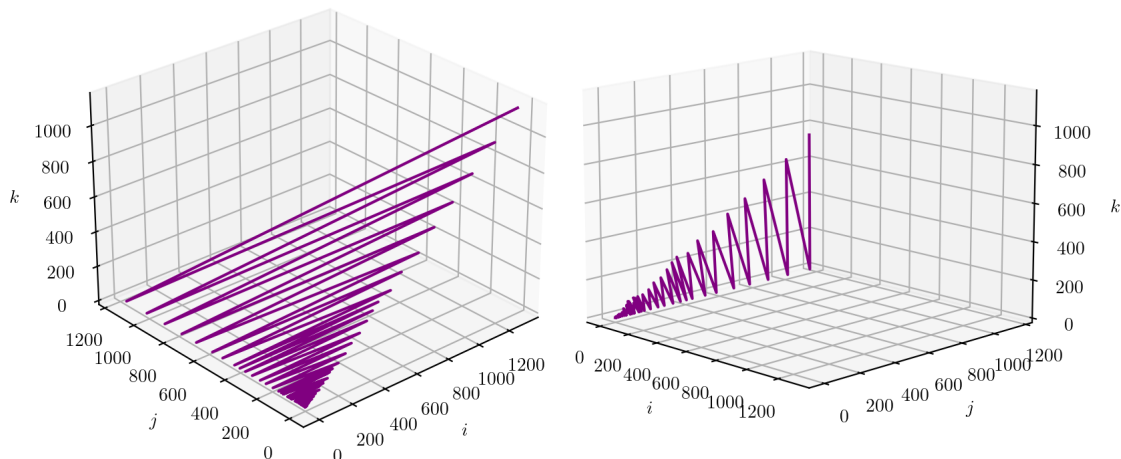


Figure 4.9 – Typical trajectory of (X_n) for the second case of Proposition 4.2.2. Here, $a = -1, b = 1.1, c = 0.5$.

Let's assume we start from a state $X_0 = (0, x, 0)$ where $x > 0$. We would then expect, on average, the next step of (X_n) to be $X_1 = (bx, 0, x)$. Continuing the reasoning : the next would, on average, be $X_2 = ((ab + c)x, bx, 0)$. Now since $ab + c < 0$, we end up with $X_2 = (0, bx, 0)$ (recall that we take the positive part of the Poisson random variable in the definition of (X_n)). To conclude, as $b > 1$, on average the Markov chain moves in two steps from a state $(0, x, 0)$ to a state $(0, y, 0)$ with $y > x$. See Figure 4.9.

Let us now formalize this proof outline. For this purpose, let $r \in (1, b)$ such that

$$ab + c + b - r < 0,$$

and let us denote by ε the quantity :

$$\varepsilon := b - r > 0.$$

We also define :

$$\beta := \frac{-a - a\lambda}{ab + c + \varepsilon}.$$

As announced, we will use Lemma 4.4.1 with

$$m_n := 2n - 1, \quad n \geq 1$$

and

$$S_n := \{(0, i, 0) \in \mathbb{N}^3 : i \geq r^n\}.$$

For the same reasons as in the previous case, Assumption (1) holds : indeed, if $X_{2n-1} = (0, i, 0) \in S_n$, then $X_{2n} = (j, 0, i)$ for some j , hence $X_{2n-1}, X_{2n} \neq (0, 0, 0)$.

96 Chapter 4. Discrete-time Hawkes process with length three memory

For the second assumption of lemma 4.4.1, note that, for every $i, j \in \mathbb{N}$, since $b > 0$, we have $(0, i, 0) \rightarrow_1 (j, 0, i)$. Then, because the probability for a Poisson random variable to be zero is positive, we have $(j, 0, i) \rightarrow_1 (0, j, 0)$ for every $j \in \mathbb{N}$. Particularly, for every $x \in S_n$, we can indeed reach S_{n+1} in two steps. Therefore, the second assumption is also satisfied.

We now establish the third assumption, asserting that there exists $n_0 \in \mathbb{N}$ such that the following holds :

$$\forall n \geq n_0, \forall x \in S_n : \mathbb{P}_x(X_2 \in S_{n+1}) \geq 1 - \frac{b + \lambda/r^n}{(b-r)^2 r^n}. \quad (4.4.4)$$

To establish (4.4.4), fix n_0 such that for all $n \geq n_0, r^n \geq \beta$. To begin with, observe that starting from $(0, i, 0) \in S_n$, we have $\tilde{X}_1 \sim \mathcal{P}(bi + \lambda)$. Subsequently, according to Bienaymé-Chebychev inequality :

$$\mathbb{P}(|\mathcal{P}(bi + \lambda) - (bi + \lambda)| > (b-r)i) \leq \frac{bi + \lambda}{(b-r)^2 i^2} \leq \frac{b + \lambda/r^n}{(b-r)^2 r^n}.$$

This means that with high probability,

$$\tilde{X}_1 \in [bi + \lambda \pm (b-r)i].$$

Furthermore, knowing \tilde{X}_1 we have :

$$\tilde{X}_2 \sim \mathcal{P}\left((a\tilde{X}_1 + ci + \lambda)_+\right).$$

However, since $i \geq \beta$, with high probability :

$$a\tilde{X}_1 + ci + \lambda \leq a(bi + \lambda + (b-r)i) + ci + \lambda = (ab + c + \varepsilon)i + a\lambda + a \leq 0,$$

and :

$$\tilde{X}_1 \geq bi + \lambda - (b-r)i = ri + \lambda > ri \geq r^{n+1}.$$

We deduce that

$$1 - \mathbb{P}_{(0,i,0)}(\tilde{X}_1 \geq r^{n+1}, \tilde{X}_2 = 0) \leq \frac{b + \lambda/r^n}{(b-r)^2 r^n}.$$

Now, (4.4.4) implies,

$$\forall x \in S_n : \mathbb{P}_x(X_3 \in S_{n+1}) \geq p_n := \left(1 - \frac{b + \lambda/r^n}{(b-r)^2 r^n}\right)_+,$$

and

$$\sum_{n \geq 1} (1 - p_n) \leq \sum_{n \geq 1} \frac{b + \lambda/r^n}{(b-r)^2 r^n} < \infty.$$

This establishes that the third assumption of Lemma 4.4.1 is satisfied. Consequently, the lemma implies that the Markov chain is transient, which concludes the proof. \square

A remark regarding this demonstration is necessary. The employed method aims to exhibit a sequence of decreasing state sets in which the Markov chain, if it reaches them, will become trapped, pulling it away from the origin. This demonstrates the transience of the Markov chain. However, a natural question arises: in practice, does the chain reach such states? This question is of interest; indeed, during the numerical investigation preceding the writing of this article, it was challenging to ascertain that the Markov chain was indeed transient when the parameters a , b , and c satisfied the assumptions of Theorem 4.2.2. This was due to the fact that, for the sake of clarity, we consistently used the same initial conditions for the process \tilde{X}_n in our simulations. In these simulations, we did not observe any process explosion phenomena, indicating that, within the scope of our numerical study, the process did not reach such states. Of course, by modifying the initial conditions to force the chain to start in these state spaces, we numerically observe the transience phenomenon described by Theorem 4.2.2.

This raises a question for the continuation of the study of this type of process. When there is no prior intuition about the parameter domain for which the Markov chain is transient, and numerical tests are conducted with arbitrary parameters, the potential dependence of the long-term behavior of the Markov chain on the chosen initial condition makes it challenging to carry out a rigorous numerical exploration. This difficulty might become more pronounced as the number of considered parameters increases.

4.A Proof of transience when $a_1, \dots, a_p \geq 0$ and $a_1 + \dots + a_p > 1$.

Proposition 4.A.1. *Let $p \geq 2$ and $a_1, \dots, a_p \geq 0$ such that $a_1 + \dots + a_p > 1$. Recall that for $n \geq 0$,*

$$\tilde{X}_{n+1} \sim \mathcal{P}(a_1 \tilde{X}_n + a_2 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p+1}).$$

Then, the Markov chain $X_n := (\tilde{X}_n, \dots, \tilde{X}_{n-p+1})$ is transient.

Proof. Let us define the polynomial

$$P(X) = X^p - a_1 X^{p-1} - \dots - a_{p-1} X - a_p.$$

Since $\lim_{x \rightarrow +\infty} P(x) = +\infty$, and $P(1) = 1 - a_1 - \dots - a_p < 0$, it exists $\theta > 1$ such that $P(\theta) = 0$ we can find $1 < r < \theta$ such that :

$$r^p - a_1 r^{p-1} - \dots - a_{p-1} r - a_p < 0. \tag{4.A.1}$$

We will use Lemma 4.4.1 to prove this result. For $n \geq p$, let us define $m_n = n$ and the sets

$$S_n = \{(i_1, \dots, i_p) \in \mathbb{N}^p \mid i_1 \geq r^n, i_2 \geq r^{n-1}, \dots, i_p \geq r^{n-p+1}\}.$$

98 Chapter 4. Discrete-time Hawkes process with length three memory

Note that since $a_1, \dots, a_p \geq 0$, assumption 1 and 2 are automatically satisfied.

For assumption 3, let $n \geq p$ and $\mathcal{I} := (i_1, \dots, i_p) \in S_n$. Then,

$$\begin{aligned} \mathbb{P}_{\mathcal{I}}(\tilde{X}_1 < r^{n+1}) &= \mathbb{P}\left(\mathcal{P}(a_1 i_1 + \dots + a_p i_p + \lambda) < r^{n+1}\right) \\ &\leq \mathbb{P}\left(\mathcal{P}(a_1 r^n + \dots + a_p r^{n-p+1}) < r^{n+1}\right) \\ &= \mathbb{P}\left[\mathcal{P}(a_1 r^n + \dots + a_p r^{n-p+1}) - (a_1 r^n + \dots + a_p r^{n-p+1})\right. \\ &\quad \left.< r^{n-p+1}(r^p - a_1 r^{p-1} - \dots - a_{p-1} r - a_p)\right]. \end{aligned}$$

From (4.A.1) and Bienaymé-Chebychev inequality, we deduce that :

$$\begin{aligned} \mathbb{P}_{\mathcal{I}}(\tilde{X}_1 < r^{n+1}) &\leq \mathbb{P}\left[\left|\mathcal{P}(a_1 r^n + \dots + a_p r^{n-p+1}) - (a_1 r^n + \dots + a_p r^{n-p+1})\right|\right. \\ &\quad \left.> r^{n-p+1}|r^p - a_1 r^{p-1} - \dots - a_{p-1} r - a_p|\right] \\ &\leq \frac{(a_1 + \dots + a_p)r^{2(p-1)}}{r^n(r^p - a_1 r^{p-1} - \dots - a_{p-1} r - a_p)^2} \end{aligned}$$

Hence, assumption 3 of lemma 4.4.1 is satisfied by considering a sequence (p_n) defined as :

$$p_n := \left(1 - \frac{(a_1 + \dots + a_p)r^{2(p-1)}}{r^n(r^p - a_1 r^{p-1} - \dots - a_{p-1} r - a_p)^2}\right)_+.$$

This concludes the proof of Proposition 4.A.1. \square

4.B Proofs of technical lemmas

4.B.1 Proof of Lemma 4.2.4

Coming back to the definition of $\text{Disc}(P)$ in (4.2.10), $\text{Disc}(P) < 0$ writes

$$a^2 b^2 + 4b^3 - 4a^3 c - 18abc - 27c^2 < 0$$

Let us look at the l.h.s. as a polynomial S of degree 2 in c

$$S(c) = -27c^2 - c(4a^3 + 18ab) + a^2 b^2 + 4b^3$$

Note that its discriminant

$$\begin{aligned} D_S &= (4a^3 + 18ab)^2 + 4 \times 27b^2(a^2 + 4b) \\ &= 16(a^2 + 3b)^3 \end{aligned}$$

If $D_S < 0$, since the dominant coefficient of S is negative, S is negative for all c .

Otherwise, S admits two real roots which are exactly c_{\pm} given by

$$c_{\pm} = \frac{1}{27}(-2a^3 - 9ab \pm 2(a^2 + 3b)^{3/2}).$$

The last condition arises simply by considering that if $S(0) = b^2(a^2 + 4b) > 0$, then the two eigenvalues have different signs.

4.B.2 Proof of Lemma 4.3.7

Recall that α_Q is the unique real root of Q and that since $Q(0) = c < 0$, α_Q is positive.

We now prove that $R(\alpha_Q) > 0$. Note that R is a polynomial in α , of discriminant $16(a^2 + 3b)$.

If $a^2 + 3b < 0$, then R is positive for all values of α , in particular $R(\alpha_Q) > 0$.

If $a^2 + 3b \geq 0$, the roots of R are

$$\alpha_{\pm} = -\frac{a}{3} \pm \frac{2}{3}\sqrt{a^2 + 3b}.$$

Note that α_- is always negative while $\alpha_+ \in [-a/3, a/3]$.

- In the case where $\alpha_+ < 0$, then $R(\alpha) > 0$ for all positive α , and as a consequence $R(\alpha_Q) > 0$.
- Let us now assume $\alpha_+ > 0$. In particular, this implies that $R(0) < 0$ which writes $a^2 + 4b > 0$.

Computing carefully, we can derive that

$$\alpha_+^2 = \frac{-a^2}{3} + \frac{4}{3}(a^2 + 3b) - \frac{4a}{9}\sqrt{a^2 + 3b},$$

$$\alpha_+^3 = \frac{8}{27}(a^2 + 3b)^{3/2} + \frac{2a^2}{9}\sqrt{a^2 + 3b} - \frac{a^3}{27} - \frac{4}{9}(a^2 + 3b),$$

and finally

$$\begin{aligned} Q(\alpha_+) &= \alpha_+^3 + a\alpha_+^2 - b\alpha_+ + c \\ &= \frac{1}{27} \left(2a^3 + 9ab + 27c + 2(a^2 + 3b)^{\frac{3}{2}} \right). \end{aligned}$$

In the following, we prove that $Q(\alpha_+) < 0$. Indeed since Q has a unique real root α_Q , we will deduce that $\alpha_Q > \alpha_+$, and as a consequence $R(\alpha_Q) > 0$.

In order to prove that $Q(\alpha_+) < 0$, it is sufficient to prove that under our assumptions on a , b and c

$$c < \frac{-1}{27}(2a^3 + 9ab + 2(a^2 + 3b)^{\frac{3}{2}}) \quad (4.B.1)$$

This derives directly from Lemma 4.2.4 since we assumed $\text{Disc}(P) < 0$, $a^2 + 4b \geq 0$ and $c < 0$,

4.B.3 Proof of Lemma 4.B.1

Lemma 4.B.1. *Let $b \in \mathbb{R}$, $a > 0$ and $c < 0$ such that $\text{Disc}(P) < 0$. Then, $ab + c < 0$.*

Let $b \in \mathbb{R}$, $a > 0$ and $c < 0$ such that $\text{Disc}(P) < 0$. We will prove that $ab + c < 0$.

Proof of Lemma 4.B.1.

- If $b < 0$, since $c < 0$ and $a > 0$ the result is obvious.
- If $b > 0$, then from Lemma 4.2.4, $\text{Disc}(P) < 0$ is equivalent to :

$$c < \frac{1}{27}(-2a^3 - 9ab) - \frac{2}{27}(a^2 + 3b)^{\frac{3}{2}}$$

which yields :

$$ab + c < \frac{1}{27}(-2a^3 + 18ab) - \frac{2}{27}(a^2 + 3b)^{\frac{3}{2}} = \frac{2}{27} \left[-a^3 + 9ab - (a^2 + 3b)^{\frac{3}{2}} \right]$$

It remains to prove that the r.h.s. is negative which can be written as

$$\begin{aligned} -a^3 + 9ab < (a^2 + 3b)^{\frac{3}{2}} &\iff (9ab - a^3)^2 < (a^2 + 3b)^3 \\ &\iff -27a^4b + 54a^2b^2 - 27b^3 < 0 \\ &\iff -27(a^2 - b)^2b < 0 \end{aligned}$$

the last inequality being obviously true when $b > 0$.

□

4.C Perspectives

We provide a proof of a partial result on the transience of the Markov chain.

Theorem 4.C.1. *Let θ be the largest real root of the polynomial $P(X) = X^3 - aX^2 - bX - c$. If $\theta > 1$, and $\theta^3 - |b|\theta - 2|c| > 0$, then the Markov chain (X_n) is transient.*

Based on the numerical study conducted on the process (\tilde{X}_n) , it appears to us that this result seems to hold whenever $\theta > 1$. The additional condition $\theta^3 - |b|\theta - 2|c| > 0$ in the statement of Theorem 4.C.1 serves as a technical requirement to successfully carry out the proof of this result.

Recall that we denoted by θ the largest real root of $P(X) = X^3 - aX^2 - bX - c$, and we supposed that $\theta > 1$ and

$$\theta^3 - |b|\theta - 2|c| > 0. \tag{4.C.1}$$

Firstly, although it is possible to numerically determine the parameter space for (a, b, c) such that $\theta > 1$ using software like Mathematica, this set appears to be difficult to grasp. This is why it is all the more challenging to determine whether the technical condition (4.C.1) is restrictive or not. However, we are certain that (4.C.1) is not automatically satisfied when $\theta > 1$ (example: for $a = 2, b = -2, c = 2$, we have $\theta \approx 1.54$ and $\theta^3 - |b|\theta - 2|c| \approx -3.4$).

Proof. Let $1 < r < \theta$ and $\varepsilon > 0$. To prove Theorem 4.C.1, we will apply Lemma 4.4.1 with $m_n = n$ and the following sets :

$$S_n := \left\{ (i, j, k) \in \mathbb{N}^3 \mid i \geq r^n, j \geq r^{n-1}, k \geq r^{n-2}, \left| \frac{i}{j} - \theta \right| < \varepsilon, \left| \frac{j}{k} - \theta \right| < \varepsilon \right\}.$$

Note that is possible to choose $\varepsilon > 0$ small enough such that

$$r^3 - a(r - \varepsilon)^2 - b(r - \varepsilon) - c < 0. \quad (4.C.2)$$

This means that if $(i, j, k) \in S_n$, then

$$ai + bj + ck + \lambda > [a(\theta - \varepsilon)^2 + b(\theta - \varepsilon) + c]k > [a(r - \varepsilon)^2 + b(r - \varepsilon) + c]k > 0.$$

So that Assumption 1 and 2 of Lemma 4.4.1 are satisfied.

For the verification of Assumption 3, we compute for $n \in \mathbb{N}$ and $(i, j, k) \in S_n$:

$$\mathbb{P}_{(i,j,k)}(X_1 \notin S_{n+1}) \leq \mathbb{P}_{(i,j,k)}(\tilde{X}_1 < r^{n+1}) + \mathbb{P}_{(i,j,k)}\left(\left|\frac{\tilde{X}_1}{i} - \theta\right| > \varepsilon\right) \quad (4.C.3)$$

We begin by bounding the first term on the right-hand side in (4.C.3) using (4.C.2) and Bienaymé-Chebychev inequality:

$$\begin{aligned} & \mathbb{P}_{(i,j,k)}(\tilde{X}_1 < r^{n+1}) \\ & \leq \mathbb{P}\left(\left|\mathcal{P}\left([a(r - \varepsilon)^2 + b(r - \varepsilon) + c]r^{n-2}\right) - [a(r - \varepsilon)^2 + b(r - \varepsilon) + c]r^{n-2}\right| > \right. \\ & \quad \left. - r^{n-2}[r^3 - a(r - \varepsilon)^2 - b(r - \varepsilon) - c]\right) \\ & \leq \frac{[a(r - \varepsilon)^2 + b(r - \varepsilon) + c]r^{n-2}}{\left[r^{n-2}[r^3 - a(r - \varepsilon)^2 - b(r - \varepsilon) - c]\right]^2} \\ & = \frac{[a(r - \varepsilon)^2 + b(r - \varepsilon) + c]r^2}{r^n [r^3 - a(r - \varepsilon)^2 - b(r - \varepsilon) - c]^2} =: \frac{C_1}{r^n}, \end{aligned} \quad (4.C.4)$$

where C_1 is a constant that does not depend on n .

For the second term on the right-hand side of (4.C.3), first note that :

$$\left|\frac{\tilde{X}_1}{i} - \theta\right| = \left|\frac{\tilde{X}_1 - \mathbb{E}_{(i,j)}[\tilde{X}_1]}{i}\right| + \left|\frac{\mathbb{E}_{(i,j)}[\tilde{X}_1]}{i} - \theta\right|$$

And, for any $(i, j, k) \in S_n$:

$$\begin{aligned} \left|\frac{\mathbb{E}_{(i,j,k)}[\tilde{X}_1]}{i} - \theta\right| & = \left|\frac{ai + bj + ck + \lambda}{i} - \theta\right| \\ & = \left|a + b\frac{j}{i} + c\frac{k}{i} + \frac{\lambda}{i} - \theta\right| \\ & \leq \underbrace{\left|a + \frac{b}{\theta} + \frac{c}{\theta^2} - \theta\right|}_{=0} + |b|\left|\frac{j}{i} - \frac{1}{\theta}\right| + |c|\left|\frac{k}{i} - \frac{1}{\theta^2}\right| + \frac{\lambda}{i} \\ & < \frac{|b|}{\theta(\theta - \varepsilon)}\varepsilon + |c|\left|\frac{k}{i} - \frac{1}{\theta^2}\right| + \frac{\lambda}{i}, \end{aligned}$$

Using that

$$|x - \theta| < \varepsilon, |y - \theta| < \varepsilon \implies \left| \frac{1}{xy} - \frac{1}{\theta^2} \right| < \frac{\varepsilon(\varepsilon + 2\theta)}{\theta^2(\theta - \varepsilon)^2},$$

we have :

$$\left| \frac{\mathbb{E}_{(i,j,k)}[\tilde{X}_1]}{i} - \theta \right| < \frac{|b|}{\theta(\theta - \varepsilon)}\varepsilon + |c| \frac{2\theta + \varepsilon}{\theta^2(\theta - \varepsilon)^2}\varepsilon + \frac{\lambda}{i}. \quad (4.C.5)$$

Using (4.C.1), it is possible to choose $\varepsilon > 0$ small enough such that :

$$\theta^2(\theta - \varepsilon)^2 - \theta(\theta - \varepsilon)|b| - (2\theta + \varepsilon)|c| > 0.$$

This implies that

$$\frac{|b|}{\theta(\theta - \varepsilon)} + |c| \frac{2\theta + \varepsilon}{\theta^2(\theta - \varepsilon)^2} < 1, \quad (4.C.6)$$

To prove that $\mathbb{P}_{(i,j,k)} \left(\left| \frac{\tilde{X}_1}{i} - \theta \right| > \varepsilon \right) \leq \frac{C_2}{r^n}$, where C_2 is a constant independent of n , we deduce from (4.C.5) that it suffices to show:

$$\mathbb{P}_{(i,j,k)} \left(\frac{\left| \tilde{X}_1 - \mathbb{E}_{(i,j,k)}[\tilde{X}_1] \right| + \lambda}{i} > \delta\varepsilon \right) \leq \frac{C_2}{r^n}, \quad (4.C.7)$$

where, by (4.C.6), we have

$$\delta := 1 - \frac{|b|}{\theta(\theta - \varepsilon)} - |c| \frac{2\theta + \varepsilon}{\theta^2(\theta - \varepsilon)^2} > 0.$$

Furthermore, since $(i, j, k) \in S_n$, we have that, for sufficiently small $\varepsilon > 0$:

$$k < k(\theta - \varepsilon) < j < j(\theta - \varepsilon) < i,$$

so that:

$$\begin{aligned} ai + bj + ck + \lambda &\leq |a|i + |b|j + |c|k + \lambda \\ &\leq (|a| + |b| + |c| + \lambda)i =: Ki. \end{aligned}$$

Using the Bienaymé-Chebyshev inequality, we obtain:

$$\begin{aligned} \mathbb{P}_{(i,j,k)} \left(\frac{\left| \tilde{X}_1 - \mathbb{E}_{(i,j,k)}[\tilde{X}_1] \right| + \lambda}{i} > \delta\varepsilon \right) &= \mathbb{P}_{(i,j,k)} \left(\left| \tilde{X}_1 - \mathbb{E}_{(i,j,k)}[\tilde{X}_1] \right| > \delta\varepsilon i - \lambda \right) \\ &\leq \frac{Ki}{(\delta\varepsilon i - \lambda)^2} \\ &= \frac{K}{(\delta\varepsilon\sqrt{i} - \lambda/\sqrt{i})^2} \\ &\leq \frac{K}{(\delta\varepsilon r^{n/2} - \lambda r^{-n/2})^2}. \end{aligned}$$

The last inequality holds for sufficiently large n . Indeed, since $i \geq r^n$, we have:

$$\delta\varepsilon\sqrt{i} - \lambda/\sqrt{i} > \delta\varepsilon r^{n/2} - \lambda r^{-n/2},$$

and for large enough n , we have:

$$\delta\varepsilon\sqrt{i} - \frac{\lambda}{\sqrt{i}} > 0.$$

This shows that (4.C.7) holds for some constant $C_2 < \infty$. Combining (4.C.4) and (4.C.7), we have:

$$\mathbb{P}_{(i,j,k)}(X_1 \in S_{n+1}) \geq \left(1 - \frac{C_1 + C_2}{r^n}\right)_+ =: p_n.$$

This will ultimately lead us to the result by applying Lemma 4.4.1 as before.

□

Critical discrete-time Hawkes processes

This chapter presents ongoing work on the critical cases of the process (\tilde{X}_n) .

Contents

5.1 Introduction.	105
5.2 The two-parameter case	107
5.2.1 A discrete martingale	107
5.2.2 Framework and theoretical results	109
5.2.3 Time and space rescaling	111
5.3 Perspectives	122

5.1 Introduction.

This work focuses on discrete-time Hawkes processes, namely auto-regressive Poisson processes defined as follows : for $p \in \mathbb{N}^*$ and initial condition (X_0, \dots, X_{-p+1}) , we defined discrete-time Hawkes process with memory of length p as :

$$\forall n \geq 1, \quad X_n \sim \mathcal{P}((a_1 X_{n-1} + \dots + a_p X_{n-p} + \lambda)_+), \quad (5.1.1)$$

where $\mathcal{P}(\rho)$ denotes the Poisson distribution with parameter ρ . The parameters a_1, \dots, a_p are real numbers, and $\lambda > 0$. The notation $(\cdot)_+$ refer to the ReLU function defined on \mathbb{R} by :

$$(x)_+ := \max(0, x).$$

In two previous chapters, we investigated the stability of these processes depending on the parameters of the model. More precisely, stability criteria for the case $p = 2$, some for $p = 3$, and $p \in \mathbb{N}$ were found and proved in Chapters 3 and 4, notably using Foster criteria through Lyapunov functions. In these chapters, the study of critical cases of the model appears to be particularly challenging, necessitating different probabilistic tools and therefore constituting a separate project. This work specifically addresses a critical case where the sum of parameters a_1, \dots, a_p equals 1, and these parameters are positive. We will demonstrate that addressing this case already requires a considerable technical background. We will thoroughly detail the case for $p = 2$ before providing directions for extending the results to higher values of p .

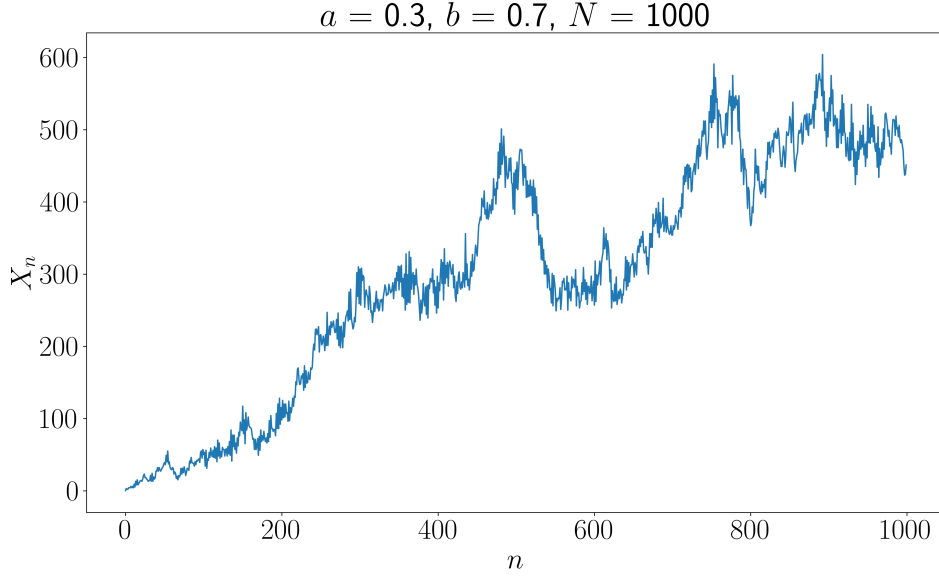


Figure 5.1 – A trajectory of the process $\{X_n\}_{n=0}^N$ in the case $p = 2$, and $a, b \geq 0$ such that $a + b = 1$. Here, $\lambda = 1$, and $(X_0, X_{-1}) = (0, 0)$. For the sake of readability, we have connected successive points of the process (X_n) .

In Figure 5.1, we present an example trajectory of X_n for the case $p = 2$, where we denote $a_1 := a, a_2 := b$. We observe erratic behavior at high frequency, which gives us the intuition that by accelerating time and after an appropriate spatial rescaling, the process (X_n) will approximately become a diffusion process on the positive real line \mathbb{R}_+ . This approach is also motivated by similar works regarding the linear continuous-time Hawkes process. Indeed, Jaisson and Rosenbaum have demonstrated the weak convergence of a nearly critical rescaled Hawkes process towards a Feller diffusion [47] or a rough fractional diffusion [48] depending on tail conditions of considered kernels. In a recent article, Horst and Xu [45], among other results, analyse the asymptotic behaviour of a rescaled critical Hawkes process in detail. More precisely, we briefly outline one of their results concerning a critical continuous linear Hawkes process denoted as $\{N(t), t \geq 0\}$ with intensity $\{\Lambda(t), t \geq 0\}$ of the form

$$\Lambda(t) := \mu_0 + \int_{(0,t)} \phi(t-s)N(ds),$$

where $\mu_0 \geq 0$ and $\phi \in L^1(\mathbb{R}_+; \mathbb{R}_+)$. With the notations

$$\sigma := \lim_{t \rightarrow +\infty} \int_0^t s\phi(s)ds, \quad \text{and } \forall n \in \mathbb{N}^*, \Lambda^{(n)}(t) := \frac{\Lambda(nt)}{n}, \quad (5.1.2)$$

the authors proved the following theorem :

Theorem 5.1.1 (2.10 in [45]). *If $\|\phi\|_1 = 1$ and $\sigma < \infty$, the sequence $(\Lambda^{(n)}(t), t \geq 0)_{n \in \mathbb{N}^*}$ converges in distribution as $n \rightarrow +\infty$ towards a process $(\Lambda^*(t), t \geq 0)$ which*

is the unique strong solution to the SDE :

$$\Lambda^*(t) = \frac{\mu_0}{\sigma} \cdot t + \frac{1}{\sigma} \int_0^t \sqrt{\Lambda^*(s)} dW(s), \quad (5.1.3)$$

where W is a Brownian motion.

We will revisit these results at the end of the chapter to compare them with our own findings.

5.2 The two-parameter case

We provide the definition of the process that will be considered in this work.

Definition 5.2.1. Let $a, b > 0$ such that $a + b = 1$ and $\lambda > 0$. We define for $X_{-1} = X_0 = 0$:

$$\forall n \geq 1, \quad X_n \sim \mathcal{P}(aX_{n-1} + bX_{n-2} + \lambda).$$

In the rest of this chapter, we will denote $\xi_{n-1} := aX_{n-1} + bX_{n-2} + \lambda \in \mathcal{F}_{n-1}$, so that $X_n \sim \mathcal{P}(\xi_{n-1})$.

5.2.1 A discrete martingale

The process (X_n) as defined does not possess any particular probabilistic properties that could help us understand its asymptotic behaviour. However, a transformation of this process satisfies a martingale property. This is the process (M_n) , which we define as follows:

Proposition 5.2.2. Let us define, for $n \in \mathbb{N}$:

$$M_n := X_n + bX_{n-1} - \lambda n, \quad \text{and} \quad \mathcal{F}_n := \sigma(X_m, m \leq n).$$

Then, the process (M_n) is a \mathcal{F}_n -martingale, and its predictable quadratic variation writes

$$\langle M \rangle_n = \sum_{k=1}^n \xi_{k-1}.$$

Proof. Let $n \geq 1$. We obtain from the definition of (X_n) and $a + b = 1$ that

$$\begin{aligned} \mathbb{E} \left[M_n \mid \mathcal{F}_{n-1} \right] &= \mathbb{E} \left[X_n + bX_{n-1} - \lambda n \mid \mathcal{F}_{n-1} \right] \\ &= aX_{n-1} + bX_{n-2} + \lambda + bX_{n-1} - \lambda n \\ &= M_{n-1}. \end{aligned}$$

Furthermore, for all $k \in \{1, \dots, n\}$:

$$\begin{aligned}
\mathbb{E} \left[\left(M_k - M_{k-1} \right)^2 \mid \mathcal{F}_{k-1} \right] &= \mathbb{E} \left[\left(X_k + bX_{k-1} - \lambda k - X_{k-1} - bX_{k-2} + \lambda(k-1) \right)^2 \mid \mathcal{F}_{k-1} \right] \\
&= \mathbb{E} \left[\left(X_k - aX_{k-1} - bX_{k-2} - \lambda \right)^2 \mid \mathcal{F}_{k-1} \right] \quad \text{since } a + b = 1 \\
&= \mathbb{E} \left[\left(X_k - \mathbb{E} \left[X_k \mid \mathcal{F}_{k-1} \right] \right)^2 \mid \mathcal{F}_{k-1} \right] \\
&= \text{Var} \left(X_k \mid \mathcal{F}_{k-1} \right) \\
&= aX_{k-1} + bX_{k-2} + \lambda = \xi_{k-1},
\end{aligned}$$

where we used that if Z is a Poisson random variable with parameter μ , then $\text{Var}(Z) = \mu$. We then deduce that for all $n \geq 1$:

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E} \left[\left(M_k - M_{k-1} \right)^2 \mid \mathcal{F}_{k-1} \right] = \sum_{k=1}^n \xi_{k-1}.$$

□

Remark 5.2.3. Recall that, by definition of $\langle M \rangle_n$, we have that $\left(M_n^2 - \langle M \rangle_n \right)_{n \in \mathbb{N}}$ is a \mathcal{F}_n -martingale.

The fact that (M_n) is a martingale already allows us to deduce non-trivial results about (X_n) and (ξ_n) :

Lemma 5.2.4. *Let $k \in \mathbb{N}^*$ and $(i, j) \in \mathbb{N}^2$.*

We denote $\mathbb{E}_{(i,j)}[\cdot] := \mathbb{E}[\cdot \mid X_0 = i, X_{-1} = j]$. Then,

1. $\mathbb{E}_{(i,j)}[\xi_{k-1}] \leq i + j + \lambda k, \quad \mathbb{E}_{(i,j)}[X_k] \leq i + j + \lambda(k+1).$
2. *It exists constants $C_1, C_2 > 0$ depending only on λ such that :*

$$\mathbb{E}_{(i,j)}[\xi_{k-1}^2] \leq (i + j)^2 + C_1 k(i + j) + C_2 k^2.$$

Proof. Proof of 1. Let $k \geq 1$, and $(i, j) \in \mathbb{N}^2$. Since $a < 1$,

$$\begin{aligned}
\xi_{k-1} &= aX_{k-1} + bX_{k-2} + \lambda \\
&\leq X_{k-1} + bX_{k-2} + \lambda \\
&= M_{k-1} + \lambda k.
\end{aligned}$$

Moreover, using the martingale property of (M_n) and $b < 1$, we have :

$$\begin{aligned}
\mathbb{E}_{(i,j)}[\xi_{k-1}] &\leq \mathbb{E}_{(i,j)}[M_{k-1}] + \lambda k \\
&= \mathbb{E}_{(i,j)}[M_0] + \lambda k \\
&\leq i + j + \lambda k.
\end{aligned}$$

With the same arguments, one can show that

$$\mathbb{E}_{(i,j)} [X_k] \leq i + j + \lambda(k + 1).$$

Proof of 2. Furthermore, employing the same calculations we obtain:

$$\xi_{k-1}^2 \leq M_{k-1}^2 + 2\lambda k M_{k-1} + \lambda^2 k^2.$$

Using 1. and Proposition 5.2.2 this leads us to :

$$\begin{aligned} \mathbb{E}_{(i,j)} [\xi_{k-1}^2] &\leq \mathbb{E}_{(i,j)} [M_{k-1}^2] + 2\lambda k \mathbb{E}_{(i,j)} [M_{k-1}] + \lambda^2 k^2 \\ &= \mathbb{E}_{(i,j)} [M_0^2] + \mathbb{E}_{(i,j)} [\langle M \rangle_{k-1}] + 2\lambda k \mathbb{E}_{(i,j)} [M_0] + \lambda^2 k^2 \\ &= (i + bj)^2 + \sum_{\ell=1}^{k-1} \mathbb{E}_{(i,j)} [\xi_{\ell-1}] + 2\lambda k(i + bj) + \lambda^2 k^2 \\ &\leq (i + j)^2 + \sum_{\ell=1}^{k-1} (i + j + \lambda \ell) + 2\lambda k(i + j) + \lambda^2 k^2 \\ &\leq (i + j)^2 + (k - 1)(i + j) + \lambda(k - 1)^2 + 2\lambda k(i + j) + \lambda^2 k^2 \\ &\leq (i + j)^2 + C_1 k(i + j) + C_2 k^2. \end{aligned}$$

This concludes the proof of the lemma. \square

As previously mentioned, we will study the convergence in distribution of a rescaled version of (X_n) with both space and time scaling. Specifically, we consider, for $T > 0$ the process $\left(\frac{X_{\lfloor nt \rfloor}}{n}, t \in [0, T] \right)_{n \in \mathbb{N}}$, which represents the natural rescaling according to Lemma 5.2.4 and corresponds to the rescaling mentioned in (5.1.2). To this end, we aim to use the theoretical framework developed by Ethier and Kurtz in their book [23] regarding diffusion approximations. In the following, we shall recall the main definitions as well as the central theorem of diffusion approximation theory.

5.2.2 Framework and theoretical results

We first place ourselves on an appropriate $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

Definition 5.2.5. Let X be a right-continuous measurable stochastic process on a metric space E , and A a subset of $\mathcal{B}(E) \times \mathcal{B}(E)$. We say that X is a **solution of the martingale problem for A** if, for all $(f, g) \in A$:

$$f(X(t)) - \int_0^t g(X(s)) ds$$

is a \mathcal{F}_t^X -martingale.

When an initial distribution $\mu \in \mathcal{P}(E)$ is specified, we say that a solution X of the martingale problem for A is a **solution of the martingale problem for (A, μ)** if $\mathbb{P}X(0)^{-1} = \mu$.

Definition 5.2.6. We say that **uniqueness** holds for solutions of the martingale problem for (A, μ) if any two solutions have the same finite-dimensional distributions.

If there exists a solution of the martingale problem for (A, μ) and uniqueness holds, we say that the martingale problem for (A, μ) is **well-posed**. If this holds for any $\mu \in \mathcal{P}(E)$, then the martingale problem for A is said to be **well-posed**.

The central theorem we will use to deal with our critical process (X_n) is the Theorem 4.1 page 354, here adapted for the case $d = 1$ as we study our process on \mathbb{R} .

Theorem 5.2.7 (Theorem 4.1 page 354, [23]). Let $\sigma^2 : \mathbb{R} \rightarrow \mathbb{R}^+$ be a continuous, non-negative function and $b : \mathbb{R} \rightarrow \mathbb{R}$ be continuous. Let

$$A := \left\{ \left(f, Gf = \frac{1}{2}\sigma^2 f'' + bf' \right), f \in \mathcal{C}_c^\infty(\mathbb{R}) \right\},$$

and suppose that the martingale problem for A is well-posed.

For $n \in \mathbb{N}^*$ let Y_n, B_n, A_n be processes with sample paths in $D_{\mathbb{R}}[0, \infty)$, such that $A_n(t) - A_n(s)$ is non-negative for $t > s \geq 0$. Set $\mathcal{F}_t^n := \sigma(Y_n(s), B_n(s), A_n(s) : s \leq t)$.

Let $\tau_n^r = \inf\{t : |Y_n(t)| \geq r\}$, and suppose that :

$$D_n := Y_n - B_n,$$

and

$$D_n^2 - A_n$$

are $\{\mathcal{F}_t^n\}$ -local martingales, and that for each $r > 0, T > 0$:

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |Y_n(t) - Y_n(t-)|^2 \right] = 0, \quad (5.2.1)$$

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |B_n(t) - B_n(t-)|^2 \right] = 0, \quad (5.2.2)$$

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |A_n(t) - A_n(t-)| \right] = 0, \quad (5.2.3)$$

$$\sup_{t \leq T \wedge \tau_n^r} \left| B_n(t) - \int_0^t b(Y_n(s)) ds \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad (5.2.4)$$

and

$$\sup_{t \leq T \wedge \tau_n^r} \left| A_n(t) - \int_0^t \sigma^2(Y_n(s)) ds \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (5.2.5)$$

Suppose that $PY_n(0)^{-1} \Rightarrow \nu \in \mathcal{P}(\mathbb{R})$. Then $\{Y_n\}$ converges in distribution to the solution of the martingale problem for (A, ν) .

Additionally, we will translate the conclusion, originally concerning the solution of a Martingale problem, into terms relevant to a specific SDE, which is our primary focus. Therefore, we will use the following:

Lemma 5.2.8. *Let $\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$ and $b : \mathbb{R} \rightarrow \mathbb{R}$ be locally bounded and Borel measurable, such that it exists $K > 0$ satisfying for all $x, y \in \mathbb{R}$:*

$$|\sigma^2(x) - \sigma^2(y)| \vee |b(x) - b(y)| \leq K|x - y|, \quad (5.2.6)$$

and :

$$|\sigma(x)|^2 \leq K(1 + |x|^2), \quad x \cdot b(x) \leq K(1 + |x|^2). \quad (5.2.7)$$

Let $\mu \in \mathcal{P}(\mathbb{R})$ and

$$A := \left\{ \left(f, Gf = \frac{1}{2}\sigma^2 f'' + bf' \right), f \in C_c^\infty(\mathbb{R}) \right\}.$$

Then,

- The Martingale Problem (A, μ) is well-posed.
- If X is a solution of the Martingale Problem (A, μ) , then X is solution of the following SDE :

$$X(t) = X(0) + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s), \quad (5.2.8)$$

where W is a Brownian motion independent of $X(0)$.

Proof.

- From (5.2.7) we can apply Theorem 3.10 page 299 which proves the existence of a solution to the SDE defined by (5.2.8). Therefore, with Corollary 3.4 page 295 we obtain the existence of a solution to the Martingale Problem for (A, μ) , for all $\mu \in \mathcal{P}(\mathbb{R})$.

Secondly, thanks to (5.2.6), we apply Theorem 3.9 page 298 which implies that the solution of the SDE (5.2.8) is pathwise unique, implying distribution uniqueness thanks to Theorem 3.6 page 296. Finally, according to Corollary 3.4, we deduce the uniqueness of solution for the Martingale Problem (A, μ) .

We thus proved that the Martingale Problem (A, μ) is well-posed.

- We simply apply Theorem 3.3 and, using that the Martingale Problem (A, μ) is well-posed, we deduce that X satisfy the SDE (5.2.8).

□

5.2.3 Time and space rescaling

Let $T > 0$. We intend to analyse the asymptotic behaviour of the sequence $\left(X_n(t) := \frac{X_{\lfloor nt \rfloor}}{n}, t \in [0, T] \right)_{n \in \mathbb{N}^*}$ as n tends to infinity, and establish its convergence

in law to a diffusion process that is solution of a specific stochastic differential equation (SDE). However, we cannot directly apply Theorem 5.2.7 to $X_n(t)$. First, we need to define an auxiliary process, denoted as $(Y_n(t), t \in [0, T])_{n \in \mathbb{N}^*}$, which is closely related to the martingale (M_n) defined in proposition 5.2.2. Using the properties established in Section 5.2.1, we will demonstrate that Theorem 5.2.7 can be applied to the process $(Y_n(t))$ defined as follows :

$$\forall t \in [0, T], \forall n \in \mathbb{N}^*, \quad Y_n(t) := \frac{X_{\lfloor tn \rfloor} + bX_{\lfloor tn \rfloor - 1}}{n}. \quad (5.2.9)$$

We also define the filtration

$$\mathcal{F}_t^n := \sigma(X_{\lfloor ns \rfloor}, s \leq t)$$

We will apply Theorem 5.2.7 on the process $(Y_n(t))$. For this purpose we introduce the functions b and σ defined by

$$\forall x \in \mathbb{R}, \quad b(x) = \lambda, \quad \text{and} \quad \sigma(x) = \sqrt{\frac{|x|}{1+b}}.$$

Theorem 5.2.9. *For all $T > 0$, the sequence $(Y_n(t), t \in [0, T])_{n \in \mathbb{N}^*}$ converges in distribution on $[0, T]$ to the solution of the following SDE :*

$$Y(t) = \lambda t + \frac{1}{\sqrt{1+b}} \int_0^t \sqrt{Y(s)} dW(s), \quad (5.2.10)$$

where W is a Brownian motion.

Proof. In order to use Theorem 5.2.7, we define the following stochastic processes :

$$\forall t \geq 0, \forall n \in \mathbb{N}^*, \quad B_n(t) := \frac{\lambda \lfloor tn \rfloor}{n}, \quad A_n(t) := \frac{\langle M \rangle_{\lfloor tn \rfloor}}{n^2},$$

as well as the following subset of $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$:

$$A := \{(f, Gf = \frac{1}{2}\sigma^2 f'' + bf'), f \in \mathcal{C}_c^\infty(\mathbb{R})\}.$$

From Proposition 5.2.2, we then have that for all $n \in \mathbb{N}^*$,

$$D_n := Y_n - B_n \quad \text{and} \quad D_n^2 - A_n \quad \text{are } \mathcal{F}_t^n\text{-martingale.}$$

Then, note that functions b and σ satisfy assumptions of Lemma 5.2.8. This means that the martingale problem for A is well-posed.

Futhermore, we show that equations (5.2.1)-(5.2.5) of theorem 5.2.7 hold.

Let $T, r > 0$.

Condition (5.2.1) Let $n \in \mathbb{N}^*$. First note that using Jensen inequality,

$$\mathbb{E} \left[\sup_{t \leq T} |Y_n(t) - Y_n(t-)|^2 \right] \leq \mathbb{E} \left[\sup_{t \leq T} |Y_n(t) - Y_n(t-)|^4 \right]^{1/2}. \quad (5.2.11)$$

Furthermore, for $t \leq T$,

$$|Y_n(t) - Y_n(t-)| = \frac{1}{n} \left| X_{[nt]} + bX_{[nt]-1} - \lim_{h \rightarrow 0^+} (X_{[n(t-h)]} + bX_{[n(t-h)]-1}) \right|. \quad (5.2.12)$$

Note that $\lim_{h \rightarrow 0^+} [n(t-h)]$ is either equal to $[nt]$ (in this case (5.2.12) reduces to 0 and the proof is over) or $[nt] - 1$. For this last case, we have using (5.2.12) and $a + b = 1$:

$$\begin{aligned} |Y_n(t) - Y_n(t-)|^4 &= \frac{1}{n^4} |X_{[nt]} + bX_{[nt]-1} - (X_{[nt]-1} + bX_{[nt]-2})|^4 \\ &= \frac{1}{n^4} |X_{[nt]} - aX_{[nt]-1} - bX_{[nt]-2}|^4 \end{aligned}$$

This yields:

$$\sup_{t \leq T} |Y_n(t) - Y_n(t-)|^4 \leq \frac{1}{n^4} \sum_{k=0}^{[nT]} |X_k - aX_{k-1} - bX_{k-2}|^4. \quad (5.2.13)$$

Recall that if Z is a Poisson random variable with parameter μ , the following holds :

$$\mathbb{E} \left[|Z - \mathbb{E}[Z]|^4 \right] = \mu(1 + 3\mu). \quad (5.2.14)$$

Using (5.2.14), we have for all $k \in \{0, \dots, [nT]\}$:

$$\begin{aligned} \mathbb{E} \left[|X_k - aX_{k-1} - bX_{k-2} - \lambda|^4 \right] &= \mathbb{E} \left[\mathbb{E} \left[\left(X_k - \mathbb{E} \left[X_k \mid \mathcal{F}_{k-1} \right] \right)^4 \mid \mathcal{F}_{k-1} \right] \right] \\ &= \mathbb{E} [\xi_{k-1}(1 + 3\xi_{k-1})] \end{aligned}$$

We use now Lemma 5.2.4, and since $X_0 = X_{-1} = 0$, we have that it exists a constant $C > 0$ such that :

$$\mathbb{E} [\xi_{k-1}(1 + 3\xi_{k-1})] \leq Ck^2 \leq Cn^2T^2. \quad (5.2.15)$$

So that for all $k \in \{0, \dots, [nT]\}$:

$$\mathbb{E} \left[|X_k - aX_{k-1} - bX_{k-2}|^4 \right] = O(n^2), \quad n \rightarrow +\infty.$$

And going back to (5.2.13) we have :

$$\begin{aligned} \mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |Y_n(t) - Y_n(t-)|^4 \right] &\leq \frac{1}{n^4} [nT] O(n^2), \quad n \rightarrow +\infty \\ &= O\left(\frac{1}{n}\right), \quad n \rightarrow +\infty \end{aligned}$$

Finally, according to (5.2.11),

$$\mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |Y_n(t) - Y_n(t-)|^2 \right] = O\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow +\infty.$$

So that (5.2.1) holds.

Conditions (5.2.2) and (5.2.4) : for all $n \in \mathbb{N}^*$, $t \leq T$,

$$|B_n(t) - B_n(t-)| = \frac{\lambda}{n} \left| [tn] - \lim_{h \rightarrow 0^+} [(t-h)n] \right| \leq \frac{\lambda}{n},$$

so that

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\sup_{t \leq T} |B_n(t) - B_n(t-)|^2 \right] = 0.$$

Analogously, we have :

$$\sup_{t \leq T} \left| B_n(t) - \int_0^t b(Y_n(s)) ds \right| = \frac{\lambda}{n} \sup_{t \leq T} |[tn] - tn| \leq \frac{\lambda}{n} \xrightarrow{n \rightarrow +\infty} 0.$$

Condition (5.2.3) : for all $n \in \mathbb{N}^*$, $t \geq 0$,

$$|A_n(t) - A_n(t-)| = \frac{1}{n^2} \left| \sum_{k=1}^{[nt]} \xi_{k-1} - \lim_{h \rightarrow 0^+} \sum_{\ell=1}^{[n(t-h)]} \xi_{\ell-1} \right| \leq \frac{\xi_{[nt]-1}}{n^2}.$$

Hence,

$$\sup_{t \leq T \wedge \tau_n^r} |A_n(t) - A_n(t-)| \leq \frac{1}{n^2} \sup_{t \leq T \wedge \tau_n^r} \xi_{[nt]}.$$

Recall that

$$\begin{aligned} \xi_{[nt]} &= aX_{[nt]} + bX_{[nt]-1} + \lambda \\ &\leq X_{[nt]} + bX_{[nt]-1} + \lambda \\ &= nY_n(t) + \lambda \end{aligned}$$

This yields,

$$\begin{aligned} \sup_{t \leq T \wedge \tau_n^r} |A_n(t) - A_n(t-)| &\leq \frac{\lambda}{n^2} + \frac{1}{n} \sup_{t \leq T \wedge \tau_n^r} Y_n(t) \\ &\leq \frac{\lambda}{n^2} + \frac{r}{n}. \end{aligned}$$

Finally,

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\sup_{t \leq T \wedge \tau_n^r} |A_n(t) - A_n(t-)| \right] = 0.$$

Condition (5.2.5) for $n \in \mathbb{N}^*$ and $t \leq T \wedge \tau_n^r$, since $X_{-1} = X_0 = 0$:

$$\begin{aligned} \int_0^t \sigma(Y_n(s))^2 ds &= \frac{1}{n} \left[\sum_{k=0}^{\lfloor nt \rfloor - 1} \sigma \left(\frac{X_k + bX_{k-1}}{n} \right)^2 + (nt - \lfloor nt \rfloor) \sigma(Y_n(t))^2 \right] \\ &= \frac{1}{n(1+b)} \left[\frac{1}{n} \sum_{k=0}^{\lfloor nt \rfloor - 1} X_k + \frac{b}{n} \sum_{k=0}^{\lfloor nt \rfloor - 1} X_{k-1} + (nt - \lfloor nt \rfloor) Y_n(t) \right] \\ &= \frac{1}{n(1+b)} \left[\frac{bX_{-1} + X_{\lfloor nt \rfloor - 1}}{n} + \frac{1+b}{n} \sum_{k=0}^{\lfloor nt \rfloor - 2} X_k + (nt - \lfloor nt \rfloor) Y_n(t) \right] \\ &= \frac{X_1 + \cdots + X_{\lfloor nt \rfloor - 2}}{n^2} + \frac{X_{\lfloor nt \rfloor - 1}}{n^2(1+b)} + \frac{(nt - \lfloor nt \rfloor) Y_n(t)}{n(1+b)} \end{aligned}$$

By definition,

$$A_n(t) = \frac{1}{n^2} \sum_{k=1}^{\lfloor nt \rfloor} \xi_{k-1} = \frac{X_1 + \cdots + X_{\lfloor nt \rfloor - 2}}{n^2} + \frac{aX_{\lfloor nt \rfloor - 1} + \lambda \lfloor nt \rfloor}{n^2}.$$

Using $a + b = 1$ and triangular inequality, this leads us to :

$$\begin{aligned} \left| A_n(t) - \int_0^t \sigma^2(Y_n(s)) ds \right| &= \left| \frac{aX_{\lfloor nt \rfloor - 1} + \lambda \lfloor nt \rfloor}{n^2} - \frac{X_{\lfloor nt \rfloor - 1}}{n^2(1+b)} - \frac{(nt - \lfloor nt \rfloor) Y_n(t)}{n(1+b)} \right| \\ &\leq \frac{1}{n^2(1+b)} \left| [a(1+b) - 1] X_{\lfloor nt \rfloor - 1} + \lambda(1+b) \lfloor nt \rfloor \right| + \frac{Y_n(t)}{n(1+b)} \\ &= \frac{1}{n^2(1+b)} \left| -b^2 X_{\lfloor nt \rfloor - 1} + \lambda(1+b) \lfloor nt \rfloor \right| + \frac{Y_n(t)}{n(1+b)} \end{aligned}$$

Recall that $b \in (0, 1)$ so that $b^2 \leq b$, and :

$$\begin{aligned} \left| A_n(t) - \int_0^t \sigma^2(Y_n(s)) ds \right| &\leq \frac{1}{n^2} (bX_{\lfloor nt \rfloor - 1} + \lambda(1+b) \lfloor nt \rfloor) + \frac{Y_n(t)}{n} \\ &\leq \frac{1}{n^2} (X_{\lfloor nt \rfloor} + bX_{\lfloor nt \rfloor - 1} + \lambda(1+b) \lfloor nt \rfloor) + \frac{Y_n(t)}{n} \\ &= \frac{1}{n^2} (nY_n(t) + \lambda(1+b) \lfloor nt \rfloor) + \frac{Y_n(t)}{n} \\ &= \frac{2Y_n(t)}{n} + \frac{\lambda(1+b) \lfloor nt \rfloor}{n^2}. \end{aligned}$$

Thus,

$$\sup_{t \leq T \wedge \tau_n^r} \left| A_n(t) - \int_0^t \sigma^2(Y_n(s)) ds \right| \leq \frac{2r + \lambda(1+b)T}{n}.$$

Finally, we proved :

$$\sup_{t \leq T} \left| A_n(t) - \int_0^t \sigma^2(Y_n(s)) ds \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

All the hypotheses of Theorem 5.2.7 are satisfied. Note that since $X_0 = X_{-1} = 0$, we also have $Y_n(0) \xrightarrow[n \rightarrow +\infty]{} 0$ a.s. By applying this theorem to the sequence $(Y_n(t))_{n \in \mathbb{N}^*}$, we conclude that (Y_n) converges in law as $n \rightarrow +\infty$ to a stochastic process $(Y(t), t \in [0, T])$ that is a solution to the Martingale problem $(A, 0)$. From Lemma 5.2.8, we deduce that $(Y(t), t \in [0, T])$ is a solution of the SDE (5.2.10). This concludes the proof. \square

Now, we aim to derive a result concerning the limit of the rescaled process $\left(X_n(t) = \frac{X_{\lfloor tn \rfloor}}{n}, t \in [0, T]\right)_{n \in \mathbb{N}^*}$. For this, note that for all $t \in [0, T], n \in \mathbb{N}^*$:

$$X_n(t) = \frac{1}{1+b} \left(Y_n(t) + b \frac{X_{\lfloor tn \rfloor} - X_{\lfloor tn \rfloor - 1}}{n} \right). \quad (5.2.16)$$

To prove the convergence in law of $(X_n(t), t \in [0, T])_{n \in \mathbb{N}^*}$ we will prove the following

Proposition 5.2.10. *For all $T > 0$,*

$$\sup_{t \leq T} \frac{X_{\lfloor tn \rfloor} - X_{\lfloor tn \rfloor - 1}}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Before presenting the proof of the proposition 5.2.10, we introduce a new martingale that will be particularly useful for the demonstration, along with some of its elementary properties.

Proposition 5.2.11. *For $n \in \mathbb{N}$, let us define :*

$$\tilde{M}_n := \left(-\frac{1}{b}\right)^n \left(X_n - X_{n-1} - \frac{\lambda}{1+b} \right).$$

The process (\tilde{M}_n) is a \mathcal{F}_n -martingale, and its quadratic variation writes :

$$\langle \tilde{M} \rangle_n = \sum_{k=1}^n \frac{1}{b^{2k}} \xi_{k-1}.$$

Proof of Proposition 5.2.11. Let $n \in \mathbb{N}^*$. Then,

$$\begin{aligned} \mathbb{E} \left[\tilde{M}_{n+1} \mid \mathcal{F}_n \right] &= \left(-\frac{1}{b}\right)^{n+1} \left(\mathbb{E} \left[X_{n+1} \mid \mathcal{F}_n \right] - X_n - \frac{\lambda}{1+b} \right) \\ &= \left(-\frac{1}{b}\right)^{n+1} \left((a-1)X_n + bX_{n-1} + \lambda - \frac{\lambda}{1+b} \right) \\ &= \left(-\frac{1}{b}\right)^{n+1} \left((-b)X_n + bX_{n-1} + \frac{b\lambda}{1+b} \right) \\ &= \left(-\frac{1}{b}\right)^n \left(X_n - X_{n-1} - \frac{\lambda}{1+b} \right) = \tilde{M}_n \end{aligned}$$

Moreover, for $k \in \{1, \dots, n\}$:

$$\begin{aligned}\tilde{M}_k - \tilde{M}_{k-1} &= \left(-\frac{1}{b}\right)^k \left(X_k - X_{k-1} - \frac{\lambda}{1+b} + b \left(X_{k-1} - X_{k-2} - \frac{\lambda}{1+b}\right)\right) \\ &= \left(-\frac{1}{b}\right)^k (X_k - aX_{k-1} - bX_{k-2} - \lambda)\end{aligned}$$

So that :

$$\mathbb{E} \left[\left(\tilde{M}_k - \tilde{M}_{k-1} \right)^2 \mid \mathcal{F}_{k-1} \right] = \frac{1}{b^{2k}} (aX_{k-1} + bX_{k-2} + \lambda) = \frac{1}{b^{2k}} \xi_{k-1},$$

which yields the result. \square

We can now prove the Proposition 5.2.10 :

Proof of Proposition 5.2.10. Let $T > 0$. We fix $\varepsilon > 0$, and we will prove that :

$$\mathbb{P} \left(\sup_{t \leq T} \frac{|X_{[tn]} - X_{[tn]-1}|}{n} > \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0.$$

Let $\delta > 0$ and $n \in \mathbb{N}$. First note that from Theorem 5.2.9, the sequence

$$(Y_j(t), t \in [0, T])_{j \in \mathbb{N}}$$

converges in law as $j \rightarrow +\infty$, which implies the tightness of this sequence. This means that it exists $K_\delta > 0$ big enough so that for all $n \in \mathbb{N}^*$,

$$\mathbb{P} \left(\sup_{t \in [0, T]} Y_n(t) \leq K_\delta \right) > 1 - \delta. \quad (5.2.17)$$

Moreover, note that if $k \in \{0, \dots, [Tn]\}$, then :

$$X_k \leq X_k + bX_{k-1} = nY_n(t_k),$$

where for all k , $t_k \in \left[\frac{k}{n}, \frac{k+1}{n}\right)$ is such that $[nt_k] = k$. Using (5.2.17), we have that for all $\delta > 0, n \in \mathbb{N}^*$, the event

$$G_\delta := \{X_k \leq K_\delta n, \forall k \in \{0, \dots, [Tn]\}\}$$

satisfies

$$\mathbb{P}(G_\delta) > 1 - \delta.$$

Hence,

$$\begin{aligned}\mathbb{P} \left(\sup_{t \leq T} \frac{|X_{[tn]} - X_{[tn]-1}|}{n} > \varepsilon \right) &\leq \mathbb{P} \left(\left\{ \sup_{t \leq T} \frac{|X_{[tn]} - X_{[tn]-1}|}{n} > \varepsilon \right\} \cap G_\delta \right) + \delta \\ &\leq \sum_{k=0}^{[Tn]} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) + \delta \quad (5.2.18)\end{aligned}$$

The idea of the proof relies on the martingale property of (\tilde{M}_n) . Indeed, since $b \in (0, 1)$, the quantity $(-1/b)^n$ will diverge exponentially fast, but the expectation of \tilde{M}_n will remain constant. Therefore, we will truncate the sum starting from a sufficiently large integer N , which will be explicitly chosen later, and then use Markov property. For all fixed $N \in \mathbb{N}^*$, we have :

$$\begin{aligned} \sum_{k=0}^{\lfloor Tn \rfloor} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) &= \sum_{k=0}^{N-1} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) \\ &\quad + \sum_{k=N}^{\lfloor Tn \rfloor} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right). \end{aligned} \quad (5.2.19)$$

Firstly, for $k \in \{0, \dots, N-1\}$, using Markov inequality and lemma 5.2.4, it exists $C_1 \geq 0$ not depending on n (recall that $X_0 = X_{-1} = 0$) such that :

$$\mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) \leq \frac{\mathbb{E} [|X_k - X_{k-1}|]}{\varepsilon n} \leq \frac{C_1 N}{\varepsilon n}$$

We then deduce :

$$\sum_{k=0}^{N-1} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) \leq \frac{C_1 N^2}{\varepsilon n}. \quad (5.2.20)$$

Moreover, note that for $k \in \{N, \dots, \lfloor Tn \rfloor\}$,

$$\mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) = \mathbb{E} \left[\mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \mid \mathcal{F}_{k-N} \right) \right], \quad (5.2.21)$$

and according to Markov property and the definition of the set G :

$$\mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \mid \mathcal{F}_{k-N} \right) \leq \max_{(i,j) \in \mathbb{N}^2} \left\{ \mathbb{P}_{(i,j)} \left(\frac{|X_N - X_{N-1}|}{n} > \varepsilon \right) \mid i, j \leq K_\delta n \right\}. \quad (5.2.22)$$

We will control the right-hand side of (5.2.22) by applying Markov's inequality to the fourth moment of $\frac{|X_N - X_{N-1}|}{n}$. To achieve this, we will use the martingale (\tilde{M}_n) defined in proposition 5.2.11 along with the following lemma:

Lemma 5.2.12. *Let $(i, j) \in \mathbb{N}^2$ such that $i, j \leq Kn$, for $K > 0$. Then, there exists a constant $C_2 \geq 0$ not depending on n such that :*

$$\mathbb{E}_{(i,j)} \left[|\tilde{M}_N - \tilde{M}_0|^4 \right] \leq \frac{C_2 (Nn)^2}{b^{4N}}.$$

Proof of Lemma 5.2.12. We use Rosenthal inequality (see [82]) on the martingale $S_n := \tilde{M}_n - \tilde{M}_0$, which satisfies $S_0 = 0$. Rosenthal inequality state that there exists a constant $\tilde{C} > 0$ such that :

$$\mathbb{E} [|S_N|^4] \leq \tilde{C} \left\{ \mathbb{E} \left[\left(\sum_{k=1}^N \mathbb{E} \left[(\tilde{M}_k - \tilde{M}_{k-1})^2 \mid \mathcal{F}_{k-1} \right] \right)^2 \right] + \sum_{k=1}^N \mathbb{E} \left[|\tilde{M}_k - \tilde{M}_{k-1}|^4 \right] \right\}, \quad (5.2.23)$$

- First, note that :

$$\left(\sum_{k=1}^N \mathbb{E} \left[(\tilde{M}_k - \tilde{M}_{k-1})^2 \mid \mathcal{F}_{k-1} \right] \right)^2 = \langle \tilde{M} \rangle_N^2.$$

And :

$$\langle \tilde{M} \rangle_N^2 = \left(\sum_{k=1}^N \frac{1}{b^{2k}} \xi_{k-1} \right)^2 \leq \frac{1}{b^{4N}} \left(\sum_{k=1}^N \xi_{k-1} \right)^2.$$

Taking expectations on the last inequality gives us :

$$\mathbb{E}_{(i,j)} \left[\langle \tilde{M} \rangle_N^2 \right] \leq \frac{1}{b^{4N}} \left(\sum_{k=1}^N \mathbb{E}_{(i,j)} [\xi_{k-1}^2] + \sum_{1 \leq \ell < m \leq N} \mathbb{E}_{(i,j)} [\xi_{\ell-1} \xi_{m-1}] \right). \quad (5.2.24)$$

Since (i, j) are such that $i, j \leq Kn$, according to lemma 5.2.4, it exists $c_1, c_2, c_3 > 0$ such that for all $q \in \{1, \dots, N\}$:

$$\mathbb{E}_{(i,j)} [\xi_{q-1}^2] \leq c_1 n^2 + c_2 Nn + c_3 N^2. \quad (5.2.25)$$

Thus, we have using Cauchy-Schwartz inequality it exists $\tilde{c}_1 > 0$ such that :

$$\mathbb{E}_{(i,j)} \left[\langle \tilde{M} \rangle_N^2 \right] \leq \frac{\tilde{c}_1 (Nn)^2}{b^{4N}} \quad (5.2.26)$$

- Furthermore, for all $i \in \{1, \dots, N\}$,

$$\begin{aligned} |\tilde{M}_k - \tilde{M}_{k-1}|^4 &= \left| \left(-\frac{1}{b} \right)^k \left(X_k - X_{k-1} - \frac{\lambda}{1+b} \right) - \left(-\frac{1}{b} \right)^{k-1} \left(X_{k-1} - X_{k-2} - \frac{\lambda}{1+b} \right) \right|^4 \\ &= \frac{1}{b^{4k}} \left| X_k - X_{k-1} - \frac{\lambda}{1+b} - (-b) \left(X_{k-1} - X_{k-2} - \frac{\lambda}{1+b} \right) \right|^4 \\ &= \frac{1}{b^{4k}} |X_k - (aX_{k-1} + bX_{k-2} + \lambda)|^4 \end{aligned} \quad (5.2.27)$$

Recall (5.2.14). We use again (5.2.25), so (5.2.27) implies, using lemma 5.2.4,

that it exists a constant $\tilde{c}_2 > 0$ such that :

$$\begin{aligned} \sum_{k=1}^N \mathbb{E}_{(i,j)} \left[|\tilde{M}_k - \tilde{M}_{k-1}|^4 \right] &= \sum_{k=1}^N \frac{1}{b^{4k}} \mathbb{E}_{(i,j)} \left[\mathbb{E} \left[\left(X_k - \mathbb{E} \left[X_k \mid \mathcal{F}_{k-1} \right] \right)^4 \mid \mathcal{F}_{k-1} \right] \right] \\ &\leq \frac{1}{b^{4N}} \sum_{k=1}^N \mathbb{E}_{(i,j)} \left[\xi_{k-1} (1 + 3\xi_{k-1}) \right] \\ &\leq \frac{\tilde{c}_2 N n^2}{b^{4N}} \end{aligned} \quad (5.2.28)$$

Finally, using (5.2.26) and (5.2.28), according to Rosenthal inequality (5.2.23), it exists a constant $C_2 > 0$ such that :

$$\mathbb{E}_{(i,j)} \left[|\tilde{M}_N - \tilde{M}_0|^4 \right] \leq \frac{C_2 (Nn)^2}{b^{4N}}. \quad (5.2.29)$$

This concludes the proof of the lemma. \square

We can now find an appropriate upper-bound for the expression (5.2.22). Indeed, it exists a constant $C > 0$ such that, for all $(i, j) \in \mathbb{N}^2$ satisfying $i, j \leq K_\delta n$, if $X_0 = i, X_{-1} = j$:

$$\left| \tilde{M}_0 \right| = \left| i - j - \frac{\lambda}{1+b} \right| \leq Cn.$$

This leads us to :

$$\begin{aligned} |X_N - X_{N-1}| &\leq b^N |\tilde{M}_N| + \frac{\lambda}{1+b} \\ &\leq b^N \left| \tilde{M}_N - \tilde{M}_0 \right| + b^N Cn + \frac{\lambda}{1+b} \end{aligned}$$

Let $n \in \mathbb{N}$ and $N \in \mathbb{N}^*$ big enough such that (recall that $b < 1$) :

$$\varepsilon - b^N C - \frac{\lambda}{(1+b)n} > 0. \quad (5.2.30)$$

Going back to (5.2.22), using lemma 5.2.12 and Markov inequality we have :

$$\begin{aligned} \mathbb{P}_{(i,j)} \left(\frac{|X_N - X_{N-1}|}{n} > \varepsilon \right) &\leq \mathbb{P}_{(i,j)} \left(\frac{b^N |\tilde{M}_N - \tilde{M}_0|}{n} + b^N C + \frac{\lambda}{(1+b)n} > \varepsilon \right) \\ &\leq \frac{b^{4N} \mathbb{E}_{(i,j)} \left[|\tilde{M}_N - \tilde{M}_0|^4 \right]}{n^4 \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4} \\ &\leq \frac{C_2 N^2}{n^2 \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4} \end{aligned}$$

Hence we just proved that :

$$\max_{(i,j) \in \mathbb{N}^2} \left\{ \mathbb{P}_{(i,j)} \left(\frac{|X_N - X_{N-1}|}{n} > \varepsilon \right) \mid i, j \leq K_\delta n \right\} \leq \frac{C_2 N^2}{n^2 \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4}.$$

Recall (5.2.21), from what precedes for $k \in \{N, \dots, \lfloor Tn \rfloor\}$,

$$\mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) \leq \frac{C_2 N^2}{n^2 \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4}.$$

This leads us to :

$$\sum_{k=N}^{\lfloor Tn \rfloor} \mathbb{P} \left(\frac{|X_k - X_{k-1}|}{n} > \varepsilon, G_\delta \right) \leq \frac{C_2 T N^2}{n \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4}.$$

To conclude, using (5.2.19) and (5.2.20) we can bound (5.2.18) the following way :

$$\mathbb{P} \left(\sup_{t \leq T} \frac{|X_{\lfloor tn \rfloor} - X_{\lfloor tn \rfloor - 1}|}{n} > \varepsilon \right) \leq \frac{C_1 N^2}{\varepsilon n} + \frac{C_2 T N^2}{n \left(\varepsilon - b^N C - \frac{\lambda}{1+b} \right)^4} + \delta$$

$$\xrightarrow{n \rightarrow +\infty} \delta. \quad (5.2.31)$$

Since (5.2.31) holds for any $\delta > 0$, this concludes our proof. \square

We thus immediately have, according to (5.2.16) :

Theorem 5.2.13. *For all $T > 0$, the sequence $(X_n(t), t \in [0, T])_{n \in \mathbb{N}}$ converges in law as $n \rightarrow +\infty$ towards a stochastic process $(X(t), t \in [0, T])$ that satisfies :*

$$\forall t \geq 0, \quad X(t) = \frac{1}{1+b} Y(t).$$

This means in particular that the process $(X(t), t \in [0, T])$ is a solution of the following SDE :

$$\forall t \in [0, T], \quad X(t) = \frac{\lambda}{1+b} t + \frac{1}{1+b} \int_0^t \sqrt{X(s)} dW(s), \quad (5.2.32)$$

where W is a Brownian motion.

Remark 5.2.14. (1) According to Theorem 5.2.13, an appropriate rescaled version of $(X(t), t \in [0, T])$ is a Bessel Squared process with dimension $m = 4(1+b)\lambda$. Specifically, setting $Z(t) := 4(1+b)^2 X(t)$, we have:

$$Z(t) = mt + 2 \int_0^t \sqrt{Z(s)} ds.$$

(2) We draw the reader's attention to the resemblance between (5.2.32) and the limiting SDE in [45], recalled in (5.1.3). Indeed, in our setting, λ plays the role of μ_0 . Furthermore, if we draw an analogy between $\|\phi\|_1$ and $\sum_{k=1}^p a_k$, for instance by considering a function ϕ with compact support, we can also relate $\lim_{t \rightarrow +\infty} \int_0^t s\phi(s)ds$ to $\sum_{k=1}^p ka_k$. Thus, in the case studied here, namely $p = 2$, we indeed have $a + b = 1$ as well as $\sigma = a + 2b = 1 + b$. Consequently, with these analogies, we establish a correspondence between (5.2.32) and (5.1.3).

5.3 Perspectives

The goal is, of course, to extend the results derived to a more general case with p becoming larger, that is, by considering

$$X_n \sim \mathcal{P}\left(a_1 X_{n-1} + a_2 X_{n-2} + \cdots + a_p X_{n-p} + \lambda\right),$$

where $p \in \mathbb{N}^*$ and $a_1, a_2, \dots, a_p \geq 0$ such that $a_1 + a_2 + \cdots + a_p = 1$. The beginning of the reasoning conducted in this chapter for the case $p = 2$ remains valid and is easily extendible to the case of an arbitrarily chosen p . Indeed, drawing inspiration from Proposition 5.2.2, we can define the process (M_n) as follows:

Proposition 5.3.1. *Let us denote for $k \in \{1, \dots, p\}$: $\rho_k := \sum_{i=k}^p a_i$. Thus, $\rho_1 = 1$ and $\rho_p = a_p$. For $n \in \mathbb{N}^*$, we define :*

$$M_n := \sum_{k=1}^p \rho_k X_{n-k+1} - \lambda n.$$

Then, (M_n) is a martingale.

Proof. Since for $k \in \{1, \dots, p-1\}$, $a_k + \rho_{k+1} = \rho_k$, we have:

$$\begin{aligned} \mathbb{E}\left[M_{n+1} \mid \mathcal{F}_n\right] &= \mathbb{E}\left[X_{n+1} + \sum_{k=2}^p \rho_k X_{n-k+2} - \lambda(n+1) \mid \mathcal{F}_n\right] \\ &= \mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] + \sum_{k=2}^p \rho_k X_{n-k+2} + -\lambda(n+1) \\ &= \sum_{k=1}^p a_k X_{n-k+1} + \lambda + \sum_{k=2}^p \rho_k X_{n-k+2} - \lambda(n+1) \\ &= \sum_{k=1}^{p-1} (a_k + \rho_{k+1}) X_{n-k+1} + a_p X_{n-p+1} - \lambda n \\ &= \sum_{k=1}^{p-1} \rho_k X_{n-k+1} - \lambda n \\ &= M_n \end{aligned}$$

□

Furthermore, we can similarly define an auxiliary process $Y_n(t)$ as follows:

$$Y_n(t) := \frac{\sum_{k=1}^p \rho_k X_{\lfloor tn \rfloor - k + 1}}{n},$$

and

$$b(x) = \lambda, \quad \sigma^2(x) = \frac{|x|}{C},$$

where $C := \sum_{k=1}^p ka_k$. It can be easily demonstrated that we can apply Theorem 5.2.7 to $(Y_n(t), t \geq 0)_{n \in \mathbb{N}^*}$ to ensure its convergence in law towards a limiting stochastic process $(Y(t), t \geq 0)$ satisfying the following SDE :

$$Y(t) = \lambda t + \frac{1}{C} \int_0^t \sqrt{Y(s)} dW(s). \quad (5.3.1)$$

The complications arise when we aim to go back to the process $\left(\frac{X_{\lfloor nt \rfloor}}{n}, t \geq 0\right)_{n \in \mathbb{N}^*}$, as the decomposition provided in (5.2.16) now yields:

$$\frac{X_{\lfloor nt \rfloor}}{n} = \frac{1}{C} \left(Y_n(t) + \sum_{k=2}^p \frac{X_{\lfloor nt \rfloor} - X_{\lfloor nt \rfloor - k + 1}}{n} \right).$$

Bibliography

- [1] Walter Appel. *Probabilités pour les non probabilistes*. H&K, 2013. ISBN: 978-2-35141-298-5.
- [2] Soeren Asmussen. *Applied Probability and Queues*. Springer Science & Business Media, 2008. ISBN: 978-0-387-21525-9.
- [3] Krishna B. Athreya, Kalyanapuram R. Parthasarathy, and Ganapathyiyer Sankaranarayanan. “Supercritical Age-Dependent Branching Processes with Immigration”. In: *Journal of Applied Probability* 11.4 (1974), pp. 695–702. ISSN: 0021-9002. DOI: [10.2307/3212553](https://doi.org/10.2307/3212553).
- [4] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. “Some limit theorems for Hawkes processes and application to financial statistics”. In: *Stochastic Processes and their Applications* 123.7 (2013), pp. 2475–2499. ISSN: 0304-4149. DOI: [10.1016/j.spa.2013.04.007](https://doi.org/10.1016/j.spa.2013.04.007).
- [5] Emmanuel Bacry and Jean-François Muzy. “Hawkes model for price and trades high-frequency dynamics”. In: *Quantitative Finance* 14.7 (2014), pp. 1147–1166. ISSN: 1469-7688. DOI: [10.1080/14697688.2014.897000](https://doi.org/10.1080/14697688.2014.897000).
- [6] Charles Bordenave and Giovanni Luca Torrisi. “Large Deviations of Poisson Cluster Processes”. In: *Stochastic Models* 23.4 (2007), pp. 593–625. DOI: [10.1080/15326340701645959](https://doi.org/10.1080/15326340701645959).
- [7] Pierre Brémaud. *Markov Chains Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 2001. ISBN: 978-1-4757-3124-8.
- [8] Pierre Brémaud and Laurent Massoulié. “Stability of Nonlinear Hawkes Processes”. In: *The Annals of Probability* 24.3 (1996), pp. 1563–1588. ISSN: 0091-1798. DOI: [10.1214/aop/1065725193](https://doi.org/10.1214/aop/1065725193).
- [9] Pierre Brémaud, Giovanna Nappo, and Giovanni L. Torrisi. “Rate of convergence to equilibrium of marked Hawkes processes”. In: *Journal of Applied Probability* 39.1 (2002), pp. 123–136. DOI: [10.1239/jap/1019737993](https://doi.org/10.1239/jap/1019737993).
- [10] Raiha Browning, Deborah Sulem, Kerrie Mengersen, Vincent Rivoirard, and Judith Rousseau. “Simple discrete-time self-exciting models can describe complex dynamic processes: A case study of COVID-19”. In: *PLOS ONE* 16.4 (2021). ISSN: 1932-6203. DOI: [10.1371/journal.pone.0250015](https://doi.org/10.1371/journal.pone.0250015).
- [11] Hermann Brunner. *Volterra Integral Equations: An Introduction to Theory and Applications*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2017. ISBN: 9781316982655.
- [12] Patrick Cattiaux, Laetitia Colombani, and Manon Costa. “Limit theorems for Hawkes processes including inhibition”. In: *Stochastic Processes and their Applications* 149 (2022), pp. 404–426. ISSN: 0304-4149. DOI: [10.1016/j.spa.2022.04.002](https://doi.org/10.1016/j.spa.2022.04.002).

- [13] Manon Costa, Carl Graham, Laurence Marsalle, and Viet-Chi Tran. “Renewal in Hawkes processes with self-excitation and inhibition”. In: *Advances in Applied Probability* 52.3 (2020), pp. 879–915. DOI: [10.1017/apr.2020.19](https://doi.org/10.1017/apr.2020.19).
- [14] Manon Costa, Pascal Maillard, and Anthony Muraro. “(Almost) complete characterization of stability of a discrete-time Hawkes process with inhibition and memory of length two”. In: *Journal of Applied Probability* (2024), pp. 1–22. DOI: [10.1017/jpr.2024.28](https://doi.org/10.1017/jpr.2024.28).
- [15] D. R. Cox. “Some Statistical Methods Connected with Series of Events”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 17.2 (1985), pp. 129–157. ISSN: 0035-9246. DOI: [10.1111/j.2517-6161.1985.tb00188.x](https://doi.org/10.1111/j.2517-6161.1985.tb00188.x).
- [16] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I*. Probability and its Applications. Springer-Verlag, 2003. ISBN: 978-0-387-95541-4. DOI: [10.1007/b97277](https://doi.org/10.1007/b97277).
- [17] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume II*. Probability and Its Applications. Springer, 2008. ISBN: 978-0-387-49835-5. DOI: [10.1007/978-0-387-49835-5](https://doi.org/10.1007/978-0-387-49835-5).
- [18] Leonard M. Delves and Joan Walsh. *Numerical solution of integral equations*. Clarendon Press, Oxford, 1974. ISBN: 9780198533429.
- [19] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. *Markov Chains*. Springer, 2018. ISBN: 978-3-319-97704-1.
- [20] Aline Duarte, Eva Löcherbach, and Guilherme Ost. “Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels”. In: *ESAIM: PS* 23 (2019), pp. 770–796. DOI: [10.1051/ps/2019005](https://doi.org/10.1051/ps/2019005).
- [21] Richard Durrett. *Probability: theory and examples*. Cambridge University Press, 1996. ISBN: 978-1108473682.
- [22] Céline Duval, Eric Luçon, and Christophe Pouzat. “Interacting Hawkes processes with multiplicative inhibition”. In: *Stochastic Processes and their Applications* 148 (2022), pp. 180–226. ISSN: 0304-4149. DOI: [10.1016/j.spa.2022.02.008](https://doi.org/10.1016/j.spa.2022.02.008).
- [23] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2009. ISBN: 978-0-470-31732-7.
- [24] Guanhua Fang, Ganggang Xu, Haochen Xu, Xuening Zhu, and Yongtao Guan. “Group Network Hawkes Process”. In: *Journal of the American Statistical Association* 0.0 (2023), pp. 1–17. ISSN: 0162-1459. DOI: [10.1080/01621459.2023.2257889](https://doi.org/10.1080/01621459.2023.2257889).
- [25] Guy Fayolle, Vadim A. Malyshev, and Mikhail V. Menshikov. *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge University Press, 1995. ISBN: 978-0-521-46197-9.

- [26] René Ferland, Alain Latour, and Driss Oraichi. “Integer-Valued GARCH Process”. In: *Journal of Time Series Analysis* 27.6 (2006), pp. 923–942. ISSN: 1467-9892. DOI: [10.1111/j.1467-9892.2006.00496.x](https://doi.org/10.1111/j.1467-9892.2006.00496.x).
- [27] Konstantinos Fokianos. “Integer Autoregressive and Moving Average Models”. In: *Annales de l’ISUP XXXXVIII.1-2* (2004), pp. 43–58.
- [28] Konstantinos Fokianos and Roland Fried. “Interventions in INGARCH processes”. In: *Journal of Time Series Analysis* 31.3 (2010), pp. 210–225. ISSN: 1467-9892. DOI: [10.1111/j.1467-9892.2010.00657.x](https://doi.org/10.1111/j.1467-9892.2010.00657.x).
- [29] Konstantinos Fokianos, Anders Rahbek, and Dag Tjøstheim. “Poisson Autoregression”. In: *Journal of the American Statistical Association* 104.488 (2009), pp. 1430–1439. ISSN: 0162-1459. DOI: [10.1198/jasa.2009.tm08270](https://doi.org/10.1198/jasa.2009.tm08270).
- [30] Konstantinos Fokianos and Dag Tjøstheim. “Log-linear Poisson autoregression”. In: *Journal of Multivariate Analysis* 102.3 (2011), pp. 563–578. ISSN: 0047-259X. DOI: [10.1016/j.jmva.2010.11.002](https://doi.org/10.1016/j.jmva.2010.11.002).
- [31] F. Gordon Foster. “On the Stochastic Matrices Associated with Certain Queuing Processes”. In: *The Annals of Mathematical Statistics* 24.3 (1953), pp. 355–360. ISSN: 0003-4851, 2168-8990. DOI: [10.1214/aoms/1177728976](https://doi.org/10.1214/aoms/1177728976).
- [32] Eric W. Fox, Martin B. Short, Frederic P. Schoenberg, Kathryn D. Coronges, and Andrea L. Bertozzi. “Modeling E-mail Networks and Inferring Leadership Using Self-Exciting Point Processes”. In: *Journal of the American Statistical Association* 111.514 (2016), pp. 564–584. ISSN: 0162-1459. DOI: [10.1080/01621459.2015.1135802](https://doi.org/10.1080/01621459.2015.1135802).
- [33] Haig Galajikian. “Non-Linear Integral Equations of the Volterra Type”. In: *Annals of Mathematics* 16 (1914), pp. 172–192. ISSN: 0003-486X. DOI: [10.2307/1968060](https://doi.org/10.2307/1968060).
- [34] Jean Gallier and Jocelyn Quaintance. *Linear Algebra and Optimization with Applications to Machine Learning: Volume I*. World Scientific, 2020. ISBN: 978-9811206399.
- [35] Antonio Galves and Eva Löcherbach. “Modeling networks of spiking neurons as interacting processes with memory of variable length”. In: *Journal de la Société Française de Statistique* 157.1 (2016), pp. 17–32. ISSN: 2102-6238.
- [36] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Boston, MA: Birkhäuser, 1994. ISBN: 978-0-8176-4771-1. DOI: [10.1007/978-0-8176-4771-1](https://doi.org/10.1007/978-0-8176-4771-1).
- [37] Matthieu Gilson and Jean-Pascal Pfister. “Propagation of Spiking Moments in Linear Hawkes Networks”. In: *SIAM Journal on Applied Dynamical Systems* 19.2 (2020), pp. 828–859. DOI: [10.1137/18M1220030](https://doi.org/10.1137/18M1220030).
- [38] Vesa Halava, Tero Harju, and Mika Hirvensalo. “Positivity of second order linear recurrent sequences”. In: *Discrete Applied Mathematics* 154.3 (2006), pp. 447–451. ISSN: 0166-218X. DOI: [10.1016/j.dam.2005.10.009](https://doi.org/10.1016/j.dam.2005.10.009).

- [39] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. “Lasso and probabilistic inequalities for multivariate point processes”. In: *Bernoulli* 21.1 (2015), pp. 83–143. ISSN: 1350-7265. DOI: [10.3150/13-BEJ562](https://doi.org/10.3150/13-BEJ562).
- [40] Alan G. Hawkes. “Hawkes processes and their applications to finance: a review”. In: *Quantitative Finance* 18.2 (2018), pp. 193–198. DOI: [10.1080/14697688.2017.1403131](https://doi.org/10.1080/14697688.2017.1403131).
- [41] Alan G. Hawkes. “Spectra of Some Self-Exciting and Mutually Exciting Point Processes”. In: *Biometrika* 58.1 (1971), pp. 83–90. ISSN: 0006-3444. DOI: [10.2307/2334319](https://doi.org/10.2307/2334319).
- [42] Alan G. Hawkes and Leonidas Adamopoulos. “Cluster models for earthquakes : regional comparisons”. In: *Bulletin of the International Statistical Institute* 45.3 (1973), pp. 454–460. ISSN: 0373-0441.
- [43] Alan G. Hawkes and David Oakes. “A Cluster Process Representation of a Self-Exciting Process”. In: *Journal of Applied Probability* 11.3 (1974), pp. 493–503. ISSN: 0021-9002. DOI: [10.2307/3212693](https://doi.org/10.2307/3212693).
- [44] Andréas Heinen. “Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model”. In: (2003). DOI: [10.2139/ssrn.1117187](https://doi.org/10.2139/ssrn.1117187).
- [45] Ulrich Horst and Wei Xu. “Functional Limit Theorems for Hawkes Processes”. In: (2024). DOI: [10.48550/arXiv.2401.11495](https://doi.org/10.48550/arXiv.2401.11495).
- [46] Lorick Huang and Mahmoud Khabou. “Nonlinear Poisson autoregression and nonlinear Hawkes processes”. In: *Stochastic Processes and their Applications* 161 (2023), pp. 201–241. ISSN: 0304-4149. DOI: [10.1016/j.spa.2023.03.015](https://doi.org/10.1016/j.spa.2023.03.015).
- [47] Thibault Jaisson and Mathieu Rosenbaum. “Limit Theorems for Nearly Unstable Hawkes Processes”. In: *The Annals of Applied Probability* 25.2 (2015), pp. 600–631. ISSN: 1050-5164. DOI: [10.1214/14-AAP1005](https://doi.org/10.1214/14-AAP1005).
- [48] Thibault Jaisson and Mathieu Rosenbaum. “Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes”. In: *The Annals of Applied Probability* 26.5 (2016), pp. 2860–2882. ISSN: 10505164. DOI: [10.1214/15-AAP1164](https://doi.org/10.1214/15-AAP1164).
- [49] Mikyoung Jun and Scott Cook. “Flexible multivariate spatio-temporal Hawkes process models of terrorism”. In: (2023). DOI: [10.48550/arXiv.2202.12346](https://doi.org/10.48550/arXiv.2202.12346).
- [50] Dmytro Karabash. “On Stability of Hawkes Process”. In: (2013). DOI: [10.48550/arXiv.1201.1573](https://doi.org/10.48550/arXiv.1201.1573).
- [51] Manuel Kauers and Peter Paule. *The Concrete Tetrahedron*. Springer, 2011. ISBN: 978-3-7091-0444-6.
- [52] Sato Ken-Iti. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. ISBN: 978-0-521-55302-5.
- [53] J. F. C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Oxford, New York: Oxford University Press, 1992. ISBN: 978-0-19-853693-2.

- [54] Matthias Kirchner. “Hawkes and INAR(∞) processes”. In: *Stochastic Processes and their Applications* 126.8 (2016), pp. 2494–2525. ISSN: 0304-4149. DOI: [10.1016/j.spa.2016.02.008](https://doi.org/10.1016/j.spa.2016.02.008).
- [55] Matthias Kirchner. “Perspectives on Hawkes Processes”. PhD thesis. ETH Zurich, 2017. DOI: [10.3929/ETHZ-B-000161487](https://doi.org/10.3929/ETHZ-B-000161487).
- [56] Junhyeon Kwon, Yingcai Zheng, and Mikyoung Jun. “Flexible spatio-temporal Hawkes process models for earthquake occurrences”. In: *Spatial Statistics* 54 (2023), p. 100728. ISSN: 2211-6753. DOI: [10.1016/j.spasta.2023.100728](https://doi.org/10.1016/j.spasta.2023.100728).
- [57] Stamatina Lamprinakou, Axel Gandy, and Emma McCoy. “Using a latent Hawkes process for epidemiological modelling”. In: *PLOS ONE* 18.3 (2023). ISSN: 1932-6203. DOI: [10.1371/journal.pone.0281370](https://doi.org/10.1371/journal.pone.0281370).
- [58] Patrick J. Laub, Thomas Taimre, and Philip K. Pollett. “Hawkes Processes”. In: (2015). DOI: [10.48550/arXiv.1507.02822](https://doi.org/10.48550/arXiv.1507.02822).
- [59] Young Lee, Kar Wai Lim, and Cheng Soon Ong. “Hawkes Processes with Stochastic Excitations”. en. In: *Proceedings of The 33rd International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, 2016, pp. 79–88.
- [60] P. A. W. Lewis and G. S. Shedler. “Simulation of nonhomogeneous poisson processes by thinning”. In: *Naval Research Logistics Quarterly* 26.3 (1979), pp. 403–413. DOI: [10.1002/nav.3800260304](https://doi.org/10.1002/nav.3800260304).
- [61] Mengya Liu, Fukang Zhu, Jianfeng Li, and Chuning Sun. “A Systematic Review of INGARCH Models for Integer-Valued Time Series”. In: *Entropy* 25.6 (2023), p. 922. ISSN: 1099-4300. DOI: [10.3390/e25060922](https://doi.org/10.3390/e25060922).
- [62] Zeqing Liu, Shin Min Kang, and Jeong Sheok Ume. “Solvability and asymptotic stability of a nonlinear functional-integral equation”. In: *Applied Mathematics Letters* 24.6 (2011), pp. 911–917. ISSN: 0893-9659. DOI: [10.1016/j.aml.2011.01.001](https://doi.org/10.1016/j.aml.2011.01.001).
- [63] Eva Löcherbach. “Spiking neurons : interacting Hawkes processes, mean field limits and oscillations”. In: *ESAIM: Proceedings and Surveys* 60 (2017), pp. 90–103. DOI: [10.1051/proc/201760090](https://doi.org/10.1051/proc/201760090).
- [64] Eric Luçon and Wilhelm Stannat. “Mean field limit for disordered diffusions with singular interactions”. In: *The Annals of Applied Probability* 24.5 (2014), pp. 1946–1993. ISSN: 10505164.
- [65] Hongyuan Mei and Jason M Eisner. “The neural hawkes process: A neurally self-modulating multivariate point process”. In: *Advances in neural information processing systems* 30 (2017). DOI: [10.48550/arXiv.1612.09328](https://doi.org/10.48550/arXiv.1612.09328).
- [66] Sylvie Méléard. *Modèles aléatoires en Ecologie et Evolution*. Springer, 2016. ISBN: 978-3-662-49454-7.
- [67] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2009. ISBN: 978-0-521-73182-9.

- [68] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. “Self-Exciting Point Process Modeling of Crime”. In: *Journal of the American Statistical Association* 106.493 (2011), pp. 100–108. ISSN: 0162-1459. DOI: [10.1198/jasa.2011.ap09546](https://doi.org/10.1198/jasa.2011.ap09546).
- [69] O. T. O’Meara. *Introduction to Quadratic Forms*. Berlin, Heidelberg: Springer, 1973. ISBN: 978-3-662-41922-9. DOI: [10.1007/978-3-662-41922-9](https://doi.org/10.1007/978-3-662-41922-9).
- [70] Yosihiko Ogata. “Space-time model for regional seismicity and detection of crustal stress changes”. en. In: *Journal of Geophysical Research: Solid Earth* 109.B3 (2004). ISSN: 2156-2202. DOI: [10.1029/2003JB002621](https://doi.org/10.1029/2003JB002621).
- [71] Yosihiko Ogata. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 9–27. ISSN: 0162-1459. DOI: [10.1080/01621459.1988.10478560](https://doi.org/10.1080/01621459.1988.10478560).
- [72] Peter Olofsson. “General Branching Processes with Immigration”. In: *Journal of Applied Probability* 33.4 (1996), pp. 940–948. ISSN: 0021-9002. DOI: [10.2307/3214975](https://doi.org/10.2307/3214975).
- [73] Mads Bonde Raad and Eva Löcherbach. “Stability for Hawkes processes with inhibition”. In: *Electronic Communications in Probability* 25 (2020), pp. 1–9. DOI: [10.1214/20-ECP312](https://doi.org/10.1214/20-ECP312).
- [74] Jakob Gulddahl Rasmussen. “Bayesian Inference for Hawkes Processes”. In: *Methodology and Computing in Applied Probability* 15.3 (2013), pp. 623–642. ISSN: 1387-5841, 1573-7713. DOI: [10.1007/s11009-011-9272-5](https://doi.org/10.1007/s11009-011-9272-5).
- [75] Patricia Reynaud-Bouret and Sophie Schbath. “Adaptive estimation for Hawkes processes; application to genome analysis”. In: *The Annals of Statistics* 38.5 (2010), pp. 2781–2822. ISSN: 0090-5364, 2168-8966. DOI: [10.1214/10-AOS806](https://doi.org/10.1214/10-AOS806).
- [76] Emmanuel Schertzer and Florian Simatos. “Height and contour processes of Crump-Mode-Jagers forests (I): general distribution and scaling limits in the case of short edges”. In: *Electronic Journal of Probability* 23 (2018), pp. 1–43. ISSN: 1083-6489, 1083-6489. DOI: [10.1214/18-EJP151](https://doi.org/10.1214/18-EJP151).
- [77] Youngsoo Seol. “Limit theorems for discrete Hawkes processes”. In: *Statistics & Probability Letters* 99 (2015), pp. 223–229. ISSN: 0167-7152. DOI: [10.1016/j.spl.2015.01.023](https://doi.org/10.1016/j.spl.2015.01.023).
- [78] Måns Thulin. “The cost of using exact confidence intervals for a binomial proportion”. In: *Electronic Journal of Statistics* 8.1 (2014), pp. 817–840. ISSN: 1935-7524, 1935-7524. DOI: [10.1214/14-EJS909](https://doi.org/10.1214/14-EJS909).
- [79] Michael Tortorella. “Numerical Solutions of Renewal-Type Integral Equations”. In: *INFORMS Journal on Computing* 17.1 (2005), pp. 66–74. DOI: [10.1287/ijoc.1030.0054](https://doi.org/10.1287/ijoc.1030.0054).
- [80] Michael Tortorella. “Numerical Solutions of Renewal-Type Integral Equations”. In: *INFORMS Journal on Computing* 17.1 (2005), pp. 66–74. ISSN: 1091-9856. DOI: [10.1287/ijoc.1030.0054](https://doi.org/10.1287/ijoc.1030.0054).

-
- [81] David Williams. *Probability with Martingales*. Cambridge University Press, 1991. ISBN: 978-0-521-40605-5.
- [82] Andrew T.A. Wood. “Rosenthal’s inequality for point process martingales”. In: *Stochastic Processes and their Applications* 81.2 (1999), pp. 231–246. ISSN: 0304-4149. DOI: [10.1016/S0304-4149\(99\)00005-8](https://doi.org/10.1016/S0304-4149(99)00005-8).
- [83] Wei Xu. “Scaling Limits for Crump-Mode-Jagers Processes with Immigration via Stochastic Volterra Equations”. In: (2018). arXiv:1809.05931 [math] version: 3. DOI: [10.48550/arXiv.1809.05931](https://doi.org/10.48550/arXiv.1809.05931).
- [84] Baichuan Yuan, Hao Li, Andrea L. Bertozzi, P. Jeffrey Brantingham, and Mason A. Porter. “Multivariate Spatiotemporal Hawkes Processes and Network Reconstruction”. In: *SIAM Journal on Mathematics of Data Science* 1.2 (2019), pp. 356–382. DOI: [10.1137/18M1226993](https://doi.org/10.1137/18M1226993).
- [85] Fukang Zhu and Dehui Wang. “Estimation and testing for a Poisson autoregressive model”. In: *Metrika* 73.2 (2011), pp. 211–230. ISSN: 1435-926X. DOI: [10.1007/s00184-009-0274-z](https://doi.org/10.1007/s00184-009-0274-z).
- [86] Lingjiong Zhu. “Central Limit Theorem for Nonlinear Hawkes Processes”. In: *Journal of Applied Probability* 50.3 (2013), pp. 760–771. ISSN: 0021-9002.
- [87] Lingjiong Zhu. “Large deviations for Markovian nonlinear Hawkes processes”. In: *The Annals of Applied Probability* 25.2 (2015), pp. 548–581. DOI: [10.1214/14-AAP1003](https://doi.org/10.1214/14-AAP1003).
- [88] Shixiang Zhu and Yao Xie. “Spatiotemporal-textual point processes for crime linkage detection”. In: *The Annals of Applied Statistics* 16.2 (2022), pp. 1151–1170. ISSN: 1932-6157, 1941-7330. DOI: [10.1214/21-AOAS1538](https://doi.org/10.1214/21-AOAS1538).

Titre : Processus de Hawkes en temps discret avec inhibition

Mots clés : Probabilités, Processus stochastiques, Chaînes de Markov

Résumé : Cette thèse porte sur les processus de Hawkes, qui sont des processus stochastiques à temps continu dont l'intensité est aléatoire et dépend de l'histoire complète du processus. Ces processus ont été introduits par Hawkes (1971) pour modéliser une dynamique auto-excitante. Une généralisation de ces processus consiste à intégrer un effet d'auto-inhibition, pour laquelle la littérature est plus réduite et souffre notamment de l'absence d'un critère nécessaire et suffisant d'existence d'une version stationnaire prenant véritablement en compte l'effet inhibiteur du modèle.

Dans un premier chapitre, nous effectuons une analyse numérique portant sur la simulation de la solution d'une équation de renouvellement non-linéaire, ayant un lien fort avec l'espérance de l'intensité conditionnelle d'un processus de Hawkes. En outre, nous présentons des exemples de simulation de ces processus dans des cadres ne respectant pas le critère mentionné précédemment, illustrant ainsi la nature restrictive de ce dernier. Dans la suite, nous nous concentrons plus spécifiquement sur une version discrète des processus de Hawkes avec inhibition. Ce modèle se présente sous la forme d'un processus auto-régressif de Poisson, où le paramètre est aléatoire et dépend des réalisations passées du processus. Nous permettons à ces paramètres de prendre des valeurs négatives pour modéliser un effet inhibiteur.

Le deuxième chapitre porte sur le cas où la mémoire de ce processus est de taille 2. Dans ce cadre, nous classifions intégralement le comportement asymptotique de ce processus en fonction des paramètres du modèle, à l'exception des cas frontières. Pour mener à bien ces résultats, nous utilisons le formalisme des chaînes de Markov tel que décrit par Douc, Moulines, Priouret (2018), et nous servons de théorèmes type critère de Foster (1953) via des fonctions de Lyapunov. Nous tirons également parti de la comparaison entre le comportement de ce processus avec le comportement des solutions d'une suite récurrente linéaire naturellement associée au modèle.

Dans le troisième chapitre, nous étendons notre étude au cas où la mémoire du processus est de taille 3, en donnant quelques résultats dans le cas général d'une mémoire de taille arbitrairement grande. Nous complétons notre étude théorique par des simulations numériques permettant d'étayer nos conjectures et donner une intuition sur le comportement en temps long de ce processus.

Le dernier chapitre est un travail en cours autour des cas critiques du processus étudié. On utilise notamment des outils d'approximation de diffusions issus du livre d'Ethier et Kurtz (1986). Nous montrons qu'après un changement adéquat d'échelle et de temps, le processus converge vers une diffusion dirigée par une EDS que nous explicitons.

Title: Discrete time Hawkes processes with inhibition

Key words: Probability, Stochastic processes, Markov chains

Abstract: This thesis focuses on Hawkes processes, which are continuous-time stochastic processes whose intensity is random and depends on the entire history of the process. These processes were introduced by Hawkes (1971) to model self-exciting dynamics. A generalization of these processes involves incorporating a self-inhibition effect, for which the literature is more limited and notably lacks a necessary and sufficient criterion for the existence of a stationary version that truly accounts for the inhibitory effect of the model.

In the first chapter, we conduct a numerical analysis on the simulation of the solution to a nonlinear renewal equation, which is closely related to the expectation of the conditional intensity of a Hawkes process. Additionally, we present examples of simulations of these processes in settings that do not meet the previously mentioned criterion, thereby illustrating the restrictive nature of this criterion.

Next, we focus more specifically on a discrete version of Hawkes processes with inhibition. This model takes the form of an autoregressive Poisson process, where the parameter is random and depends on the past realizations of the process. We allow these parameters to take negative values to model an inhibitory effect.

The second chapter examines the case where the memory of this process is of size 2. In this context, we classify the asymptotic behavior of this process for the whole range of parameters, except for boundary cases. To achieve these results, we use the formalism of Markov chains as described by Douc, Moulines and Priouret (2018), and employ Foster's (1953) criterion-type theorems via Lyapunov functions. We also take advantage of comparing the behavior of this process with the behavior of solutions to a naturally associated linear recurrence sequence.

In the third chapter, we extend our study to the case where the memory of the process is of size 3, and provide some results in the general case of arbitrarily large memory. We complement our theoretical study with numerical simulations to support our conjectures and provide intuition about the long-term behavior of this process.

The final chapter is ongoing work on the critical cases of the studied process. We notably use diffusion approximation tools from the book by Ethier and Kurtz (1986). We show that after an appropriate time and space rescaling, the process converges to a diffusion governed by a stochastic differential equation, which we explicitly describe.