



**HAL**  
open science

# Inférence non-paramétrique pour des interactions poissonniennes

Laure Sansonnet

► **To cite this version:**

Laure Sansonnet. Inférence non-paramétrique pour des interactions poissonniennes. Mathématiques générales [math.GM]. Université Paris Sud - Paris XI, 2013. Français. NNT : 2013PA112084 . tel-00835427

**HAL Id: tel-00835427**

**<https://theses.hal.science/tel-00835427v1>**

Submitted on 18 Jun 2013

**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.

## UNIVERSITÉ PARIS-SUD

École Doctorale de Mathématiques de la Région Paris-Sud  
Laboratoire de Mathématiques de la Faculté des Sciences d'Orsay

## THÈSE DE DOCTORAT SUR TRAVAUX

présentée pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS-SUD

*Spécialité : Mathématiques*

par

**Laure SANSONNET**

---

### **Inférence non-paramétrique pour des interactions poissoniennes**

---

Soutenue le 14 juin 2013 devant le jury composé de :

Béatrice	LAURENT-BONNEAU	Institut de Mathématiques de Toulouse	<i>Examinatrice</i>
Pascal	MASSART	Université Paris-Sud	<i>Examinateur</i>
Richard	NICKL	University of Cambridge	<i>Rapporteur</i>
Patricia	REYNAUD-BOURET	Université Nice Sophia-Antipolis	<i>Directrice</i>
Vincent	RIVOIRARD	Université Paris Dauphine	<i>Directeur</i>
Stéphane	ROBIN	AgroParisTech	<i>Rapporteur</i>

Département de  
 mathématiques  
d'Orsay

Thèse préparée au  
Laboratoire de Mathématiques, CNRS UMR 8628  
Bât. 425, Université Paris-Sud  
15, rue Georges Clémenceau, 91405 Orsay Cedex, France

## Inférence non-paramétrique pour des interactions poissonniennes

**Résumé :** L'objet de cette thèse est d'étudier divers problèmes de statistique non-paramétrique dans le cadre d'un modèle d'interactions poissonniennes. De tels modèles sont, par exemple, utilisés en neurosciences pour analyser les interactions entre deux neurones au travers leur émission de potentiels d'action au cours de l'enregistrement de l'activité cérébrale ou encore en génomique pour étudier les distances favorisées ou évitées entre deux motifs le long du génome. Dans ce cadre, nous introduisons une fonction dite de reproduction qui permet de quantifier les positions préférentielles des motifs et qui peut être modélisée par l'intensité d'un processus de Poisson. Dans un premier temps, nous nous intéressons à l'estimation de cette fonction que l'on suppose très localisée. Nous proposons une procédure d'estimation adaptative par seuillage de coefficients d'ondelettes qui est optimale des points de vue oracle et minimax. Des simulations et une application en génomique sur des données réelles provenant de la bactérie *E. coli* nous permettent de montrer le bon comportement pratique de notre procédure. Puis, nous traitons les problèmes de test associés qui consistent à tester la nullité de la fonction de reproduction. Pour cela, nous construisons une procédure de test optimale du point de vue minimax sur des espaces de Besov faibles, qui a également montré ses performances du point de vue pratique. Enfin, nous prolongeons ces travaux par l'étude d'une version discrète en grande dimension du modèle précédent en proposant une procédure adaptative de type Lasso.

**Mots-clés :** Processus de Poisson, estimation et tests adaptatifs, seuillage de coefficients d'ondelettes, inégalités oracle,  $U$ -statistiques, vitesse de séparation uniforme, modèle d'interactions, processus de Hawkes, espaces de Besov, Lasso.

## Adaptive nonparametric inference for Poissonian interactions

**Abstract:** The subject of this thesis is the study of some adaptive nonparametric statistical problems in the framework of a Poisson interactions model. Such models are used, for instance, in neurosciences to analyze interactions between two neurons through their spikes emission during the recording of the brain activity or in genomics to study favored or avoided distances between two motifs along a genome. In this setting, we naturally introduce a so-called reproduction function that allows to quantify the favored positions of the motifs and which is considered as the intensity of a Poisson process. Our first interest is the estimation of this function assumed to be well localized. We propose a data-driven wavelet thresholding estimation procedure that is optimal from oracle and minimax points of view. Simulations and an application to genomic data from the bacterium *E. coli* allow us to show the good practical behavior of our procedure. Then, we deal with associated problems on tests which consist in testing the nullity of the reproduction function. For this purpose, we build a minimax optimal testing procedure on weak Besov spaces and we provide some simulations showing good practical performances of our procedure. Finally, we extend this work with the study of a high-dimensional discrete setting of our previous model by proposing an adaptive Lasso-type procedure.

**Keywords:** Poisson process, adaptive estimation and tests, wavelet thresholding rules, oracle inequalities,  $U$ -statistics, uniform separation rate, interactions model, Hawkes processes, Besov spaces, Lasso.



## Remerciements

Comme le veut la tradition, je m'attelle avec le plus grand plaisir à l'écriture de mes remerciements afin d'exprimer toute ma reconnaissance aux personnes qui ont influencé de près ou de loin ma thèse.

Mes premiers remerciements s'adressent à mes directeurs de thèse, Patricia Reynaud-Bouret et Vincent Rivoirard, pour leur encadrement exceptionnel et parfait. Patricia et Vincent, je vous remercie très chaleureusement pour votre confiance en me proposant un sujet de recherche aussi riche et passionnant, vos nombreux conseils, votre disponibilité malgré vos agendas bien chargés et la distance géographique, vos encouragements dans les périodes de doute et votre soutien permanent. Merci de m'avoir fait découvrir le monde de la recherche et de m'avoir transmis une infime part de votre grande culture mathématique à travers votre enthousiasme, votre esprit curieux, votre expérience, votre intuition et votre rigueur. Je suis honorée et fière d'être votre première thésarde puisque, ensemble ou séparément, vous êtes des directeurs exemplaires aux innombrables qualités scientifiques mais aussi humaines.

J'exprime toute ma reconnaissance à mes rapporteurs, Richard Nickl et Stéphane Robin, pour avoir consacré de leur temps à la lecture de ce manuscrit. Richard, thank you for the meticulous attention you have brought to the reading of this manuscript, your consideration and your presence at my PhD defense in France. Stéphane, merci pour ta sympathie et l'intérêt que tu as manifesté pour mon travail tout au long de ma thèse. Je tiens également à remercier Béatrice Laurent-Bonneau et Pascal Massart de me faire l'honneur de participer à mon jury. Pascal, je te remercie aussi de m'avoir fait découvrir l'univers des statistiques et de m'avoir orientée vers l'équipe de choc « Patricia et Vincent » pour mon stage de Master 2.

Une partie de ma thèse a été enrichie par une collaboration avec Christine Tuleau-Malot. Christine, je te remercie pour ta gentillesse, ta bonne humeur, ta persévérance, ton soutien et ta présence à ma soutenance. J'en profite pour remercier le Laboratoire J. A. Dieudonné pour leur accueil constant et sympathique lors de mes séjours (professionnels!) à Nice. This thesis also gave birth to a recent collaboration with Rebecca Willett and I thank her. Merci à Gwendal Restoux, Karine Tribouley, Frédérique Letué, Robin Ryder, Xavier Gendre, Clément Marteau, Cathy Maugis et Tristan Mary-Huard de m'avoir offert l'opportunité de diffuser mes travaux lors de séminaires.

Je souhaite témoigner ma reconnaissance aux enseignants (trop nombreux pour tous les citer sans en oublier ...) de mon cursus scolaire et universitaire qui m'ont permis d'accroître mon intérêt pour les mathématiques mais aussi à prendre un peu plus confiance en moi. J'ai eu la chance de poursuivre mes études doctorales au sein du Laboratoire de Mathématiques d'Orsay et plus particulièrement de l'Équipe Probabilités et Statistiques où j'ai pu côtoyer de nombreux chercheurs talentueux à qui je voue une gratitude éternelle pour leur savoir-faire professionnel, leurs conseils et leur amitié. Un merci spécial à Nathalie Castelle, Christine Kéribin et Marie-Anne Poursat, auprès desquelles j'ai effectué mes premiers pas d'enseignante. À leur côté et dans la convivialité j'ai beaucoup appris sur l'enseignement des statistiques, en particulier dans la filière Ingénierie Mathématique et cela a confirmé mon plaisir d'enseigner et de transmettre mes connaissances. Merci à Catherine Ardin, Christine Bailleul, Pascale Starck, Françoise Vasseur et un grand merci à Valérie Blandin-Lavigne pour leur aide, leur efficacité et leur gentillesse en me rendant la vie plus simple et plus agréable lors des nombreuses tâches administratives.

La vie de doctorante à Orsay est ponctuée par des moments de la vie de tous les jours tels que des discussions philosophiques sur tout et rien à la fois, des pauses-café qui s'éternisent ou des trajets en RER B interminables. Pour tous les bons moments partagés en cette période de dur labeur, je souhaite remercier tous les doctorants et en particulier : mes amis du bureau 114 Caroline, Lionel, Lucie et Thierry (114 en force!), mes compères Aurélien et Olivier, Arthur, Clément, Nina, Pierre-Antoine, Tristan et Vincent pour leur bonne humeur et les anciens Eddy, Igor, Jérémie, Maud, Pierre, Ramla, Robin et Sébastien que j'essaye d'imiter au mieux. Les nombreux séminaires et colloques auxquels j'ai participé m'ont permis de rencontrer ou de revoir des jeunes statisticiens. Un grand merci à Aurélie, Benjamin, Cyrielle, Florian, Gaëlle, Sarah et Sylvain pour les bons moments passés à Paris ou ailleurs (Marseille, Fréjus, Tunis, Aussois, Toulouse, ...). Et un merci spécial à Christophe, ami statisticien de longue date avec qui j'ai affronté les obstacles de la thèse et pris grand plaisir à découvrir les joies de la recherche.

Je tiens à remercier mes chers amis parisiens : Anne-Claire (amie chercheuse qui m'a toujours soutenue), Paul (qui a préféré le privé à la recherche), Geoffrey (je ne voulais pas faire une case spéciale Nord!), Matthias et Mélanie (qui eux ont préféré l'enseignement en collège et lycée ... bon courage!) et Croissant et Flory (à quand notre prochain voyage tous les 4?). J'ai toujours plaisir de rentrer au Mans pour boire un verre avec Marion, Romain et Stéphane. Et je remercie infiniment mes amies de cœur, pour certaines depuis toutes petites, qui m'ont toujours encouragée et soutenue : Ninie, Liliq, Yso, Aurore et Cécile.

Je remercie ma famille et ma belle-famille qui m'ont soutenue et nourrie de bien des façons. Merci à Christian, Kiki, Maria, Christiane et Jean-Philippe d'avoir fait tout ce trajet pour assister à ma soutenance. Merci à Papy, Mémé et Pépé pour leur affection sans faille. Merci pour tout à mon grand frère Julien, à ma petite sœur Pauline et à mes parents que j'aime. Les moments passés tous ensemble, même s'ils se font de plus en plus rares, riment toujours avec bonheur, joie et rires. Merci Maman et Papa pour ce délicieux pot! Pauline, je suis triste que tu ne sois pas là le jour de ma soutenance mais c'est pour la bonne cause : amuse-toi bien à Kpalimé au Togo avec les chèvres et les moutons! Et merci à Toundra et Step pour leurs ronrons.

Enfin, j'adresse mes plus tendres remerciements à Jérémy. Je te remercie mille fois pour ton amour infini, ton inébranlable confiance en moi et ton soutien inconditionnel. Merci de me supporter lorsque je te parle de mes petits « Poisson », merci pour tout ...

*À Marc et Alice*





# Table des matières

- 1 Introduction** **11**
- 1.1 Inférence non-paramétrique pour un processus de Poisson . . . . . 11
- 1.2 Biologie et interactions . . . . . 14
- 1.3 Interactions poissoniennes et processus de Hawkes . . . . . 17
- 1.4 Vue d'ensemble de nos résultats . . . . . 22
- 1.5 Perspectives . . . . . 34
  
- 2 Mathematical tools** **37**
- 2.1 Wavelet thresholding on classical and weak Besov spaces . . . . . 37
- 2.2 An introduction to the theory of  $U$ -statistics . . . . . 41
  
- 3 Adaptive estimation in a Poissonian interactions model** **45**
- 3.1 Introduction . . . . . 46
- 3.2 Estimation procedure . . . . . 49
- 3.3 Implementation . . . . . 53
- 3.4 Application to genomic data . . . . . 60
- 3.5 Conclusion . . . . . 63
- 3.6 Proofs . . . . . 65
- 3.A Supplement: Proof of Theorem 3.2 . . . . . 71
  
- 4 Detection of dependence in a Poissonian interactions model** **97**
- 4.1 Introduction . . . . . 98
- 4.2 Description of our testing procedure . . . . . 102
- 4.3 Main theoretical results . . . . . 104
- 4.4 Simulation study . . . . . 107
- 4.5 Conclusion . . . . . 111
- 4.6 Proofs . . . . . 111
  
- 5 Lasso estimation for Poissonian interactions on the circle** **127**
- 5.1 Introduction . . . . . 128
- 5.2 Formulation of the problem . . . . . 129
- 5.3 Isometry properties . . . . . 130
- 5.4 Lasso-type procedure . . . . . 132
- 5.5 Perspectives . . . . . 134
- 5.6 Proofs . . . . . 134

<b>A Programs coded in Scilab</b>	<b>153</b>
A.1 Implementation of the estimation procedure of Chapter 3 . . . . .	153
A.2 Implementation of the testing procedure of Chapter 4 . . . . .	160
<b>Bibliographie</b>	<b>165</b>

# Chapitre 1

## Introduction

Ce chapitre est une introduction au domaine statistique dans lequel s'inscrit cette thèse et une exposition de ses principaux résultats. Après avoir présenté brièvement le cadre, qui est celui des processus de Poisson, nous introduisons notre modèle, qui peut être vu comme une alternative aux processus de Hawkes. Puis nous résumons les contributions principales des Chapitres 3 à 5 et concluons en présentant les perspectives de recherche faisant suite aux travaux de cette thèse.

### Sommaire

---

<b>1.1</b>	<b>Inférence non-paramétrique pour un processus de Poisson . . . . .</b>	<b>11</b>
1.1.1	Processus de Poisson . . . . .	11
1.1.2	Estimation de l'intensité d'un processus de Poisson . . . . .	12
1.1.3	Test d'hypothèses sur l'intensité d'un processus de Poisson . . . . .	13
<b>1.2</b>	<b>Biologie et interactions . . . . .</b>	<b>14</b>
1.2.1	Contexte de la génomique . . . . .	14
1.2.2	Contexte des neurosciences . . . . .	16
<b>1.3</b>	<b>Interactions poissoniennes et processus de Hawkes . . . . .</b>	<b>17</b>
1.3.1	Processus de Hawkes bivarié . . . . .	17
1.3.2	Notre modèle d'interactions poissoniennes . . . . .	20
<b>1.4</b>	<b>Vue d'ensemble de nos résultats . . . . .</b>	<b>22</b>
1.4.1	Procédure de test adaptatif . . . . .	23
1.4.2	Procédure d'estimation adaptative . . . . .	27
1.4.3	Estimation Lasso et interactions poissoniennes sur le cercle . . . . .	31
<b>1.5</b>	<b>Perspectives . . . . .</b>	<b>34</b>

---

## 1.1 Inférence non-paramétrique pour un processus de Poisson

### 1.1.1 Processus de Poisson

Le processus de Poisson est l'exemple le plus simple des processus ponctuels. Il constitue la brique essentielle pour bâtir des processus plus complexes. On rappelle qu'un processus ponctuel  $N$  est un ensemble aléatoire au plus dénombrable de points d'un espace mesurable  $\mathbb{X}$ . On note  $dN$  la mesure ponctuelle associée à  $N$  définie par  $dN = \sum_{X \in N} \delta_X$ , où  $\delta_X$  est la mesure de Dirac au point  $X$ .

**Définition 1.1.** Soit  $(\mathbb{X}, \mathcal{X})$  un espace mesurable. Soit  $N$  un processus ponctuel. On dit que  $N$  est un processus de Poisson sur  $(\mathbb{X}, \mathcal{X})$  si

- pour tout  $A \in \mathcal{X}$ , le nombre de points de  $N$  se trouvant dans l'ensemble  $A$  est une variable aléatoire, notée  $N_A$ , qui suit une loi de Poisson de paramètre  $\mu(A)$ , où  $\mu : \mathcal{X} \rightarrow \mathbb{R}_+$  est une mesure sans atome, appelée mesure moyenne de  $N$ , et
- pour toute famille finie d'ensembles disjoints  $A_1, \dots, A_n$  de  $\mathcal{X}$ , les variables aléatoires  $N_{A_1}, \dots, N_{A_n}$  sont indépendantes.

Dans le cas où  $\mathbb{X} = \mathbb{R}^d$  et  $\mu$  absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , on peut écrire  $d\mu(x) = h(x) dx$  et  $h$  est appelée l'**intensité du processus de Poisson**  $N$  ( $h$  est une fonction déterministe). Si  $h$  est constante, on dit que  $N$  est un processus de Poisson homogène, c'est donc un processus stationnaire. Sinon, on dit qu'il s'agit d'un processus de Poisson inhomogène.

L'étude des processus de Poisson a fait l'objet d'une littérature considérable. Les livres de Kingman [67] et de Daley et Vere-Jones [29] en donnent une bonne vue d'ensemble. Les processus de Poisson peuvent modéliser un grand nombre de situations. En effet, ils permettent de modéliser mathématiquement et simplement des phénomènes caractérisés par des événements aléatoires très localisés qui peuvent donc être vus comme des points. À titre d'exemple, on peut considérer les phénomènes temporels suivants : l'émission de particules radioactives, les appels dans un central téléphonique, les arrivées de clients devant un guichet, ou bien les buts marqués lors d'un match de football. La version spatiale des processus de Poisson peut quant à elle modéliser par exemple la répartition d'arbres dans une forêt tropicale, les épicentres de tremblements de terre, ou bien les positions d'étoiles d'une nébuleuse. L'intensité qui gouverne chacune de ces modélisations poissonniennes, dans le temps ou l'espace, peut tout à fait être une fonction discontinue, irrégulière et présenter des pics très localisés.

### 1.1.2 Estimation de l'intensité d'un processus de Poisson

Il est naturel de chercher à estimer l'intensité  $h$  à partir de  $n$  réalisations indépendantes du processus de Poisson  $N$ . On mesure la qualité d'un estimateur  $\hat{h}$  de  $h$ , construit à partir de ces  $n$  réalisations, à travers son risque quadratique  $\mathbb{E}(\|h - \hat{h}\|_2^2)$ . En effet, on espère que le risque de  $\hat{h}$  décroisse lorsque le nombre de réalisations  $n$  augmente (plus on dispose d'information, mieux on estime la fonction). Plus précisément, étant donné  $\mathcal{F}$  un ensemble de fonctions possibles pour  $h$ , on dit que l'estimateur  $\hat{h}$  atteint la vitesse de convergence  $v_n$  s'il existe une constante  $C$  strictement positive telle que

$$\sup_{h \in \mathcal{F}} \mathbb{E}(\|h - \hat{h}\|_2^2) \leq C v_n.$$

On cherche alors à construire un estimateur  $\hat{h}$  de  $h$  optimal du **point de vue minimax** sur une classe de fonctions  $\mathcal{F}$ , c'est-à-dire ayant une vitesse de convergence sur  $\mathcal{F}$  qui coïncide, à une perte logarithmique près, avec les vitesses minimax sur  $\mathcal{F}$  :

$$R(\mathcal{F}) = \inf_{\hat{h}} \sup_{h \in \mathcal{F}} \mathbb{E}(\|h - \hat{h}\|_2^2),$$

où l'infimum est pris sur tous les estimateurs possibles de  $h$ . Ce risque représente le risque du meilleur estimateur possible pour la pire fonction à estimer dans  $\mathcal{F}$ . Dans cette thèse, les classes fonctionnelles considérées sont des boules de l'espace de Besov  $\mathcal{B}_{2,\infty}^\delta$  avec  $\delta > 0$  (voir Triebel [107]

ou encore DeVore et Lorentz [34]) et leurs versions faibles  $\mathcal{W}_p^*$  avec  $p > 0$  (voir Cohen *et al.* [24] ou encore Rivoirard [94]). Une succincte description de ces espaces est proposée à la Section 2.1.4 du Chapitre 2.

L'estimation non-paramétrique adaptative a pour but d'estimer  $h$  en faisant le moins possible d'hypothèses sur  $h$ , en particulier sur sa régularité qui apparaît dans les vitesses de convergence classiques. Pour cela, de nouvelles procédures dites adaptatives se sont développées dans les années 90 permettant de trouver un estimateur de  $h$  construit uniquement à partir des données (« data-driven » en anglais), c'est-à-dire ne nécessitant aucune connaissance au préalable de la régularité de  $h$ . Ces procédures ont l'avantage de s'adapter à cette régularité inconnue, tout en étant aussi performantes que les méthodes connaissant précisément la régularité de  $h$ , en atteignant les mêmes vitesses minimax mais parfois à une perte logarithmique près. Les trois grandes familles de procédures adaptatives sont les procédures par pénalisation (voir Tibshirani [105], Candès et Tao [19] pour des méthodes pénalisées  $\ell_1$ , Birgé et Massart [11], Barron *et al.* [7], Massart [77] pour des méthodes par sélection de modèles (pénalisation  $\ell_0$ ) ou encore Tsybakov [108]), les estimateurs à noyau avec sélection de fenêtre (voir Lepski [70, 71, 72] ou encore Goldenshluger et Lepski [49]) et les procédures par seuillage de coefficients d'ondelettes (voir Donoho et Johnstone [37, 38], Donoho *et al.* [41] ou encore Härdle *et al.* [57]). Ce sont ces dernières qui sont au coeur de cette thèse comme on le verra à la Section 1.4.2. On peut alors considérer un autre critère théorique d'optimalité d'estimation, souvent mis en avant dans la littérature tout comme le point de vue minimax, qui est l'optimalité du **point de vue oracle**. Cela consiste à établir une inégalité de type oracle qui prouve que l'estimateur a un risque aussi bon que celui de l'oracle (à quelques pertes éventuelles près), où l'oracle représente le meilleur « estimateur » parmi une famille de procédures, si l'on connaît la fonction  $h$  à estimer.

En remarquant simplement que conditionnellement à l'événement «  $N_{\mathbb{R}^d} = n$  » les points du processus  $N$  ont la même loi qu'un  $n$ -échantillon de densité  $f = h / \int_{\mathbb{R}^d} h(x) dx$ , estimer l'intensité d'un processus de Poisson ou estimer une densité est un travail similaire. C'est pourquoi de nombreuses procédures d'estimation non-paramétrique d'une densité  $f$  ont été transcrites dans le contexte poissonien. Citons par exemple les travaux de Rudemo [100], Donoho [35], Kolarczyk [68], Cavalier et Koo [22], Reynaud-Bouret [90], Willett et Nowak [114], Zhang *et al.* [116] et Reynaud-Bouret et Rivoirard [91]. Reynaud-Bouret et Rivoirard [91] ont proposé une procédure par seuillage qui est « data-driven », optimale des points de vue oracle et minimax et qui, contrairement aux autres, ne nécessite pas de supposer que le support de  $h$  est connu ou même borné.

### 1.1.3 Test d'hypothèses sur l'intensité d'un processus de Poisson

On cherche à tester des hypothèses sur l'intensité  $h$ , à partir de l'observation du processus de Poisson  $N$ . Un test de l'hypothèse nulle  $\mathcal{H}_0 : \langle h \in \mathcal{F}_0 \rangle$  contre l'alternative  $\mathcal{H}_1 : \langle h \notin \mathcal{F}_0 \rangle$  est une règle de décision  $\Phi$ , qui dépend de l'observation  $N$ , à valeurs dans  $\{0, 1\}$ . Par convention, on accepte l'hypothèse nulle  $\mathcal{H}_0$  lorsque  $\Phi = 0$ , sinon on rejette  $\mathcal{H}_0$ . Étant donné  $\alpha \in ]0, 1[$  fixé, on dit que le test  $\Phi$  est de niveau  $\alpha$  si  $\sup_{h \in \mathcal{F}_0} \mathbb{P}_h(\Phi = 1) \leq \alpha$ , où  $\mathbb{P}_h$  est la loi du processus de Poisson  $N$  d'intensité  $h$  (le niveau de test peut être vu comme le risque maximal que l'on accepte de prendre en rejetant l'hypothèse nulle  $\mathcal{H}_0$  à tort). Ensuite, on mesure la qualité du test  $\Phi$  à travers sa fonction puissance  $h \notin \mathcal{F}_0 \mapsto \mathbb{P}_h(\Phi = 1)$ , c'est-à-dire la capacité du test à détecter les alternatives. Plus précisément, étant donné  $\beta \in ]0, 1[$  fixé (petit de préférence) et une classe de

fonctions  $\mathcal{F}_1$ , on définit la vitesse de séparation du test  $\Phi$  sur  $\mathcal{F}_1$  (voir Baraud [4]) :

$$\rho(\Phi, \mathcal{F}_1, \beta) = \inf \left\{ \rho > 0 : \inf_{h \in \mathcal{F}_1, d(h, \mathcal{F}_0) > \rho} \mathbb{P}_h(\Phi = 1) \geq 1 - \beta \right\},$$

où  $d$  est une distance sur les classes de fonctions de  $\mathbb{L}_2$ . Comme pour l'estimation, on cherche donc à construire un test  $\Phi$  de niveau  $\alpha$  qui a une vitesse de séparation sur une classe fonctionnelle  $\mathcal{F}_1$  (typiquement des espaces de Besov classiques ou faibles) du même ordre (à un possible terme de perte près) que la vitesse minimax de test sur  $\mathcal{F}_1$  définie par

$$\underline{\rho}(\mathcal{F}_1, \alpha, \beta) = \inf_{\Phi_\alpha} \rho(\Phi_\alpha, \mathcal{F}_1, \beta),$$

où l'infimum est pris sur tous les tests  $\Phi_\alpha$  de niveau  $\alpha$ . Finalement, les critères d'optimalité de test d'hypothèses sont les analogues de ceux adoptés pour l'estimation, avec la notion de vitesse de séparation de test présentée ci-dessus.

À la suite du développement des procédures adaptatives d'estimation, se sont naturellement déployées des procédures adaptatives de test qui ont pour but de tester une hypothèse sur l'intensité  $h$  en faisant le moins possible d'hypothèses sur les alternatives, en particulier sur leur régularité qui apparaît dans les vitesses minimax de test classiques, tout en restant performantes. Pour construire ces tests, les mêmes méthodes que celles utilisées en estimation et les problèmes associés entrent en jeu.

Donnons à présent quelques exemples de tests d'hypothèses sur l'intensité  $h$  d'un processus de Poisson. Dans un premier temps, on peut s'intéresser à tester l'homogénéité du processus. Un large éventail d'articles, à la fois d'un point de vue théorique et d'un point de vue pratique, traite de ce problème (voir Bain *et al.* [3], Cohen et Sackrowitz [25] ou encore Bhattacharjee *et al.* [10] où les alternatives sont des fonctions monotones). Pour ce même problème, Fromont *et al.* [45] ont proposé un test adaptatif au sens minimax, basé sur des méthodes de sélection de modèles et de seuillage. Un autre problème est celui de tester l'hypothèse nulle simple suivante « un processus ponctuel est un processus de Poisson avec une intensité donnée ». Dans le cadre paramétrique (ou semi-paramétrique), on peut citer les travaux de Fazli et Kutoyants [44] où l'alternative est aussi un processus de Poisson avec une intensité donnée et de Dachian et Kutoyants [28] où les alternatives sont des processus ponctuels auto-excitants. Dans le cadre non-paramétrique, Ingster et Kutoyants [63], qui considèrent des alternatives qui sont également des processus de Poisson avec des intensités non-paramétriques dans des classes fonctionnelles de type boule de Besov ou boule de Sobolev, ont prouvé que leur procédure était adaptative au sens minimax sur ces classes d'alternatives. On peut également s'intéresser à tester l'égalité des intensités de deux processus de Poisson indépendants. Ce problème classique a par exemple été étudié par Fromont *et al.* [46] en proposant des procédures adaptatives de tests multiples au sens minimax sur des classes d'alternatives basées, en particulier, sur les espaces de Besov classiques et faibles.

## 1.2 Biologie et interactions

### 1.2.1 Contexte de la génomique

Le génome est l'ensemble du matériel génétique d'un individu ou d'une espèce codé dans son ADN (à l'exception de certains virus dont le génome est porté par des molécules d'ARN). Le génome est souvent comparé à une encyclopédie dont les différents volumes seraient les chromosomes. Les gènes seraient les phrases contenues dans ces volumes et ces phrases seraient écrites

dans l'alphabet  $\mathcal{A} = \{a, c, g, t\}$  où les lettres correspondent aux quatre bases (ou nucléotides) : adénine, cytosine, guanine et thymine. La taille du génome se mesure en nombre de bases. La plupart du temps, on parle de paire de bases, puisque la majorité des génomes est constituée de doubles brins d'ADN (ou bien d'ARN) se faisant face et formant une double hélice (avec une complémentarité des nucléotides : en face d'une adénine, il y a toujours une thymine et inversement ; en face d'une cytosine, il y a toujours une guanine et inversement), ce qui confère la structure 3D de l'ADN. Les génomes sont constitués de régions codantes qui correspondent aux gènes et de régions non-codantes qui sont constituées des segments entre les gènes mais aussi de portions de gène. Le séquençage de l'ADN (qui date des années 70) permet d'établir l'enchaînement des nucléotides des brins d'ADN, afin de cartographier le génome.

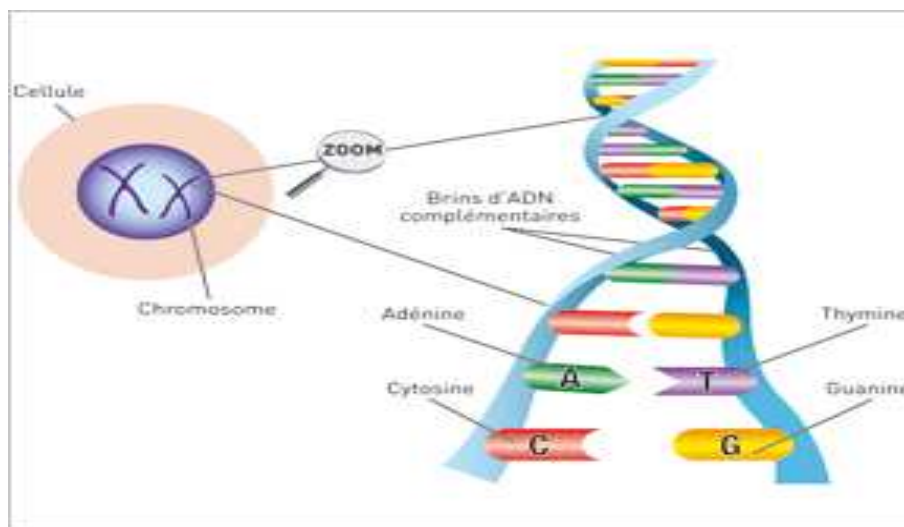


FIGURE 1.1 – Structure de la molécule d'ADN.

Grâce aux progrès de la biologie moléculaire mais aussi de la (bio)informatique ces dernières décennies, les laboratoires scientifiques ont à présent un accès facilité à une multitude de séquences d'ADN, provenant de divers espèces ou individus (par exemple, le génome humain a été complètement séquencé il y a une dizaine d'années). Le (bio)statisticien joue un rôle primordial dans l'analyse de ces séquences et par conséquent, dans l'évolution de la biologie moléculaire.

Les travaux de Schbath et ses coauteurs (par exemple, [89, 96, 103]) concernent, entre autres, la distribution des motifs le long des séquences d'ADN. Un motif est défini par une courte suite de lettres (généralement pas plus d'une quinzaine) dans l'alphabet  $\mathcal{A}$  et on s'intéresse plus particulièrement aux motifs dits fonctionnels (appelés simplement motifs par la suite), c'est-à-dire un motif dont les occurrences sur le génome seront reconnues par une protéine qui se fixera alors sur l'ADN pour entrer en action. Les occurrences d'un motif d'intérêt sont en général soit très fréquentes, soit des événements rares, le long des séquences d'ADN. On s'intéresse ici en particulier à ces motifs dits rares. Ainsi, et comme les génomes sont de longueur 1 million de bases environ, les occurrences de ce motif le long du génome peuvent être modélisées par un processus de Poisson homogène.

Dans cette thèse, on s'intéresse en particulier à l'action ou à l'influence réciproque entre deux motifs donnés le long du génome. Le but est alors de modéliser cette dépendance entre les motifs en vue d'identifier des distances favorisées ou évitées entre eux, suggérant de possibles



interactions à un niveau moléculaire, comme des interactions entre l'ADN et les protéines qui sont à l'origine de processus biologiques tels que la transcription de gène ou la recombinaison, la réparation ou la réplication d'ADN, etc. Gusto et Schbath [54] et Reynaud-Bouret et Schbath [93] ont étudié ce problème en utilisant un processus de Hawkes (défini à la Section 1.3.1.1). Le modèle d'interactions poissonniennes décrit à la Section 1.3.2.1 permet de modéliser les éventuelles interactions entre deux motifs, où l'un des motifs serait là indépendamment de l'autre (via un processus de Poisson homogène par exemple) et l'autre ne serait que le résultat de la présence du premier (via un processus de Poisson dont l'intensité dépend des positions du premier motif).

Mentionnons qu'une application sur des données réelles provenant de la bactérie *Escherichia coli* est donnée à la Section 1.4.2.4 (ou à la Section 3.4 du Chapitre 3 pour une étude détaillée). Cette application a pour but d'analyser l'influence entre les gènes et les sites promoteurs (plus précisément, le motif « tataat ») le long du génome d'*E. coli*.

Le lecteur intéressé par un aperçu succinct des enjeux des statistiques appliquées à la génomique pourra se reporter au n° 130 de la Gazette des Mathématiciens [86].

### 1.2.2 Contexte des neurosciences

Un neurone est une cellule excitable constituant l'unité fonctionnelle de base du système nerveux. Les neurones assurent la transmission d'un signal bioélectrique appelé potentiel d'action (« spike » en anglais). Les neurones ont deux propriétés physiologiques : l'excitabilité, c'est-à-dire la capacité de répondre aux stimulations et de convertir celles-ci en impulsions nerveuses, et la conductivité, c'est-à-dire la capacité de transmettre les impulsions. Le potentiel d'action correspond à une dépolarisation transitoire, locale, brève et stéréotypée de la membrane plasmique des neurones. La genèse du potentiel d'action a lieu au niveau du cône d'émergence, à la base du corps cellulaire du neurone qui fait la sommation des potentiels gradués provenant des synapses situées le long des dendrites et sur le corps cellulaire. Si cette somme dépasse un certain seuil d'excitabilité du neurone, un potentiel d'action est créé et se propage, sans perte d'amplitude, le long de l'axone. Ainsi, les neurones (au nombre estimé de 86 à 100 milliards dans le cerveau humain) peuvent communiquer entre eux à travers l'émission des potentiels d'action et l'activité neuronale peut donc être réduite aux différents potentiels d'action émis par chacun des neurones. Cette activité peut se résumer par la collection des temps de potentiels d'action, appelée « train de spikes », qui est alors assimilée à un processus ponctuel temporel.

Les travaux de Grammont et Riehle [50] et de Grün *et al.* [53] portant sur l'étude d'une paire de neurones en particulier ont mis en évidence un phénomène appelé la synchronisation : la présence d'un spike sur le train de spikes d'un neurone peut participer à l'apparition d'un spike, avec un délai, sur le train de spikes du second neurone. En effet, pour atteindre le seuil d'excitabilité qui permettra l'émission d'un spike, deux phénomènes sont possibles : soit la fréquence des spikes reçus augmente, soit les potentiels d'action arrivent en nombre moindre, mais au même moment, ce qui est moins coûteux en énergie. Les neurobiologistes s'intéressent donc à la détection de ce phénomène de synchronisation ; autrement dit, ils souhaitent savoir si les deux neurones considérés dans l'étude évoluent indépendamment ou non. Le modèle d'interactions poissonniennes décrit à la Section 1.3.2.1 permet de modéliser les éventuelles interactions entre deux neurones. À noter que Krumin *et al.* [69] et Pernice *et al.* [84, 85] ont récemment étudié ce type de problème en utilisant un modèle de Hawkes.

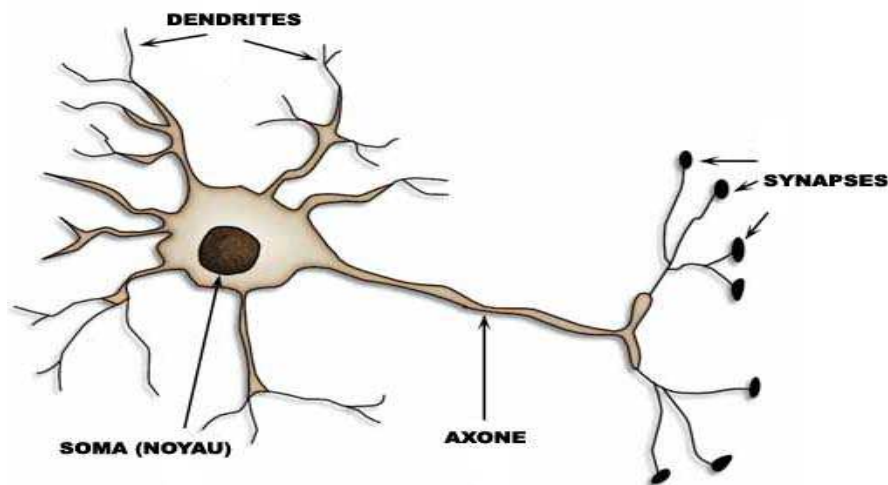


FIGURE 1.2 – Schéma d'un neurone.

## 1.3 Interactions poissoniennes et processus de Hawkes

### 1.3.1 Processus de Hawkes bivarié

#### 1.3.1.1 Définition et remarques

Le processus de Hawkes, introduit par Hawkes [59] en 1971, est très utilisé en pratique pour modéliser des phénomènes issus de la sismologie ou de l'épidémiologie, ou plus récemment, des phénomènes issus de la génomique ou des neurosciences comme on l'a évoqué précédemment. Un processus de Hawkes bivarié est défini par deux processus ponctuels  $(N_t)_{t \in \mathbb{R}}$  et  $(N'_t)_{t \in \mathbb{R}}$  d'intensité respective conditionnellement au passé

$$\lambda(t) = \nu + \int_{-\infty}^{t^-} g(t-u) dN_u + \int_{-\infty}^{t^-} h(t-u) dN'_u \quad (1.1)$$

et

$$\lambda'(t) = \nu' + \int_{-\infty}^{t^-} g'(t-u) dN'_u + \int_{-\infty}^{t^-} h'(t-u) dN_u, \quad (1.2)$$

où  $\nu$  et  $\nu'$  sont des constantes positives et  $g, h$  et  $g', h'$  sont des fonctions positives à support dans  $\mathbb{R}_+$ . On suppose que le rayon spectral de  $\Lambda = \begin{pmatrix} \int_0^{+\infty} g(u) du & \int_0^{+\infty} h(u) du \\ \int_0^{+\infty} h'(u) du & \int_0^{+\infty} g'(u) du \end{pmatrix}$  (c'est-à-dire la plus grande valeur propre de  $\Lambda$  en valeur absolue) est strictement inférieur à 1. Cette condition permet d'assurer l'existence d'une version stationnaire du processus de Hawkes défini ci-dessus (voir Brémaud et Massoulié [13]).

Pour décrire simplement le processus de Hawkes ainsi défini, considérons le contexte de la génomique présenté à la Section 1.2.1, à savoir l'étude de l'influence entre deux motifs donnés : motif 1 et motif 2, le long du génome. Interprétons en détails l'intensité  $\lambda(t)$  du processus  $N$  définie par (1.1). La droite réelle tient lieu de génome et  $N, N'$  et leurs points représentent respectivement le motif 1, le motif 2 et leurs occurrences (ou de manière analogue, dans le contexte des neurosciences, le neurone 1, le neurone 2 et leurs spikes, au cours de l'enregistrement de

l'activité neuronale). L'intensité  $\lambda(t)$  représente la probabilité d'avoir une occurrence du motif 1 à la position  $t$  le long du génome connaissant tout le passé, le passé étant les occurrences antérieures à la position  $t$  des motifs 1 et 2 :

$$\mathbb{P}(\text{occurrence du processus } N \in [t; t + dt] \text{ conditionnellement au passé avant } t) = \lambda(t) dt.$$

Autrement dit, une occurrence du motif 1 peut apparaître à la position  $t$  le long du génome :

- de manière **spontanée**, sans cause extérieure (via  $\nu$  dans (1.1)),
- par une influence des occurrences passées du motif 1 (via  $g$  dans (1.1)) traduisant un phénomène d'**auto-excitation** du motif 1 sur lui-même, ou
- par une influence des occurrences passées du motif 2 (via  $h$  dans (1.1)) traduisant un phénomène d'**interaction** du motif 2 sur le motif 1.

Puisque les fonctions  $g$  et  $h$  sont ici des fonctions positives, les influences sur l'apparition d'une occurrence du motif 1 sont positives et se quantifient selon l'ampleur de  $g$  et  $h$ , ce qui permet de définir la notion de distance favorisée. On pourrait considérer des fonctions  $g$  et  $h$  qui peuvent prendre des valeurs négatives (et on prendrait par conséquent la partie positive de (1.1) pour définir l'intensité). Ainsi lorsque  $g$  est négative, les influences sur l'apparition d'une occurrence du motif 1 sont négatives (on parle alors d'auto-inhibition au lieu d'auto-excitation) et se quantifient selon l'ampleur de  $|g|$ , ce qui permet de définir la notion de distance évitée.

À l'aide de la définition de la mesure ponctuelle associée au processus, on peut réécrire l'intensité (1.1) du processus de Hawkes comme suit :

$$\lambda(t) = \nu + \sum_{X \in N, X < t} g(t - X) + \sum_{X' \in N', X' < t} h(t - X').$$

Cela permet de mieux apprécier la contribution de chaque occurrence passée des motifs 1 et 2 à l'incrémentement de l'intensité. Ainsi, si on se donne une position  $t$  sur le génome, la distance  $d = t - X$ , où les  $X$  sont les occurrences passées du motif 1, est d'autant plus favorisée que  $g(d)$  est grande.

Pour plus de détails, on se référera par exemple au livre de Daley et Vere-Jones [29] et aux travaux de Brémaud et Massoulié [13, 14] pour l'aspect théorique des processus de Hawkes, aux travaux de Vere-Jones et Ozaki [113] pour une application en sismologie, de Mitchell et Cates [79] pour une application sur les interactions sociales, de Bacry *et al.* [2] pour une application en finance et aux travaux de Gusto et Schbath [54] et Carstensen *et al.* [21] pour l'aspect génomique. Le dernier article cité utilise le processus de Hawkes multivarié qui est une généralisation naturelle du processus de Hawkes bivarié.

À noter que l'estimation des fonctions  $g$  et  $h$  fait l'objet d'une littérature fournie dans le cadre paramétrique (voir par exemple Ozaki [83], Ogata et Akaike [82] ou encore Daley et Vere-Jones [29]).

Dans la suite, on décrit trois approches non-paramétriques :

- la méthode FADO, une méthode pénalisée de type AIC dans un modèle de Hawkes bivarié, proposée par Gusto et Schbath [54],
- une méthode d'estimation adaptative par sélection de modèles pour des processus de Hawkes univariés (i.e.  $h = 0$  dans (1.1)) proposée par Reynaud-Bouret et Schbath [93], et
- une procédure adaptative de type Lasso pour des processus de Hawkes multivariés proposée par Hansen *et al.* [56].

Précisons que les travaux de Gusto et Schbath [54] et de Reynaud-Bouret et Schbath [93] s'inscrivent dans le contexte de la génomique et s'intéressent plus particulièrement au problème de détection de distances favorisées ou évitées le long d'une séquence d'ADN et que les travaux de Hansen *et al.* [56] (article actuellement soumis) sont postérieurs à la majeure partie des résultats de cette thèse.

### 1.3.1.2 Inférence non-paramétrique pour un processus de Hawkes

Gusto et Schbath [54] considèrent le processus de Hawkes bivarié (défini par (1.1)). Leur procédure d'estimation des fonctions d'interaction  $g$  et  $h$ , par maximum de vraisemblance, s'effectue dans une base de splines à noeuds régulièrement espacés et sélectionne le nombre de noeuds par un critère de type AIC validé asymptotiquement.

La méthode FADO produit des estimateurs lisses et obtient de bons résultats. Cependant, elle présente quelques inconvénients. La méthode se comporte extrêmement mal pour des familles complexes de modèles, ce qui vient probablement du choix de la pénalité par un critère de type AIC. De plus, elle présente des problèmes de parcimonie (« sparsity » en anglais). En effet, dans le contexte de la génomique, s'il y a une interaction, disons à une distance de  $d$  bases, la fonction  $g$  ou  $h$  doit prendre de grandes valeurs aux alentours de  $d$  et, s'il n'y a pas de raison biologique pour qu'il y ait une autre interaction, cette fonction doit être nulle partout ailleurs. Ainsi, même si les oscillations de l'estimation par splines de cette fonction sont petites sur ces plages, l'estimée de la fonction n'est jamais vraiment nulle et cela pourrait entraîner des erreurs d'interprétation biologique.

Afin de résoudre les problèmes soulevés, Reynaud-Bouret et Schbath [93] ont proposé, dans le cadre des processus de Hawkes univariés, une méthode basée sur de la sélection de modèles pénalisée mais dans sa version non-asymptotique, afin d'obtenir un estimateur optimal du point de vue oracle et adaptatif au sens minimax sur certaines classes fonctionnelles. Ainsi, leur méthode solutionne le problème de parcimonie de la méthode FADO et permet de discuter pour la première fois les aspects minimax des processus de Hawkes. Cependant, en considérant un processus de Hawkes univarié, on ne prend pas en compte l'interaction avec un autre processus extérieur, bien que ce soit le but recherché à terme. De plus, la procédure pratique est très coûteuse en temps de calculs et engendre des problèmes de taille mémoire.

Ces deux procédures, avec leurs avantages et leurs inconvénients, sont à l'origine du modèle simple introduit à la section suivante, directement lié au problème de détection de distances favorisées ou évitées entre deux motifs le long d'une séquence d'ADN.

Se limiter aux interactions entre deux processus peut être restrictif pour la compréhension globale de certains phénomènes. Par exemple, en génomique on peut vouloir trouver les distances favorisées ou évitées entre plusieurs éléments régulateurs de la transcription (voir les données traitées par Carstensen *et al.* [21]) et en neurosciences il n'est pas rare d'enregistrer simultanément plus d'une centaine de neurones. Néanmoins, dans ce cadre multivarié, on espère qu'il y ait peu de fonctions d'interaction estimées comme non nulles et que de plus ces fonctions soient parcimonieuses (ce qui motive entre autres la construction de tests statistiques). Les deux méthodes citées ci-dessus ne sont donc pas envisageables dans ce contexte pour des problèmes soit de parcimonie, soit algorithmique. C'est pourquoi, dans le cadre des processus de Hawkes multivariés, Hansen *et al.* [56] ont récemment élaboré une procédure de type Lasso (voir Tibshirani [105] ou encore Section 1.4.3 et Chapitre 5) avec des poids adaptatifs dans la pénalité  $\ell_1$  (par exemple, voir Ber-

tin *et al.* [9]), calibrés par des inégalités de type Bernstein pour les martingales. En plus d'établir des inégalités oracles pour leur estimateur Lasso (sous certaines conditions), Hansen *et al.* [56] ont proposé des simulations (sur des données de type neuronal) qui mettent en avant les bonnes performances et la robustesse (par rapport à la durée de l'enregistrement) de la procédure pratique associée.

Mentionnons un dernier inconvénient du processus de Hawkes. Ce dernier impose que les fonctions  $g$  et  $h$  soient à support dans  $\mathbb{R}_+^*$ . Autrement dit, les occurrences d'un motif à une position sur le génome ne dépendent que des occurrences passées des deux motifs. La notion de passé, présent et futur est primordiale dans le cadre des processus de Hawkes, mais cet ordre temporel semble artificiel quand on parle de données génomiques même si le sens de lecture de l'ADN en impose un plus faible.

### 1.3.2 Notre modèle d'interactions poissonniennes

Cette section présente le modèle d'interactions poissonniennes sur lequel porte l'essentiel des travaux de cette thèse.

#### 1.3.2.1 Description du modèle

On se donne un réel  $T$  strictement positif, qui peut représenter la longueur du génome ou le temps d'enregistrement de l'activité neuronale et un entier naturel  $n$  non nul, qui peut représenter le nombre d'occurrences du premier motif d'intérêt dans le cadre génomique ou le nombre de spikes sur le premier neurone d'intérêt dans le cadre des neurosciences. On s'intéresse à deux événements modélisés par des processus ponctuels, pour lesquels on présuppose une interaction entre eux (voir Section 1.2) et on privilégie un sens d'interaction : un premier processus est là indépendamment du second et on explique la présence de ce second processus par le premier.

Le premier processus représente les « parents » et est modélisé par un  $n$ -échantillon  $U_1, \dots, U_n$  de variables aléatoires uniformes sur  $[0; T]$ . Ensuite, chaque parent  $U_i$  donne naissance indépendamment les uns des autres à des « enfants » selon un processus de Poisson  $N^i$  dont l'intensité est une fonction inconnue  $h$  recentrée sur le parent  $U_i$ . À cela s'ajoute un processus de Poisson homogène  $N^0$  d'intensité constante  $\nu$  représentant des enfants dits « orphelins ». On observe alors le processus de Poisson agrégé des enfants, qui représente le second processus :

$$N = \sum_{i=0}^n N^i \quad \text{dont l'intensité est donnée par la fonction} \quad t \mapsto \nu + \sum_{i=1}^n h(t - U_i). \quad (1.3)$$

Ainsi, si on peut distinguer les points qui sont des parents des points qui sont des enfants, on ne peut pas déterminer le parent de chaque enfant, ni savoir si un enfant est orphelin. La fonction  $h$  est naturellement appelée fonction de reproduction et notre but est alors d'estimer  $h$  ou de tester des hypothèses sur cette fonction, à partir de l'observation des parents  $U_1, \dots, U_n$  et de la réalisation du processus des enfants  $N$ .

On peut voir ce modèle comme un cas très particulier du modèle de Hawkes bivarié défini à la Section 1.3.1.1. En effet, avec  $g' = h' = 0$  dans (1.2), le processus  $N'$  est un processus de Poisson homogène et, conditionnellement au nombre  $n$  de points de  $N'$  sur  $[0; T]$ , cela revient à se donner un  $n$ -échantillon de variables uniformes qui correspond alors aux parents de notre modèle. Puis, le processus associé aux enfants est alors défini par (1.1) avec  $g = 0$  et  $h$  à support dans  $\mathbb{R}_+^*$ .

Contrairement aux processus de Hawkes, le support de la fonction d'interaction  $h$  dans notre modèle est tout  $\mathbb{R}$  et cela permet une dépendance entre les processus selon leurs événements passés mais aussi leurs événements présents et futurs.

### 1.3.2.2 Lien avec le contexte de la génomique et remarques

Dans le contexte de la génomique, on s'intéresse à l'influence du motif 2 sur le motif 1. On observe les occurrences de ces deux motifs le long du génome. Ainsi,  $n$  représente le nombre d'occurrences du motif 2 et  $T$  la longueur normalisée du génome considéré (le génome étant lu dans les deux sens de transcription, la longueur normalisée du génome est typiquement deux fois la longueur du génome).

Le choix de la modélisation des occurrences du motif 2 par des variables uniformes indépendantes n'est pas anodine et résulte de la constatation suivante. Les motifs d'intérêt le long du génome sont typiquement des événements rares et peuvent donc être modélisés par un processus de Poisson homogène  $N'$  sur  $[0; T]$  d'intensité constante  $\nu'$  (voir Section 1.2.1). Ainsi, conditionnellement à l'événement «  $N'_{[0;T]} = n$  », les points du processus  $N'$  (c'est-à-dire le motif 2 ou encore les parents) ont la même loi qu'un  $n$ -échantillon de variables aléatoires uniformes sur  $[0; T]$ . De plus, comme  $\mathbb{E}(N'_{[0;T]}) = \nu'T$  et que l'on s'est fixé le nombre  $n$  de parents en amont de l'étude, on a qu'avec grande probabilité,  $n$  et  $T$  sont du même ordre de grandeur. Le cas où  $n$  et  $T$  sont proportionnels semble être alors le cas générique.

Ensuite, on cherche à expliquer les occurrences du motif 1, les enfants, à partir des occurrences du motif 2, les parents, c'est-à-dire à l'aide d'un phénomène d'interaction des parents sur les enfants et également à l'aide d'un phénomène d'apparition spontanée. En effet, contrairement au modèle de Hawkes bivarié complet (voir (1.1)), dans notre modèle, chaque enfant provient soit d'un parent, soit est orphelin : il n'y a pas d'enfant qui peut également être un parent, c'est-à-dire on ne considère pas le phénomène d'auto-excitation des enfants sur eux-mêmes.

Quels sont les avantages et les inconvénients d'une telle modélisation ?

L'inconvénient majeur est surtout du point de vue biologique car notre modèle d'interactions poissoniennes est moins réaliste, en ne prenant pas en compte le phénomène d'auto-excitation et en ne reliant pas la présence des parents à celles des enfants (via (1.2) par exemple). Cela engendre une dissymétrie quant aux rôles des parents et des enfants. Il est donc important de bien définir quel motif sera le parent (ici le motif 2) et quel motif sera l'enfant (ici le motif 1). En outre, si on souhaite s'intéresser à l'influence du motif 1 sur le motif 2, on inversera les rôles parents/enfants dans le modèle et l'interaction du motif 1 sur le motif 2 sera mesurée grâce à une autre fonction de reproduction. Cependant, la procédure d'estimation de la fonction de reproduction  $h$  proposée au Chapitre 3 a permis de révéler les distances favorisées ou évitées entre les gènes et les sites promoteurs le long du génome de la bactérie *E. coli* auxquelles le biologiste pouvait s'attendre (voir Section 1.4.2.4 ou Section 3.4 du Chapitre 3). De plus, l'intensité  $h$  est à support dans tout  $\mathbb{R}$ , ou plus exactement dans  $[-A; A]$ , où  $A$  représente la mémoire maximale des occurrences le long des séquences d'ADN et peut être prise égale à 10000 bases (au-delà, la structure 3D de l'ADN viendrait interférer et les distances réelles physiques pourraient être plus courtes que les distances lues « en ligne » sur la séquence d'ADN). Dans le contexte des neurosciences, cette borne  $A$  représenterait le temps limite de synchronisation durant l'accomplissement d'une tâche, selon les neurobiologistes. Autrement dit, les occurrences futures du motif 2 peuvent influencer ce qui se passe dans le présent pour le motif 1. Cela permet donc de regarder en même temps

les deux directions : le passé et le futur (et ainsi de s'affranchir de l'hypothèse artificielle d'ordre temporel le long des séquences d'ADN).

D'autres avantages majeurs sont surtout liés à la parcimonie et à la mise en oeuvre algorithmique. En effet, on a développé au Chapitre 3 une procédure par seuillage de coefficients d'ondelettes, optimale et complètement « data-driven », pour le problème de l'estimation de cette fonction de reproduction  $h$  (voir Section 1.4.2). Ainsi, le seuillage permet de pallier les problèmes de parcimonie rencontrés par la méthode FADO de Gusto et Schbath [54] et l'algorithme qui découle de cette procédure (à savoir un algorithme de descente, voir Mallat [74]) est peu coûteux en temps de calculs, contrairement à celui de Reynaud-Bouret et Schbath [93].

Hors du contexte génomique, le découplage des paramètres  $n$  (le nombre de parents) et  $T$  (la longueur normalisée du génome étudié) provoque alors une double asymptotique qui n'est pas commune et engendrera des vitesses inhabituelles.

### 1.3.2.3 Lien avec le processus de Cox

Le processus de Cox (ou le double processus de Poisson stochastique), introduit pour la première fois par Cox [27] et très utilisé dans le domaine des mathématiques financières, de l'actuariat et de la médecine, est obtenu par randomisation de la mesure moyenne d'un processus de Poisson. Plus précisément, si on se donne  $(\mathbb{X}, \mathcal{X})$  un espace mesurable et  $\mu$  une mesure aléatoire sur  $\mathcal{X}$ , alors on dit qu'un processus ponctuel  $N$  est un processus de Cox gouverné par  $\mu$  lorsque, conditionnellement à  $\mu$ , les réalisations de  $N$  sont celles d'un processus de Poisson  $N(\cdot|\mu)$  sur  $\mathcal{X}$  de mesure moyenne  $\mu$  (voir [29]).

Dans le modèle d'interactions poissonniennes défini par (1.3), la fonction  $t \mapsto \nu + \sum_{i=1}^n h(t-U_i)$  est aléatoire (car dépend des parents  $U_1, \dots, U_n$ ) et peut être vue comme l'intensité d'un processus de Cox où les covariables seraient les parents  $U_1, \dots, U_n$ .

De nombreux travaux sur l'estimation de la mesure  $\mu$  d'un processus de Cox ont été effectués dans le cadre paramétrique (voir par exemple Andersen *et al.* [1] ou encore Møller *et al.* [80]). Dans le cadre non-paramétrique, citons par exemple Comte *et al.* [26] qui ont proposé un estimateur original de l'intensité conditionnelle d'un processus de Cox, construit à partir des méthodes de sélection de modèle, qui est optimal des points de vue oracle et minimax. Malheureusement, ces procédures engendrent de sérieuses difficultés pour l'implémentation algorithmique, comme on a pu déjà le souligner pour les travaux de Reynaud-Bouret et Schbath [93].

## 1.4 Vue d'ensemble de nos résultats

Nos travaux sont essentiellement centrés sur l'estimation de la fonction de reproduction  $h$  et la détection de dépendance dans le modèle d'interactions poissonniennes (1.3). Les procédures adaptatives d'estimation et de test par seuillage de coefficients d'ondelettes et les  $U$ -statistiques sont au coeur de cette thèse. C'est pourquoi le Chapitre 2 propose des rappels sur les bases d'ondelettes et donne une petite introduction aux  $U$ -statistiques, un outil abondamment utilisé dans les preuves des résultats de cette thèse. Les contributions principales des Chapitres 3, 4 et 5 sont détaillées ci-après. Quelques précisions :

- le Chapitre 3 fait l'objet d'un article qui a été accepté pour publication : SANSONNET, L. Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data. *Accepté pour publication au Scandinavian Journal of Statistics*, 2012,

- le Chapitre 4 fait l'objet d'un article qui est actuellement soumis : SANSONNET, L. ET TULEAU-MALOT, C. A model of Poissonian interactions and detection of dependence. *Soumis*, 2013, et
- le Chapitre 5 est un travail en cours en collaboration avec Patricia Reynaud-Bouret, Vincent Rivoirard et Rebecca Willett.

Une Annexe présentant les programmes codés en Scilab des principaux algorithmes mis en place dans cette thèse termine ce manuscrit.

### 1.4.1 Procédure de test adaptatif

Les résultats présentés dans cette section sont issus du Chapitre 4 et ont été obtenus en collaboration avec Christine Tuleau-Malot (Université Nice Sophia-Antipolis). Avant d'élaborer une procédure d'estimation de  $h$  du modèle (1.3), il semble pertinent de valider ce modèle et de savoir si l'introduction de cette fonction de reproduction  $h$  est justifiée et donc s'il y a une dépendance entre les deux événements considérés.

#### 1.4.1.1 Cadre et méthodologie

Dans le but de déceler de la dépendance ou non entre les deux événements considérés, on propose de tester la nullité de  $h$  dans le modèle (1.3).

Dans la suite, on suppose que  $h$  est à support compact connu (hypothèse justifiée dans les cadres de la génomique et des neurosciences) et sans perte de généralité, on suppose ici que le support de  $h$  est strictement inclus dans  $[-1; 1]$  et que l'on observe les  $U_i$  (les parents) sur  $[0; T]$  et une réalisation du processus  $N$  (les enfants) sur  $[-1; T + 1]$ .

On décompose alors la fonction  $h$  de  $\mathbb{L}_2(\mathbb{R})$  dans la base de Haar (voir Section 2.1.2 du Chapitre 2) :

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda \quad \text{avec} \quad \beta_\lambda = \int_{\mathbb{R}} h(x) \varphi_\lambda(x) dx,$$

où

$$\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}.$$

Comme on veut détecter un signal et non le reconstruire, la base de Haar convient tout à fait.

On veut tester l'hypothèse nulle  $\mathcal{H}_0 : \ll h = 0 \gg$  (le phénomène d'apparition des enfants ne dépend pas de celui des parents) contre l'alternative  $\mathcal{H}_1 : \ll h \neq 0 \gg$ . Par traduction de ces hypothèses en terme de coefficients d'ondelettes, on effectue d'abord des tests simples de  $\mathcal{H}_0 : \ll h = 0 \gg$  contre  $\mathcal{H}_1^\lambda : \ll \beta_\lambda \neq 0 \gg$ , pour  $\lambda \in \Gamma$ , où  $\Gamma$  est le sous-ensemble de  $\Lambda$  suivant :

$$\Gamma = \{\lambda = (j, k) \in \Lambda : j \geq 0, -2^j \leq k \leq 2^j - 1\}.$$

Cet ensemble  $\Gamma$  contient les indices  $\lambda = (j, k)$  avec  $j \geq 0$ , de sorte que  $\int_{-1}^1 \varphi_\lambda(t) dt = 0$  pour tout  $\lambda \in \Gamma$ . Remarquons que si un coefficient  $\beta_\lambda$  est non nul pour  $\lambda \notin \Gamma$ , alors au moins un coefficient  $\beta_\lambda$  est non nul pour  $\lambda \in \Gamma$  (voir Section 4.2 du Chapitre 4 pour plus de détails).

Pour tout  $\lambda \in \Gamma$ , le test simple associé se construit de la manière suivante :

1. On définit la statistique de test par  $\hat{T}_\lambda = |\hat{\beta}_\lambda|$ , avec

$$\hat{\beta}_\lambda = \frac{\mathcal{G}(\varphi_\lambda)}{n}, \quad \text{avec} \quad \mathcal{G}(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] dN_t, \quad (1.4)$$



où  $\pi$  désigne la loi uniforme sur  $[0; T]$ . Le premier terme de  $\mathcal{G}(\varphi_\lambda)$  est un estimateur naturel de  $n\beta_\lambda$  (sous la condition  $\int_{-1}^1 \varphi_\lambda(t) dt = 0$ ) et le second terme permet de le rendre sans biais (voir Proposition 4.1 du Chapitre 4).

2. Sous  $\mathcal{H}_0$ , conditionnellement aux  $U_1, \dots, U_n$  et à  $N_{[-1; T+1]} = m$ , les points du processus  $N$  suivent une loi uniforme sur  $[-1; T+1]$  et ainsi, la loi de  $\hat{T}_\lambda$  ne dépend que de quantités observables. On introduit alors le  $(1 - \alpha)$ -quantile  $q_\lambda^{[U_1, \dots, U_n; m]}(\alpha)$  de  $\hat{T}_\lambda$  sous  $\mathcal{H}_0$  conditionnellement aux  $U_1, \dots, U_n$  et à  $N_{[-1; T+1]} = m$  vérifiant

$$\mathbb{P}_0 \left( \hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; m]}(\alpha) \mid U_1, \dots, U_n, N_{[-1; T+1]} = m \right) \leq \alpha.$$

Notons que ce quantile est complètement implémentable.

3. Par conséquent, on rejette  $\mathcal{H}_0$  lorsque  $\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{[-1; T+1]}]}(\alpha)$ .

### 1.4.1.2 Résultats théoriques

Par construction, les test simples définis ci-dessus sont de niveau  $\alpha$ . De plus, le Théorème 4.1 du Chapitre 4 donne un seuil pour  $|\beta_\lambda|$  à partir duquel le test simple associé est capable de détecter un signal en ayant la puissance requise.

Pour tester à présent  $\mathcal{H}_0 : \ll h = 0 \gg$  contre  $\mathcal{H}_1 : \ll h \neq 0 \gg$ , on propose une procédure de test multiple basée sur l'agrégation de tests simples en les pondérant grâce à une collection de nombres positifs  $\{w_\lambda, \lambda \in \Gamma\}$  telle que  $\sum_{\lambda \in \Gamma} e^{-w_\lambda} \leq 1$ . On considère alors le test qui rejette  $\mathcal{H}_0$  lorsqu'il existe au moins un indice  $\lambda$  dans  $\Gamma$  tel que

$$\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{[-1; T+1]}]}(u_\alpha^{[U_1, \dots, U_n; N_{[-1; T+1]}]} e^{-w_\lambda}),$$

où

$$u_\alpha^{[U_1, \dots, U_n; N_{[-1; T+1]}]} = \sup \left\{ u > 0 : \mathbb{P} \left( \max_{\lambda \in \Gamma} \left( \hat{T}_{\lambda, N_{[-1; T+1]}}^0 - q_\lambda^{[U_1, \dots, U_n; N_{[-1; T+1]}]}(ue^{-w_\lambda}) \right) > 0 \mid U_1, \dots, U_n; N_{[-1; T+1]} \right) \leq \alpha \right\}.$$

La fonction de test correspondante est définie par

$$\Phi_\alpha = \mathbf{1}_{\max_{\lambda \in \Gamma} \left( \hat{T}_\lambda - q_\lambda^{[U_1, \dots, U_n; N_{[-1; T+1]}]}(u_\alpha^{[U_1, \dots, U_n; N_{[-1; T+1]}]} e^{-w_\lambda}) \right) > 0}.$$

On introduit les nombres réels positifs  $R_1$  et  $R_\infty$  tels que  $\|h\|_1 \leq R_1$  et  $\|h\|_\infty \leq R_\infty$ .

**Théorème 1.1** (cf. Théorème 4.2). *Soient  $\alpha, \beta$  des niveaux fixés dans  $]0; 1[$ . Alors, le test multiple  $\Phi_\alpha$  est de niveau  $\alpha$ . De plus, s'il existe au moins un sous-ensemble  $L$  de  $\Gamma$  tel que*

$$\|h_L\|_2^2 > C \left\{ (D_L + \sum_{\lambda \in L} w_\lambda) \left[ \frac{1}{n} + \frac{n}{T^2} \right] + (D_L + \sum_{\lambda \in L} w_\lambda + \sum_{\lambda \in L} w_\lambda^2) \left[ \frac{j_L}{n^2} + \frac{j_L^2 2^{j_L}}{n^3} + \frac{1}{n^2 T^2} \right] \right\}, \quad (1.5)$$

où  $h_L$  est la projection orthogonale de  $h$  sur l'espace engendré par  $\{\varphi_\lambda, \lambda \in L\}$ ,  $D_L$  le cardinal de  $L$ ,  $j_L = \max\{j \geq 0 : (j, k) \in L \text{ avec } k \in \mathcal{K}_j\}$  et  $C$  une constante positive dépendant de  $\alpha, \beta, \nu, R_1$  et  $R_\infty$ , alors

$$\mathbb{P}_h(\Phi_\alpha = 0) \leq \beta. \quad (1.6)$$

Le Théorème 1.1, avec la présence de la double asymptotique  $n$  et  $T$ , présente une vitesse inhabituelle  $[\frac{1}{n} + \frac{n}{T^2}]$  qui sera plus discutée à la Section 1.4.2.

L'équation (1.6) signifie que l'erreur de seconde espèce est contrôlée, ce qui montre que le test est puissant si la Condition (1.5) est vérifiée.

On se place dans le cas où  $n$  et  $T$  sont proportionnels, qui est le cas générique permettant par exemple de faire le lien avec le contexte génomique (voir Section 1.3.2.2.). De plus, cela permet d'interpréter aisément ce résultat. Avec de bonnes hypothèses sur  $j_L$  et sur  $\{w_\lambda, \lambda \in \Gamma\}$  (pour que ces quantités ne soient pas trop grandes), la vitesse de séparation du test  $\Phi_\alpha$  entre l'hypothèse nulle et l'alternative est de l'ordre de  $D_L/n$ , ce qui est typique des procédures de test basées sur une approche par seuillage (voir Fromont *et al.* [45, 46] par exemple). Ce théorème va permettre d'aboutir à un résultat d'adaptation au sens minimax sur des espaces de Besov classiques et faibles qui permettent de mesurer la parcimonie d'un signal décomposé dans une base d'ondelettes, avec une faible régularité. En effet, le choix d'une procédure par seuillage de coefficients d'ondelettes permet de détecter des fonctions parcimonieuses et ce sont ces fonctions que l'on s'attend à retrouver en génomique et en neurosciences. Le théorème suivant établit les vitesses de séparation uniformes du test  $\Phi_\alpha$  sur  $\mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R')$ . On trouvera une définition de ces espaces à la Section 2.1.4 du Chapitre 2. L'indice  $\delta$  mesure la régularité et l'indice  $\gamma$  la parcimonie.

**Théorème 1.2** (cf. Théorème 4.3). *Soient  $\alpha, \beta$  des niveaux fixés dans  $]0; 1[$ . On suppose que  $n$  et  $T$  sont proportionnels. En considérant un bon choix de poids  $w_\lambda$  et la fonction de test  $\Phi_\alpha$  associée, alors, pour  $\delta > 0, \gamma > 0, R > 0, R' > 0$ , si  $2\delta > \gamma/(1 + 2\gamma)$*

$$\rho(\Phi_\alpha, \mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R'), \beta) \leq C \left( \frac{\ln n}{n} \right)^{\frac{\gamma}{1+2\gamma}},$$

où  $C$  est une constante positive qui dépend de  $\delta, \gamma, R, R', \alpha, \beta, \nu, R_1$  et  $R_\infty$

Ce résultat illustre l'optimalité de notre procédure de test multiple du point de vue minimax, puisque les vitesses de séparation du test  $\Phi_\alpha$  correspondent aux vitesses minimax de test obtenues par Fromont *et al.*, [45, 46], dans le cas où  $2\delta > \gamma/(1 + 2\gamma)$  et aussi  $\delta < \gamma/2$  et  $\gamma > 1/2$ . Notons que cette vitesse de test correspond à la vitesse minimax d'estimation par seuillage (voir [66, 91, 95] par exemple), ce qui signifie qu'il est aussi difficile de tester que d'estimer sur de telles classes de fonctions, ce qui n'est pas usuel.

### 1.4.1.3 Résultats pratiques

Le Chapitre 4, dont sont issus les résultats précédemment énoncés, présente également des simulations, avec une comparaison au test de Kolmogorov-Smirnov (voir Darling [30]), noté dans la suite KS et à une méthode paramétrique de détection de dépendance en neurosciences, notée dans la suite GAUE, proposée par Tuleau-Malot *et al.* [109], qui utilise la technique des « Unitary Events » introduite par Grün [51].

Nous travaillons sur des jeux de données simulées qui reflètent une certaine réalité neurobiologique. Les parents sont simulés selon le modèle (1.2) avec  $\nu' = 50$  et  $g' = h' = 0$  et les enfants selon le modèle (1.1) avec  $\nu = 20, g = 0$  et  $h = \theta \mathbf{1}_{[\xi; 0.01]}$ , pour différents paramètres  $\theta$  et  $\xi$ . Alors que le paramètre  $\theta$  représente l'influence des parents sur les enfants, le paramètre  $\xi$  introduit un éventuel délai à la synchronisation. On considère les 9 jeux de données suivants :  $Data_0, Data_{10}$ ,

Jeu de donnée	notre procédure	KS	GAUE
$Data_0$	0.0486	0.051	0.0446/0.0510/0.0548

TABLE 1.1 – Niveaux de test empiriques, avec un niveau théorique  $\alpha = 0.05$ , pour notre procédure, KS et GAUE (minimum/médiane/maximum calculés sur la grille de  $\delta$ ).

Jeu de donnée	notre procédure	KS	GAUE
$Data_{10}$	0.095	0.040	0.068/0.1085/0.168
$Data_{30}$	0.478	0.051	0.154/0.3795/0.707
$Data_{50}$	0.864	0.087	0.278/0.6645/0.953
$Data_{80}$	0.993	0.113	0.451/0.9160/0.998
$Data_{10r}$	0.073	0.054	0.047/0.0575/0.077
$Data_{30r}$	0.282	0.059	0.050/0.1415/0.277
$Data_{50r}$	0.664	0.053	0.053/0.2825/0.589
$Data_{80r}$	0.968	0.073	0.048/0.4900/0.879

TABLE 1.2 – Puissances empiriques évaluées pour différentes interactions, avec un niveau théorique  $\alpha = 0.05$ , pour notre procédure, KS et GAUE (minimum/médiane/maximum calculés sur la grille de  $\delta$ ).

$Data_{30}$ ,  $Data_{50}$ ,  $Data_{80}$ ,  $Data_{10r}$ ,  $Data_{30r}$ ,  $Data_{50r}$  et  $Data_{80r}$ , où  $\forall k \in \{0, 10, 30, 50, 80\}$ ,  $Data_k$  est simulé avec  $\theta = k$  et  $\xi = 0$  et  $Data_{kr}$  est simulé avec  $\theta = k$  et  $\xi = 0.005$ . Le premier jeu de données permet d'évaluer le niveau de test empirique (en effectuant 5000 réalisations indépendantes) alors que les autres permettent d'évaluer la puissance empirique (en effectuant 1000 réalisations indépendantes de chaque jeu).

La méthode GAUE, méthode couramment utilisée en neurosciences, est basée sur le comptage des coïncidences avec un délai  $\delta$  sur  $[0; T]$ , c'est-à-dire le nombre de paires parent/enfant dont la distance entre eux est inférieure à  $\delta$ , où  $T$  est la durée simulée de l'enregistrement de l'activité neuronale ( $T = 2(\text{secondes})$ ). Cela introduit un paramètre  $\delta$  à choisir au préalable, que l'on fait varier sur une grille régulière  $[0.001; 0.04]$  de pas 0.001 pour nos simulations (voir Section 4.4.3 du Chapitre 4 ou Tuleau-Malot *et al.* [109] pour plus de précisions sur cette méthode).

Pour l'application de notre procédure de test, nous effectuons une remise à l'échelle sur  $[0; 100]$  des données pour des raisons algorithmiques. On choisit également un niveau de résolution maximal  $j_0 = 3$  pour limiter le nombre de tests simples à effectuer. Les Tableaux 1.1 et 1.2 résument les résultats de nos simulations.

Les 3 méthodes semblent avoir un niveau de test empirique correct. Le test KS présente des puissances empiriques petites (typiquement proches de  $\alpha$ ), ce qui signifie que le test KS ne parvient pas à détecter clairement la dépendance. Notre procédure et la méthode GAUE, quant à elles, offrent des puissances empiriques nettement meilleures. Plus  $\theta$  est grand, plus le test est puissant, ce qui s'explique par une meilleure séparation des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . Les 2 méthodes semblent comparables lorsque  $\xi = 0$  (i.e. la synchronisation peut être immédiate). Par contre, lorsque  $\xi \neq 0$  notre méthode semble meilleure avec de plus grandes puissances empiriques.

Ces simulations montrent ainsi le bon comportement pratique de notre procédure.

### 1.4.2 Procédure d'estimation adaptative

Les résultats présentés dans cette section sont issus du Chapitre 3. Nous proposons une procédure par seuillage qui est « data-driven » et optimale des points de vue oracle et minimax pour l'estimation de la fonction de reproduction  $h$  du modèle d'interactions poissoniennes défini par (1.3) avec  $\nu = 0$ .

#### 1.4.2.1 Cadre et méthodologie

On suppose que  $h$  est à support compact dans  $[-A; A]$ , avec  $A > 0$ . On décompose  $h$  dans une base d'ondelettes  $\{\varphi_\lambda, \lambda \in \Lambda\}$  gouvernée par le système d'ondelettes père/mère  $(\mathbf{1}_{[0;1]}, \psi)$  :

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda \quad \text{avec} \quad \beta_\lambda = \int_{\mathbb{R}} h(x) \varphi_\lambda(x) dx,$$

où

$$\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}.$$

Pour les aspects théoriques de cette thèse, on a considéré une base d'ondelettes biorthogonales qui a l'avantage de présenter des fonctions d'analyse qui sont constantes par morceaux facilitant l'implémentation algorithmique, différentes des fonctions de reconstruction qui elles sont régulières permettant des reconstructions lisses. On trouvera à la Section 2.1.2 du Chapitre 2 une description détaillée de cette base d'ondelettes biorthogonales spécifique due à Cohen *et al.* [23]. Ici, on fait l'amalgame entre les fonctions d'analyse et de reconstruction afin de simplifier les notations (ce qui est licite pour la base de Haar puisque le système d'ondelettes père/mère qui la génère est le même pour les fonctions d'analyse que pour les fonctions de reconstruction).

Pour tout  $\lambda \in \Lambda$ , on estime le coefficient d'ondelettes  $\beta_\lambda$  par

$$\hat{\beta}_\lambda = \frac{G(\varphi_\lambda)}{n}, \quad \text{avec} \quad G(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] dN_t. \quad (1.7)$$

Il s'agit d'un estimateur sans biais de  $\beta_\lambda$  (voir Proposition 3.1 du Chapitre 3) pour les mêmes raisons évoquées pour montrer le caractère sans biais de l'estimateur défini par (1.4). De plus, la variance de  $\hat{\beta}_\lambda$  est bornée :

$$\text{Var}(\hat{\beta}_\lambda) \leq C \left\{ \frac{1}{n} + \frac{n}{T^2} \right\}, \quad (1.8)$$

où  $C$  est une constante dépendant de  $R_1$ ,  $R_\infty$ ,  $\|\psi\|_1$  et  $\|\psi\|_2$ . Cette majoration présente un comportement peu habituel de la variance et cela est dû à l'intervention de deux paramètres de l'asymptotique :  $n$  le nombre de parents et  $T$  la longueur normalisée du génome. Toutefois, lorsque l'on est dans le cas où  $n$  et  $T$  sont proportionnels, la variance est bornée par  $1/T$  à une constante près, comme pour les processus de Hawkes (voir Reynaud-Bouret et Schbath [93]).

La procédure de seuillage s'effectue de la manière suivante :

1. On introduit un sous-ensemble déterministe  $\Gamma$  de  $\Lambda$  qui permet de définir un niveau de résolution maximale :  $\Gamma = \{\lambda = (j, k) \in \Lambda : -1 \leq j \leq j_0, k \in \mathcal{K}_j\}$ .
2. On définit le seuil qui dépend d'un paramètre  $\gamma > 0$  :

$$\eta_\lambda = \sqrt{2\gamma j_0 \tilde{V} \left( \frac{\varphi_\lambda}{n} \right)} + \frac{\gamma j_0}{3} B \left( \frac{\varphi_\lambda}{n} \right) + \Delta \frac{N_{\mathbb{R}}}{n},$$

où  $\Delta$  est définie par (3.8) et  $B(\varphi_\lambda)$  et  $\tilde{V}(\varphi_\lambda)$  (un estimateur de la variance de  $n\hat{\beta}_\lambda$ ), définies respectivement par (3.9) et (3.10), sont des quantités qui ne dépendent que des observations. Ce seuil, défini certes de manière compliqué, peut être mis en relation avec le seuil universel introduit par Donoho et Johnstone [37] dans un cadre de régression gaussienne (voir Section 2.1.3 du Chapitre 2 pour plus de détails).

3. Finalement, on définit l'estimateur seuillé de  $\beta_\lambda$  :  $\tilde{\beta}_\lambda = \hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma}$ , et par conséquent, l'estimateur de  $h$  :  $\tilde{h} = \sum_{\lambda \in \Lambda} \tilde{\beta}_\lambda \varphi_\lambda$ . Cet estimateur dépend uniquement du choix de  $\gamma$  et de  $j_0$ .

### 1.4.2.2 Résultats théoriques

Deux types de résultats sont établis : une inégalité de type oracle, de laquelle sont déduites des vitesses de convergence.

**Théorème 1.3** (cf. Théorème 3.1). *On suppose que  $n \geq 2$ ,  $j_0 \in \mathbb{N}^*$  tel que  $2^{j_0} \leq n < 2^{j_0+1}$  et  $\gamma > 2 \log 2$ . Alors l'estimateur  $\tilde{h}$  satisfait à*

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \left[ (\log n)^4 \times \frac{1}{n} + (\log n)^2 \times \frac{n}{T^2} \right] |m| \right\} + C_2 \left[ \frac{1}{n} + \frac{n}{T^2} \right], \quad (1.9)$$

où  $|m|$  est le cardinal de l'ensemble  $m$ ,  $C_1$  est une constante positive dépendant de  $\gamma$ ,  $R_1$ ,  $R_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  et  $\|\psi\|_\infty$  et  $C_2$  est une constante positive dépendant de  $A$ ,  $R_1$ ,  $R_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$ ,  $\|\psi\|_\infty$  et du support compact de  $\psi$ .

Comme le terme entre crochets dans (1.9) est du même ordre de grandeur que la majoration de  $\text{Var}(\hat{\beta}_\lambda)$  (voir (1.8)) à un terme logarithmique près, ce résultat montre que l'estimateur  $\tilde{h}$  satisfait à une inégalité oracle classique.

En particulier, si on considère le cas où  $n$  et  $T$  sont proportionnels, l'inégalité (1.9) du Théorème 1.3 devient

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{(\log T)^4}{T} |m| \right\} + \frac{C_2}{T}.$$

Cette inégalité de type oracle est similaire à celle obtenue par Reynaud-Bouret et Schbath [93] pour un processus de Hawkes et aussi aux inégalités oracles classiques obtenues en sélection de modèles, à un terme logarithmique près (voir par exemple Massart [77]).

**Théorème 1.4** (cf. Corollaire 3.2). *Soient  $R > 0$  et  $\delta \in \mathbb{R}$  tel que  $0 < \delta < r + 1$  (où  $r$  est le paramètre de régularité des ondelettes de reconstruction de la base biorthogonale, voir Section 2.1.2 du Chapitre 2). On suppose que  $h \in \mathcal{B}_{2,\infty}^\delta(R)$ ,  $n$  et  $T$  sont proportionnels et que les hypothèses du Théorème 1.3 sont vérifiées. Alors l'estimateur  $\tilde{h}$  satisfait à*

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C_3 \left( \frac{(\log n)^4}{n} \right)^{\frac{2\delta}{2\delta+1}},$$

où  $C_3$  est une constante positive dépendant de  $\gamma$ ,  $A$ ,  $R_1$ ,  $R_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$ ,  $\|\psi\|_\infty$ ,  $R$  et du support compact de  $\psi$ .

Ce résultat illustre l'optimalité de notre estimateur  $\tilde{h}$  du point de vue minimax, puisque la vitesse de convergence ainsi exhibée correspond à la vitesse minimax d'estimation, à un terme logarithmique près, obtenue par exemple par Reynaud-Bouret [90] pour l'estimation de l'intensité à support compact d'un processus de Poisson ou pour l'estimation de densité à support compact à partir de l'observation d'un  $n$ -échantillon par Donoho *et al.* [41].

### 1.4.2.3 Résultats pratiques

L'implémentation de la procédure, avec en particulier le calcul du seuil, nécessite l'usage d'un algorithme de descente (voir Mallat [74]) qui a été programmé pour la base de Haar et une base de Spline (voir Figure 3.1 du Chapitre 3). D'une part des simulations ont permis de calibrer les paramètres qui apparaissent dans la définition du seuil du point de vue pratique. Une fois les paramètres « optimaux » pratiques choisis, nous proposons d'autres simulations afin de montrer la robustesse et le bon comportement pratique de notre procédure d'estimation. Les Figures 1.3 et 1.4 en donnent un exemple avec les reconstructions respectives des signaux :

$$\text{'Signal1'} : 4 \times \mathbf{1}_{[0;1]} \quad \text{et} \quad \text{'SignalN'} : 4 \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Pour ces reconstructions, on a pris  $T = 10000$ ,  $A = 10$  (on rappelle que  $[-A; A]$  est le support de  $h$ ) et le niveau de résolution maximal  $j_0 = 5$ . Les parents ont été simulés selon un processus de Poisson d'intensité 0.1 sur  $[0; T]$  et les enfants selon le modèle (1.3) avec  $\nu = 0$  et  $h$  la fonction 'Signal1' ou 'SignalN'.

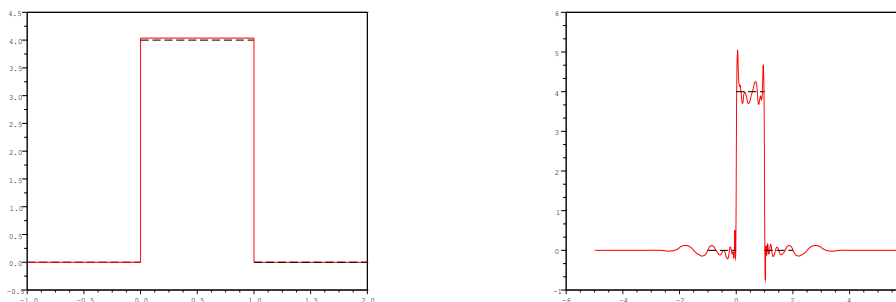


FIGURE 1.3 – Reconstruction de 'Signal1' (vraie : en pointillé, estimée : trait plein) : à gauche : dans la base de Haar ; à droite dans la base de Spline.

Les reconstructions sont satisfaisantes. Le caractère intrinsèque des deux bases considérées ainsi que des deux signaux à reconstruire donne une meilleure reconstruction de 'Signal1' (resp. 'SignalN') par la décomposition dans la base de Haar (resp. la base de Spline). Néanmoins, dans tous les cas, on retrouve l'allure générale de la fonction à estimer. La Section 3.3 du Chapitre 3 propose d'autres simulations et établit ainsi la robustesse et le bon comportement pratique de notre procédure, avec en particulier une capacité à gérer quelques enfants orphelins (via une apparition spontanée) et à retrouver des fonctions négatives si l'on introduit un phénomène d'inhibition (voir Section 3.3.5 pour plus de détails).

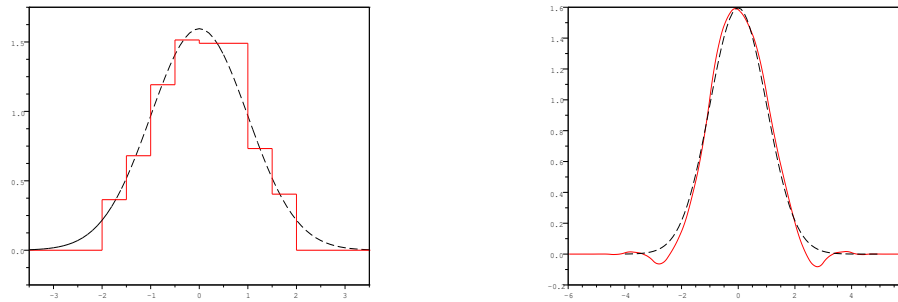


FIGURE 1.4 – Reconstruction de 'SignalN' (vraie : en pointillé, estimée : trait plein) : à gauche : dans la base de Haar ; à droite dans la base de Spline.

#### 1.4.2.4 Application à des données génomiques

Comme mentionné à la Section 1.2.1, nous appliquons notre procédure à l'étude de la bactérie *E. coli* afin d'analyser l'influence entre les gènes et les sites promoteurs, plus précisément, le motif « tataat ». La plupart des gènes sont précédés à très courte distance de ce motif. Afin de valider notre procédure, on souhaite détecter ces distances. On dispose pour cela de la position de 4290 gènes (représentée par la première base codant la séquence) et de 1036 occurrences du motif tataat, le long du génome de longueur normalisée  $T = 9288442$  (en prenant  $A = 10000$  la mémoire maximale le long de la séquence d'ADN). La Figure 1.5 donne l'allure de l'estimateur  $\tilde{h}$  dans le modèle où les parents sont les tataat d'une part et les gènes d'autre part.

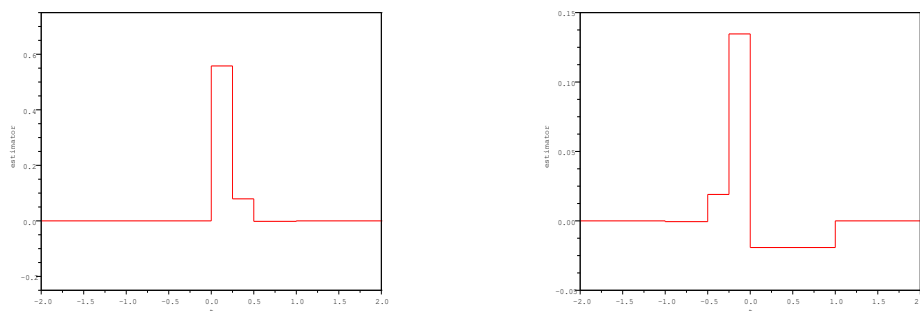


FIGURE 1.5 – Estimateur  $\tilde{h}$  pour les données *E. coli* à l'échelle 1 : 1000, avec à gauche : parents=tataat et enfants=gènes et à droite : parents=gènes et enfants=tataat.

La figure de gauche, qui explique l'influence des tataat sur les gènes, signifie que

- l'estimateur étant nul avant 0 base et au-delà de 500 bases, les gènes ne semblent pas être influencés par les tataat sur ces intervalles, et
- l'estimateur étant positif entre 0 et 500 bases, les courtes distances sont favorisées et plus la distance est petite, plus grande est la préférence.

La figure de droite, qui explique l'influence des gènes par les tataat, nous permet de tirer les mêmes conclusions avec en plus un phénomène intéressant. En effet, entre 0 et 1000 bases,

l'estimateur est négatif, ce qui signifie que les occurrences de **tataat** sont évitées sur de telles distances. Cela peut s'expliquer du point de vue pratique par le fait que les gènes, de longueur environ 1000 bases, ne se chevauchent pas. Autrement dit, le long du génome, l'apparition du motif **tataat** est évitée lors du codage d'un gène. À noter que le cas  $h$  négatif n'a pas été étudié d'un point de vue théorique.

En conclusion, cette application sur des données réelles a permis de détecter les distances favorisées ou évitées entre les gènes et les sites promoteurs le long du génome de la bactérie *E. coli* attendues et donc de valider notre procédure.

### 1.4.3 Estimation Lasso et interactions poissoniennes sur le cercle

Les résultats présentés dans cette section sont issus du Chapitre 5 et ont été obtenus en collaboration avec Patricia Reynaud-Bouret (Université Nice Sophia-Antipolis), Vincent Rivoirard (Université Paris Dauphine) et Rebecca Willett (Duke University). Ces travaux se situent à la frontière entre statistique et traitement du signal dans le champ du « compressed sensing » (voir Donoho [36]) et constituent un problème inverse poissonien.

#### 1.4.3.1 Cadre et méthodologie

La procédure par seuillage proposée à la Section 1.4.2 pour estimer la fonction de reproduction  $h$  dans un modèle d'interactions poissoniennes, a montré ses performances d'un point de vue théorique mais également d'un point de vue pratique, en offrant une meilleure précision et une plus grande rapidité de calcul que celle proposée par Reynaud-Bouret et Schbath [93]. Cependant, elle nécessite la connaissance en amont d'une borne  $A$  sur la portée d'interaction, tout comme pour le processus de Hawkes (voir [56, 93] par exemple). On souhaite alors s'affranchir de la connaissance de cette borne  $A$  dans un modèle discret en grande dimension, qui peut être vu comme une version discrète du modèle (1.3) avec  $\nu = 0$ .

Soient  $U_1, \dots, U_n$   $n$  variables aléatoires uniformes i.i.d. sur  $\{0, \dots, M-1\}$ , représentant des points sur un cercle (on travaille modulo  $M$ ). Considérer un modèle circulaire permet d'éviter les effets de bord mais permet également de refléter une réalité biologique puisque certains génomes, comme ceux des bactéries, sont circulaires. Chaque parent  $U_i$  donne indépendamment naissance à des enfants en nombre poissonien, la position des enfants dépendant de l'écart aux parents. Plus précisément, si  $h^* = (h_0^*, \dots, h_{M-1}^*)^T$  est un vecteur de  $\mathbb{R}_+^M$ , alors le nombre d'enfants du parent  $U_i$  à distance  $j$ ,  $N_{U_i+j}^i$ , suit une loi de Poisson  $\mathcal{P}(h_j^*)$ . Le but est alors d'estimer  $h^*$  à partir de l'observation des parents  $U_i$  et du nombre total d'enfants à chaque site  $k \in \{0, \dots, M-1\}$  :

$$Y_k = \sum_{i=1}^n N_k^i, \quad \text{dont la loi conditionnellement aux } U_i \text{ est } \mathcal{P}\left(\sum_{i=1}^n h_{k-U_i}^*\right).$$

On peut reformuler le problème comme suit :

$$Y = (Y_0, \dots, Y_{M-1})^T \sim \mathcal{P}(Gh^*)$$

et ainsi, estimer  $h^*$  est équivalent à un problème inverse où l'opérateur  $G$  est aléatoire et dépend uniquement des  $U_i$  (il est donc observable). On s'oriente vers une estimation de type Lasso (introduite par Tibshirani [105] en régression linéaire) afin de capter la parcimonie de  $h^*$ . En particulier, notre estimateur Lasso nécessitera une propriété de type RIP (Restricted Isometry



Property, introduite par Candès et Tao [19]) sur la matrice aléatoire  $G$  avec grande probabilité et l'introduction de poids aléatoires « data-driven » dans sa définition (par exemple, voir Zou [117], Huang *et al.* [62], Bertin *et al.* [9] ou encore Hansen *et al.* [56]).

### 1.4.3.2 Résultats théoriques

On établit au préalable une propriété de type RIP sur  $G$ , avec une approche similaire à celle de Rudelson et Vershynin [99].

**Théorème 1.5** (cf. Théorème 5.2). *Soit  $\delta \in ]0; 1[$  et soit  $K \in \mathbb{N}$  tel que  $K < M$ . On suppose que*

$$c := \kappa \delta^{-2} \max \left( \frac{K^{3/4}}{M^{1/2}}, \frac{1}{n} \right) K^2 [\log(M) + \log(n)]^2 < 1, \quad (1.10)$$

où  $\kappa$  est une constante positive absolue. Alors avec probabilité plus grande que  $1 - \delta$ , pour tout vecteur  $x$   $K$ -parcimonieux (i.e. avec au plus  $K$  coordonnées non nulles),

$$r \|x\|_2^2 := (1 - c)n \|x\|_2^2 \leq \|Gx\|_r^2 \leq (1 + c)n \|x\|_2^2 =: R \|x\|_2^2, \quad (1.11)$$

où  $\|Gx\|_r^2 = \|Gx\|_2^2 - (n^2 - n) \frac{(\mathbf{1}^T x)^2}{M}$ . En particulier, si  $K = O(M^\alpha)$  et  $K = O(n^\beta)$  quand  $\min(M, n) \rightarrow +\infty$  avec  $\alpha < 2/11$  et  $\beta < 1/2$ , alors pour toute constante absolue  $c \in ]0; 1[$ , il existe  $\delta(n, M)$  qui tend vers 0 à une vitesse polynomiale tel que l'inégalité (1.11) soit vraie avec une probabilité plus grande que  $1 - \delta(n, M)$ .

Cette propriété va nous permettre d'obtenir des inégalités de type oracle pour notre estimateur de type Lasso. Pour l'instant, on définit un estimateur modifié de type Lasso que l'on force à être parcimonieux : pour  $K_0 \in \mathbb{N}^*$ ,

$$\hat{h}_L = \hat{h}_L(K_0) = \underset{h \in \mathbb{R}^M \text{ } K_0\text{-parcimonieux}}{\operatorname{argmin}} \left\{ \|Y - Gh\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k |h_k| \right\}, \quad (1.12)$$

pour  $\eta > 0$  et des poids aléatoires  $d_k$  à calibrer. Cette approche avec des poids aléatoires a déjà été utilisée par exemple par Bertin *et al.* [9] et Hansen *et al.* [56]. D'une certaine manière, ces poids jouent le même rôle que le seuil aléatoire des procédures d'estimation par seuillage (voir par exemple notre procédure d'estimation adaptative à la Section 1.4.2 ou encore Reynaud-Bouret et Rivoirard [91]). Précisons que l'estimateur  $\hat{h}_L$  n'est pas implémentable pour de grandes valeurs de  $M$  en un temps de calculs raisonnable.

Le choix des poids aléatoires  $d_k$  est donc crucial. Il faut qu'ils ne dépendent que de quantités observables (pour pouvoir les calculer en pratique), qu'ils vérifient avec grande probabilité  $|G^H(Y - Gh^*)|_k \leq d_k$  (via une inégalité de type Bernstein) afin d'obtenir des inégalités oracles avec grande probabilité (voir Théorème 1.8) et qu'ils soient petits (voir l'inégalité (1.15)). Un choix est donné par le théorème suivant.

**Théorème 1.6** (cf. Théorème 5.3). *Soit  $\delta > 0$  (tel que  $\delta < 3$ ). On définit pour tout  $k$  dans l'ensemble  $\{0, \dots, M - 1\}$ ,*

$$d_k = \sqrt{2 \ln(3/\delta) \tilde{V}_k} + \frac{1}{3} \ln(3/\delta) B_k, \quad (1.13)$$

avec

$$\tilde{V}_k = \hat{V}_k + \sqrt{2 \ln(3/\delta) \hat{V}_k B_k^2} + 3 \ln(3/\delta) B_k^2,$$

$$\hat{V}_k = \sum_{1 \leq i, j \leq n} \mathbf{1}_{U_i=U_j} Y_{k+U_i} \quad \text{et} \quad B_k = \sup_{t \in \mathcal{C}_{\{0, \dots, M-1\}}} \sum_{i=1}^n \mathbf{1}_{U_i+k=t},$$

où  $\mathcal{C}_{\{0, \dots, M-1\}}$  désigne le cercle discret associé à l'ensemble  $\{0, \dots, M-1\}$ . Alors, avec probabilité plus grande que  $1 - \delta$ ,

$$\forall k \in \{0, \dots, M-1\}, \quad |G^H(Y - Gh^*)|_k \leq d_k. \quad (1.14)$$

Le théorème suivant donne un ordre de grandeur des poids aléatoires  $d_k$ .

**Théorème 1.7** (cf. Théorème 5.4). *Pour tout  $\delta \in ]0; 1[$ ,*

$$\mathbb{E}(d_k^2) \leq C \left\{ \log(3/\delta) \left( n + \frac{n^3}{M^2} \right) + (\log(3/\delta))^2 \left( n + \frac{n^2}{M^2} \right) \right\},$$

où  $C$  est une constante positive dépendant de  $h_k^*$  et  $\|h^*\|_{\ell_1}$ .

On s'intéresse aux propriétés théoriques de l'estimateur Lasso  $\hat{h}_L$  du point de vue oracle. Le théorème suivant propose une inégalité oracle pour la perte  $\ell_2$ .

**Théorème 1.8** (cf. Théorème 5.5). *Soit  $\Omega$  un événement sur lequel la borne inférieure de (1.11) est vraie avec  $K \geq 2K_0$  et  $c < 1/2$  et tel que (1.14) soit vérifiée. Cela signifie que sur  $\Omega$ , pour tout vecteur  $x$  au moins  $2K_0$ -parcimonieux dans  $\mathbb{C}^M$*

$$\|Gx\|_2^2 \geq \|Gx\|_r^2 \geq (1-c)n\|x\|_{\ell_2}^2$$

et

$$|G^H(Y - Gh^*)|_k \leq d_k \quad \forall k.$$

Sur  $\Omega$  (i.e. avec probabilité au moins  $1 - 2\delta$ ), si  $\eta > 0$  et si  $h^*$  est  $K_0$ -parcimonieux,

$$\|\hat{h}_L - h^*\|_{\ell_2}^2 \leq \frac{8(2+\eta)^2}{(1-c)^2 n^2} \sum_{k \in S(h^*)} d_k^2, \quad (1.15)$$

où  $S(h)$  est le support de  $h$ .

Par le Théorème 1.7, l'espérance du membre de droite de (1.15) est de l'ordre  $\frac{1}{n} + \frac{n}{M^2}$ , à un terme logarithmique près. Notons que cette vitesse a déjà été exhibée aux Sections 1.4.1 et 1.4.2.

Si l'on suppose que  $n = O(M)$ , le Théorème 1.5 montre que si  $K = O(M^\alpha)$  et  $K = O(n^\beta)$  quand  $\min(M, n) \rightarrow +\infty$  avec  $\alpha < 2/11$  et  $\beta < 1/2$ , alors on peut choisir  $\delta = M^{-\gamma}$  avec  $\gamma > 0$  tel que (1.11) avec probabilité plus grande que  $1 - M^{-\gamma}$ . Avec ce choix de  $\delta$ , le Théorème 1.7 montre alors que les  $d_k$  sont de l'ordre  $\log M \sqrt{n}$  si  $n = O(M)$ . Le membre de droite de (1.15) étant ainsi de l'ordre  $(\log M)^2/n$ , l'inégalité (1.15) est une inégalité de type oracle classique, à un terme logarithmique près, pour notre estimateur Lasso (voir par exemple [9, 16, 17, 110]).

Des perspectives immédiates à ce travail inachevé sont les suivantes (voir Section 5.5 pour plus de détails) :

- On souhaite mener une étude pratique sur des données simulées. L'estimateur de type Lasso  $\hat{h}_L$  n'étant pas implémentable pour de grandes valeurs de  $M$  en un temps raisonnable, on devra effectuer des simulations pour l'estimateur Lasso classique

$$\tilde{h}_L := \operatorname{argmin}_{h \in \mathbb{R}^M} \left\{ \|Y - Gh\|_{\ell_2}^2 + 2 \sum_{k=0}^{M-1} d_k |h_k| \right\}.$$

- On souhaite également établir avec grande probabilité que l'estimateur Lasso classique  $\tilde{h}_L$  satisfait à la propriété de  $K_0$ -parcimonie si  $h^*$  est  $K_0$ -parcimonieux afin d'établir des inégalités de type oracle pour cet estimateur.
- On souhaite également étudier l'estimateur Dantzig associé défini par

$$\hat{h}_D := \operatorname{argmin}_{h \in \mathbb{R}^M} \{ \|h\|_{\ell_1} : \forall k \in \{0, \dots, M-1\}, |G^H(Y - Gh^*)|_k \leq d_k \}.$$

En effet, les propriétés de type RIP permettent d'établir des inégalités oracles avec grande probabilité pour la perte  $\ell_2$  pour l'estimateur Dantzig (voir Candès et Tao [20]). Malheureusement le Théorème 1.5 ne présente pour l'instant un tel résultat que pour la norme  $\|\cdot\|_r$ , au lieu de  $\|\cdot\|_2$ .

## 1.5 Perspectives

Outre les perspectives immédiates du Chapitre 5 que nous venons de proposer, des axes de recherche faisant suite à ces travaux sont envisageables.

- Tout au long de ces travaux, nous nous sommes restreints au cas de la dimension 1. Il serait donc intéressant d'étudier, toujours dans le cadre de la statistique non-paramétrique adaptative, des processus similaires à notre modèle d'interactions poissoniennes (1.3) dans le cadre spatial et de les relier, par exemple, au processus de Neymann-Scott. Ce processus est de type « Poisson cluster process ». Il s'agit d'un mécanisme en deux étapes :
  1. On génère un processus de Poisson homogène qui donne la première génération :  $X_1, \dots, X_n$ .
  2. Autour de chaque  $X_i$ , on génère un processus ponctuel  $N_{X_i}$  de même loi qu'un processus  $N_0$  et de manière indépendante des autres points.

Le processus de Neyman-Scott est le résultat du second tirage (c'est donc l'équivalent des enfants dans notre modèle). Ce processus a été essentiellement étudié dans le cadre paramétrique (voir Tanaka *et al.* [104] par exemple) et est utilisé pour modéliser des données en astronomie et des structures forestières. Voir Section 6.3 de Daley et Vere-Jones [29] pour plus de détails.

- Le modèle (1.3) postule que la fonction  $h$  est la même sur tout l'intervalle  $[0; T]$  (et  $\nu$  constante), ce qui n'est pas très réaliste des points de vue génomique et neurobiologique. En effet, en neurosciences, où  $T$  représente le temps d'enregistrement de l'activité neuronale, les neurones peuvent avoir des comportements très différents et du coup des influences inégales selon que le sujet soit dans une phase d'attente, de repos ou de travail, par exemple. Par contre, cette hypothèse de stationnarité peut se faire sur des intervalles de temps plus courts (même s'ils ne sont pas facilement identifiables). On pourrait donc coupler notre procédure de test adaptatif proposée à la Section 1.4.1 avec une approche semblable à celle de Benjamini et Hochberg [8]. Cela consisterait à se donner une collection  $\mathcal{W}$  de  $K$  fenêtres de  $[0; T]$  (qui peuvent se chevaucher) et d'effectuer sur chaque fenêtre  $W$  de  $\mathcal{W}$  un test  $\Phi_\alpha^W$  (construit avec une méthode similaire à celle de la Section 1.4.1) pour détecter la dépendance sur cette fenêtre ou non. La procédure de Benjamini et Hochberg [8] permettrait alors de contrôler le FDR (False Discovery Rate).
- La procédure d'estimation adaptative proposée à la Section 1.4.2 est optimale du point de vue oracle. Mais, on a montré son optimalité du point de vue minimax uniquement lorsque

$n$  et  $T$  sont proportionnels (voir Théorème 1.4). Plus généralement, ce résultat est valide lorsque  $n = O(T)$ , c'est-à-dire lorsque les parents sont suffisamment espacés pour rendre leurs interactions insignifiantes pour l'analyse statistique. Par contre le régime  $n \gg T$  n'est pour l'instant pas étudié. Ce régime signifie qu'il y a beaucoup de parents sur l'intervalle d'observation  $[0; T]$ , ce qui rend très difficile la détermination du parent de chaque enfant. Cela revient alors à un problème particulier de déconvolution statistique (la convolution se faisant avec des lois uniformes) qui est dans ce cas un problème inverse mal-posé. De plus, lorsque  $n \gg T$ , le terme principal de l'inégalité oracle (1.9) du Théorème 1.3 est en  $n/T^2$  (à un terme logarithmique près) et il faudrait donc étudier la borne inférieure du risque minimax associé afin de déterminer si la vitesse  $n/T^2$  est optimale ou non.

- Du point de vue de la génomique, il serait intéressant d'étudier un modèle prenant en compte la vraie structure 3D de l'ADN. Pour l'instant, on a seulement eu accès à des données génomiques lues « en ligne » et donc on a introduit une mémoire maximale de 10000 bases (établie par les biologistes) afin d'éviter que la structure 3D de l'ADN vienne interférer et que les distances réelles physiques soient éventuellement plus courtes que les distances lues « en ligne ». Cependant, il se pourrait que le type d'analyse mené au Chapitre 5 puisse se généraliser à des graphes plus complexes que le cercle, qui représenteraient la structure réelle de l'ADN. Avec les données adéquates, cela permettrait alors de comprendre plus finement les relations en terme de distance physique réelle entre, par exemple, différents éléments régulateurs de la transcription. D'ailleurs, de récentes technologies permettent d'étudier la conformation des chromosomes en 3D (voir [101]).



# Chapitre 2

## Mathematical tools

This chapter succinctly presents the main mathematical tools used in this manuscript. In the first section, we recall the basics of wavelet bases and in particular biorthogonal wavelet bases and we briefly describe wavelet thresholding procedures and Besov spaces. The second section provides an overview of the  $U$ -statistics theory. Part of the following material is based on the book of Härdle *et al.* [57] for the first section and on the book of de la Peña and Giné [32] for the second section.

### Contents

---

<b>2.1</b>	<b>Wavelet thresholding on classical and weak Besov spaces</b>	<b>37</b>
2.1.1	Multiresolution analysis	37
2.1.2	The Haar basis and a specific biorthogonal wavelet basis	38
2.1.3	Wavelet thresholding estimation	39
2.1.4	Classical and weak Besov bodies	40
<b>2.2</b>	<b>An introduction to the theory of <math>U</math>-statistics</b>	<b>41</b>
2.2.1	Definition and elementary examples	41
2.2.2	Some exponential and moment inequalities for $U$ -statistics	42

---

## 2.1 Wavelet thresholding on classical and weak Besov spaces

In this section, we give the basics of the wavelet bases construction. Then, we present a specific biorthogonal wavelets basis, due to Cohen *et al.* [23], used in this manuscript. Finally, we describe wavelet thresholding rules and the functional spaces considered in this work. For more details, see for instance Daubechies [31], Meyer [78] and Mallat [75].

### 2.1.1 Multiresolution analysis

The definition of wavelet bases is linked to the fundamental concept of multiresolution analysis.

**Definition 2.1.** A multiresolution analysis of  $\mathbb{L}_2(\mathbb{R})$  is a nested sequence of spaces  $\{V_j, j \in \mathbb{Z}\}$  that satisfies the following properties:

- $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ ,

- $\bigcup_{j \in \mathbb{Z}} V_j$  is dense in  $\mathbb{L}_2(\mathbb{R})$ ,
- $\forall f \in \mathbb{L}_2(\mathbb{R}), \forall j \in \mathbb{Z}, \forall x \in \mathbb{R}, f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$ ,
- $\forall f \in \mathbb{L}_2(\mathbb{R}), \forall k \in \mathbb{Z}, \forall x \in \mathbb{R}, f(x) \in V_0 \Leftrightarrow f(x - k) \in V_0$ , and
- there exists a function  $\phi \in V_0$ , named father wavelet, such that  $\{\phi(\cdot - k), k \in \mathbb{Z}\}$  is an orthonormal basis of  $V_0$ .

Consequently, at each resolution level  $j$ , the space  $V_j$  of a multiresolution analysis  $\{V_j, j \in \mathbb{Z}\}$  is generated by an orthonormal basis obtained by dilatations and translations of the father wavelet  $\phi$ :  $\{\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), k \in \mathbb{Z}\}$ . For any  $j \in \mathbb{Z}$ , we define the space  $W_j$  such that

$$V_{j+1} = V_j \oplus W_j,$$

namely  $W_j$  is the orthogonal complement of  $V_j$  in  $V_{j+1}$ . We can build a function  $\psi$ , named mother wavelet, such that  $\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), k \in \mathbb{Z}\}$  is an orthonormal basis of  $W_j$ . By definition of the multiresolution analysis  $\{V_j, j \in \mathbb{Z}\}$  and the  $W_j$ 's, for any  $j_0 \in \mathbb{Z}$ , one obtains

$$\mathbb{L}_2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j \geq j_0} W_j.$$

This means that any  $f \in \mathbb{L}_2(\mathbb{R})$  can be represented as a series: for any  $j_0 \in \mathbb{Z}$ ,

$$f = \sum_{k \in \mathbb{Z}} \alpha_{j_0,k} \phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k}, \quad (2.1)$$

where the wavelet coefficients are

$$\alpha_{j_0,k} = \int_{\mathbb{R}} f(x) \phi_{j_0,k}(x) dx \quad \text{and} \quad \beta_{j,k} = \int_{\mathbb{R}} f(x) \psi_{j,k}(x) dx.$$

Without loss of generality, we can choose  $j_0 = 0$  and we denote  $\forall k \in \mathbb{Z}, \forall x \in \mathbb{R}$ ,

$$\psi_{-1,k}(x) = \phi_{0,k}(x) \quad \text{and} \quad \beta_{-1,k} = \alpha_{0,k}.$$

This multiresolution analysis yields orthonormal bases whose elements are localized both in frequency and time domain and obtained by dyadic dilatations and translations of a "father/mother wavelets" system  $(\phi, \psi)$ .

### 2.1.2 The Haar basis and a specific biorthogonal wavelet basis

The Haar basis wavelet system, defined by

$$\phi = \mathbf{1}_{[0;1]} \quad \text{and} \quad \psi = \mathbf{1}_{] \frac{1}{2}; 1]} - \mathbf{1}_{[0; \frac{1}{2}]},$$

has the properties of the multiresolution analysis described above with

$$V_j = \overline{\text{Span}\{\phi_{j,k}, k \in \mathbb{Z}\}}.$$

The functions associated with the Haar basis are discontinuous and irregular with few fluctuations. This is why the Haar system is not very convenient for approximation of smooth functions. But here, in the biological context, the reproduction function to estimate (see Section 1.3.2 of Chapter 1) can be very irregular, for instance it can have some localized spikes. So the choice of the Haar basis may be suitable here.

If we are interested in the approximation of smooth functions, we can consider a specific biorthogonal wavelet basis, built by Cohen *et al.* [23], that we can describe as follows. We set  $\phi = \mathbf{1}_{[0,1]}$  the analysis father wavelet. For any  $r > 0$ , there exist three functions  $\psi$ ,  $\tilde{\phi}$  and  $\tilde{\psi}$  with the following properties:

- $\tilde{\phi}$  and  $\tilde{\psi}$  are compactly supported,
- $\tilde{\phi}$  and  $\tilde{\psi}$  belong to  $C^{r+1}$ , where  $C^{r+1}$  denotes the Hölder space of order  $r + 1$ ,
- $\psi$  is compactly supported and is a piecewise constant function,
- $\psi$  is orthogonal to polynomials of degree no larger than  $r$ ,
- $\left\{ (\phi_k, \psi_{j,k})_{j \geq 0, k \in \mathbb{Z}}, (\tilde{\phi}_k, \tilde{\psi}_{j,k})_{j \geq 0, k \in \mathbb{Z}} \right\}$  is a biorthogonal family: for any  $j, j' \geq 0$ , for any  $k, k' \in \mathbb{Z}$ ,

$$\int_{\mathbb{R}} \phi_k(x) \tilde{\psi}_{j',k'}(x) dx = \int_{\mathbb{R}} \psi_{j,k}(x) \tilde{\phi}_{k'}(x) dx = 0,$$

$$\int_{\mathbb{R}} \phi_k(x) \tilde{\phi}_{k'}(x) dx = \mathbf{1}_{\{k=k'\}}, \quad \int_{\mathbb{R}} \psi_{j,k}(x) \tilde{\psi}_{j',k'}(x) dx = \mathbf{1}_{\{j=j', k=k'\}},$$

where for any  $x \in \mathbb{R}$ ,

$$\phi_k(x) = \phi(x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$$

and

$$\tilde{\phi}_k(x) = \tilde{\phi}(x - k), \quad \tilde{\psi}_{j,k}(x) = 2^{j/2} \tilde{\psi}(2^j x - k).$$

This implies the following wavelet decomposition of  $f \in \mathbb{L}_2(\mathbb{R})$ :

$$f = \sum_{k \in \mathbb{Z}} \alpha_k \tilde{\phi}_k + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{j,k} \tilde{\psi}_{j,k},$$

where for any  $j \geq 0$  and any  $k \in \mathbb{Z}$ ,

$$\alpha_k = \int_{\mathbb{R}} f(x) \phi_k(x) dx, \quad \beta_{j,k} = \int_{\mathbb{R}} f(x) \psi_{j,k}(x) dx.$$

The Haar basis can be viewed as a particular biorthogonal wavelet basis, by setting  $\tilde{\phi} = \phi$  and  $\tilde{\psi} = \psi = \mathbf{1}_{[\frac{1}{2}, 1]} - \mathbf{1}_{[0, \frac{1}{2}]}$ , with  $r = 0$  (even if the second property is not satisfied with such a choice). The Haar basis is an orthonormal basis, which is not true for general biorthogonal wavelet bases. Another example of biorthogonal wavelet basis is given at Section 3.3.1 of Chapter 3.

The interest of biorthogonal wavelet bases lies in the fact that they can have the following property: analysis wavelets  $\phi_k$  and  $\psi_{j,k}$  are piecewise constant functions and reconstruction wavelets  $\tilde{\phi}_k$  and  $\tilde{\psi}_{j,k}$  are typically smooth functions. This nice property allows us both to implement low computational complexity algorithms, as cascade algorithm (see Mallat [74]) which has been implemented in our context (see Section 3.3.2 of Chapter 3) and to provide smooth reconstructions.

### 2.1.3 Wavelet thresholding estimation

Thresholding estimation procedures have been introduced by Donoho and Johnstone [37] in 1994. Since we expect that only very few wavelet coefficients contribute to the main part of the signal, we consider thresholding rules that retain only observed data that exceed a certain threshold. More precisely, Donoho and Johnstone [37] display two thresholding estimators:



- an estimator by hard thresholding:  $d_t^H(x) = x\mathbf{1}_{|x|>t}$ , and
- an estimator by soft thresholding:  $d_t^S(x) = \text{sign}(x)(|x| - t)_+$ ,

with the threshold  $t$ . Therefore, the choice of the threshold  $t$  is crucial and has been the subject of numerous studies: Donoho et Johnstone [37, 38], Nason [81] for instance or more recently, Juditsky and Lambert-Lacroix [65], Reynaud-Bouret and Rivoirard [91] (concerning in particular the estimation of the intensity of one Poisson process) and Reynaud-Bouret *et al.* [92]. Furthermore, Donoho and Johnstone [37] defined the universal threshold  $t = \sqrt{2\sigma^2 \log n}$  in the Gaussian regression framework, where  $\sigma^2$  (assumed to be known) is the variance of each noisy wavelet coefficient. With the universal threshold, Donoho and Johnstone [37] show that their estimation procedure constructed by hard or soft thresholding is near optimal from the minimax point of view on Besov balls. Consequently, the wavelet thresholding estimation procedures are very effective from a theoretical point of view as evidenced by the work of Donoho and collaborators [37, 38, 40, 41, 42].

### 2.1.4 Classical and weak Besov bodies

Besov spaces, denoted  $\mathcal{B}_{p,q}^\delta$ , are classically defined using modulus of continuity (see [57] for instance). Here, we just recall the sequential characterization of Besov spaces by using a wavelet basis.

Let  $1 \leq p, q \leq \infty$  and  $0 < \delta < r + 1$ , where  $r > 0$  denotes the wavelet smoothness parameter. The  $\mathcal{B}_{p,q}^\delta$ -norm of the wavelet decomposed function  $f$  given in (2.1) is equivalent to the norm

$$\|f\|_{\delta,p,q} = \begin{cases} \|(\beta_{-1,k})_k\|_{\ell_p} + \left[ \sum_{j \geq 0} 2^{jq(\delta + \frac{1}{2} - \frac{1}{p})} \|(\beta_{j,k})_k\|_{\ell_p}^q \right]^{1/q} & \text{if } q < \infty, \\ \|(\beta_{-1,k})_k\|_{\ell_p} + \sup_{j \geq 0} 2^{j(\delta + \frac{1}{2} - \frac{1}{p})} \|(\beta_{j,k})_k\|_{\ell_p} & \text{if } q = \infty. \end{cases}$$

In this manuscript, we restrain our attention to the Besov space  $\mathcal{B}_{2,\infty}^\delta$ . Indeed, this space includes Sobolev spaces  $S_2^\delta$ , Besov spaces  $\mathcal{B}_{2,q}^\delta$  for  $q \geq 1$  and Besov spaces  $\mathcal{B}_{p,\infty}^\delta$  for  $p \geq 2$  for which the functions are compactly supported (see Triebel [107] or DeVore and Lorentz [34]). The class of Besov spaces provides a useful tool to classify wavelet decomposed signals with respect to their regularity and sparsity properties (see Härdle *et al.* [57]).

We also consider a weaker version of the Besov spaces defined for  $p > 0$  by

$$\mathcal{W}_p^* = \left\{ f = \sum_{j \geq -1} \sum_k \beta_{(j,k)} \psi_{j,k} \in \mathbb{L}_2(\mathbb{R}) : (\beta_{(j,k)})_{j,k} \in w\ell_p \right\},$$

where

$$w\ell_p = \left\{ \theta = (\theta_\ell)_\ell : \sup_{s>0} s^p \sum_\ell \mathbf{1}_{|\theta_\ell|>s} < \infty \right\}.$$

Whereas the spaces  $\mathcal{B}_{2,\infty}^\delta$  constitute an ideal class to measure the regularity of a wavelet decomposed signal  $f$ , the spaces  $\mathcal{W}_p^*$  constitute an ideal class to measure the sparsity of the function  $f$ . Indeed, we can rewrite

$$w\ell_p = \left\{ \theta = (\theta_\ell)_\ell : \sup_{\ell \in \mathbb{N}^*} \ell^{1/p} |\theta|_{(\ell)} < \infty \right\},$$

where the sequence  $(|\theta|_{(\ell)})_\ell$  is the non-increasing rearrangement of  $\theta$ :  $|\theta|_{(1)} \geq |\theta|_{(2)} \geq \dots \geq |\theta|_{(\ell)} \geq \dots$ . This condition gives a polynomial control of the decreasing rate of the sequence

$(|\beta|_{(\ell)})_{\ell}$ . Thus, a function belonging to  $\mathcal{B}_{2,\infty}^{\delta} \cap \mathcal{W}_p^*$  is more regular when  $\delta$  increases and is sparser when  $p$  decreases. We mention that pioneering works about the role of weak Besov spaces in statistics are due to Donoho [35], Johnstone [64] and Donoho and Johnstone [39]. There exists an embedding between Besov and weak Besov spaces:

$$\mathcal{B}_{2,\infty}^{\delta} \subset \mathcal{W}_{\frac{2}{1+2\delta}}^* \cap \mathcal{B}_{2,\infty}^{\frac{\delta}{1+2\delta}}.$$

See Kerkycharian and Picard [66] and Rivoirard [94, 95] for more details and for extensions in a more general setting.

In this manuscript, we actually consider Besov balls and weak Besov bodies, defined as follows. For any  $R > 0$ , Besov balls with radius  $R$  are defined by

$$\mathcal{B}_{2,\infty}^{\delta}(R) = \{f \in \mathbb{L}_2(\mathbb{R}) : \|f\|_{\delta,2,\infty} \leq R\}$$

and for  $R' > 0$ , weak Besov bodies with radius  $R'$  are defined by

$$\mathcal{W}_p^*(R') = \left\{ f \in \mathbb{L}_2(\mathbb{R}) : \sup_{s>0} s^p \sum_{(j,k)} \mathbf{1}_{|\beta_{(j,k)}|>s} \leq R'^p \right\}.$$

## 2.2 An introduction to the theory of $U$ -statistics

$U$ -statistics are a useful tool. The basic theory of  $U$ -statistics was developed by Hoeffding [60], one of the pioneers of nonparametric statistics, in connection with unbiased statistics. Detailed expositions of the general topic may be found in the books of van der Vaart [112] and de la Peña and Giné [32].

### 2.2.1 Definition and elementary examples

**Definition 2.2.** Given a sequence of independent and identically distributed (i.i.d.) random variables  $(X_i)_{i \geq 1}$  with values in a measurable space  $(\mathbb{X}, \mathcal{X})$  and a measurable function  $h : \mathbb{X}^m \rightarrow \mathbb{R}$ , the  $U$ -statistics of order  $m$  and kernel  $h$  based on the sequence  $(X_i)_{i \geq 1}$  are

$$U_n(h) = \frac{(n-m)!}{n!} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \quad m \leq n, \quad (2.2)$$

where  $I_n^m = \{(i_1, \dots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}$ .

In particular, if  $m = 1$ , then

$$U_n(h) = \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (2.3)$$

is a sum of i.i.d. random variables. Therefore, the  $U$ -statistics defined by (2.2) can be regarded as functional algebraic generalizations of an ordinary mean value of the form (2.3).

If the kernel  $h$  is symmetric in its arguments,  $U_n(h)$  has the equivalent form

$$U_n(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in SI_n^m} h(X_{i_1}, \dots, X_{i_m}), \quad m \leq n,$$

where  $SI_n^m = \{(i_1, \dots, i_m) : i_j \in \mathbb{N}, 1 \leq i_1 < i_2 < \dots < i_m \leq n\}$ .

We can cite some elementary examples of  $U$ -statistics.

**Example 2.1.** If  $m = 1$ ,  $U_n(h)$  is a simple sum of i.i.d. variables. For instance, if  $h(x) = x^k$  for  $k = 1, 2, \dots$ , we find the empirical mean and the empirical moments of order  $k$  for  $k \geq 2$ :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 2, 3, \dots$$

**Example 2.2.** For estimating  $\mu^m$ , where  $\mu = \mathbb{E}(X_1)$ , for  $m \in \mathbb{N}^*$ , we use the symmetric kernel  $h(x_1, \dots, x_m) = x_1 \dots x_m$  and we obtain the following  $U$ -statistic, unbiased for  $\mu^m$ :

$$U_n = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in SI_n^m} X_{i_1} \dots X_{i_m}.$$

**Example 2.3.** Since the variance  $\sigma^2 = \text{Var}(X_1)$  can be written  $\sigma^2 = \frac{1}{2} \mathbb{E}[(X_1 - X_2)^2]$ , the  $U$ -statistic defined by

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \frac{1}{2} (X_i - X_j)^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

is an unbiased estimator for  $\sigma^2$ , based on the symmetric kernel  $h(x_1, x_2) = (x_1 - x_2)^2/2$ .

From the very beginning, the extensive studies concerning  $U$ -statistics have lead to a manifold of results, including limit theorems, tail inequalities and also statistical applications, influenced by the classical theory of independent random variables. We propose, in the sequel, some exponential and moment inequalities for  $U$ -statistics and specifically for degenerate  $U$ -statistics. The  $U$ -statistic  $U_n(h)$  (or its kernel function  $h$ ) is called degenerate if for all  $1 \leq i \leq m$ ,

$$\mathbb{E}(h(X_1, \dots, X_m) | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m) = 0.$$

More precisely, in this case, we say that the  $U$ -statistic  $U_n(h)$  is canonical or completely degenerate. Indeed, we can also define nondegenerate  $U$ -statistics and degenerate  $U$ -statistics with an order of degeneracy  $r$  such that  $2 \leq r < m$  (for instance, see [32, 112]).

## 2.2.2 Some exponential and moment inequalities for $U$ -statistics

### 2.2.2.1 $U$ -statistics of order one

Since a  $U$ -statistic of order one is quite simply a sum of i.i.d. variables, classical concentration inequalities are available (see [77] for a review). In this manuscript, we have for instance opted for exponential Bernstein's inequality (for instance, see Proposition 2.9 of [77]) to control in probability a  $U$ -statistic of order one and for classical Rosenthal's inequality (for instance, see Proposition 10.2 of [57]) to control the moments of a  $U$ -statistic of order one. The following result is a particular case of Rosenthal's inequality, with a well known moment bound for sums of centered independent random variables.

**Theorem 2.1** (Lemma 8.1.4 of [32]). *Let  $(X_i)_{i \geq 1}$  be a sequence of i.i.d. real random variables such that  $\mathbb{E}(|X_i|^p) < \infty$  for some  $p \geq 2$  and  $\mathbb{E}(X_i) = 0$ . Then,*

$$\mathbb{E} \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right|^p \right) \leq C_p n^{p/2} \mathbb{E}(|X_1|^p),$$

for a positive universal constant  $C_p$  depending only on  $p$ .

We precise that Bernstein and Rosenthal inequalities have been extended to degenerate  $U$ -statistics of any order by Giné *et al.* [48]. Most of exponential and moment inequalities are first obtained for  $U$ -statistics of order two and are generalized to higher orders while this is possible.

### 2.2.2.2 $U$ -statistics of order two

For degenerate  $U$ -statistics of order two, we propose an exponential inequality obtained by Houdré and Reynaud-Bouret [61]. Let  $T_1, \dots, T_n$  be independent real random variables. We consider for all integer  $n \geq 2$ ,

$$U_n = \sum_{i=2}^n \sum_{j=1}^{i-1} g_{i,j}(T_i, T_j),$$

where the  $g_{i,j} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  are Borel measurable functions satisfying

$$\mathbb{E}(g_{i,j}(T_i, T_j)|T_i) = 0 \quad \text{and} \quad \mathbb{E}(g_{i,j}(T_i, T_j)|T_j) = 0,$$

namely  $U_n$  is a degenerate  $U$ -statistic of order two. We introduce the following quantities  $A$ ,  $B$ ,  $C$  and  $D$ :

- $A = \max_{i,j} \|g_{i,j}\|_\infty$ ,
- $B^2 = \max \left\{ \sup_{t,i} \left( \sum_{j=1}^{i-1} \mathbb{E}(g_{i,j}(t, T_j)^2 | T_j) \right), \sup_{t,j} \left( \sum_{i=j+1}^n \mathbb{E}(g_{i,j}(T_i, t)^2 | T_i) \right) \right\}$ ,
- $C^2 = \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}(g_{i,j}(T_i, T_j)^2)$  (that represents the true variance of  $U_n$ ), and
- $D = \sup \left\{ \mathbb{E} \left( \sum_{i=2}^n \sum_{j=1}^{i-1} g_{i,j}(T_i, T_j) a_i(T_i) b_j(T_j) \right) : \mathbb{E} \left( \sum_{i=2}^n a_i(T_i)^2 \right) \leq 1, \mathbb{E} \left( \sum_{j=1}^{n-1} b_j(T_j)^2 \right) \leq 1 \right\}$ .

The following result is a sharp exponential inequality, with precise constants.

**Theorem 2.2** (Theorem 3.4 of [61]). *For all  $\varepsilon, u > 0$ ,*

$$\mathbb{P}(U_n \geq 2(1 + \varepsilon)^{3/2} C \sqrt{u} + \eta(\varepsilon) D u + \beta(\varepsilon) B u^{3/2} + \gamma(\varepsilon) A u^2) \leq 2.77 e^{-u},$$

where  $\eta$ ,  $\beta$  and  $\gamma$  are positive functions of  $\varepsilon$  (exactly defined in [61]).

There exists also a Poisson version of this theorem (see Theorem 4.2 of [61]).

### 2.2.2.3 $U$ -statistics of upper order

We can extend the moment inequality established in Theorem 2.1 to degenerate  $U$ -statistics.

**Theorem 2.3** (Theorem 8.1.6 of [32]). *Let  $(X_i)_{i \geq 1}$  be a sequence of i.i.d. random variables with values in a measurable space  $(\mathbb{X}, \mathcal{X})$ . Let  $p \geq 2$  and let  $h : \mathbb{X}^k \rightarrow \mathbb{R}$  be a measurable function of  $k$  variables for some  $k \geq 2$  such that*

$$\mathbb{E}(|h(X_1, \dots, X_k)|^p) < \infty \quad \text{and} \quad \mathbb{E}(h(X_1, \dots, X_k) | X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k) = 0,$$

for all  $1 \leq j \leq k$ . Then,

$$\mathbb{E} \left( \max_{k \leq n \leq m} \left| \sum_{(i_1, \dots, i_k) \in I_n^k} h(X_{i_1}, \dots, X_{i_k}) \right|^p \right) \leq C_{k,p} m^{kp/2} \mathbb{E}(|h(X_1, \dots, X_k)|^p),$$

for all  $m \geq k$  and for some positive universal constant  $C_{k,p}$  depending only on  $k$  and  $p$ .

### 2.2.2.4 Decoupling for $U$ -statistics

In probability and statistics, decoupling is a reduction of a sample statistic to an average of the statistic evaluated on several independent sequences of the random variable. This sum, conditioned on all but one of the independent sequences becomes a sum of independent random variables. Consequently, a decoupling inequality permits the transfer of some results for sums of independent random variables, named also decoupled  $U$ -statistics, to the case of  $U$ -statistics.

By considering sequences  $(X_i^{(k)})_{i \geq 1}$ ,  $k = 1, \dots, m$  that are independent copies of the original sequence  $(X_i)_{i \geq 1}$ , a decoupling inequality consists in the establishment of an inequality between the quantities

$$\mathbb{E} \left( \Phi \left( \left| \sum_{I_n^m} h(X_{i_1}, \dots, X_{i_m}) \right| \right) \right) \quad \text{and} \quad \mathbb{E} \left( \Phi \left( \left| \sum_{I_n^m} h(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}) \right| \right) \right),$$

up to a constant that depends only on  $m$ , where  $\Phi$  is a nonnegative function. Usually, we consider the convex function  $\Phi(x) = |x|^p$  for  $p \geq 1$  to control the moments of a  $U$ -statistic, or  $\Phi(x) = \mathbf{1}_{|x| \geq t}$  to control in probability a  $U$ -statistic.

For instance, the following moment inequality proposed by Giné *et al.* [48] is stated for decoupled  $U$ -statistics.

**Theorem 2.4** (Theorem 3.3 of [48]). *There exists a universal constant  $K < \infty$  such that, if  $g_{i,j}$  are bounded degenerated kernels of two variables for the independent random variables sequences  $(X_i^{(1)})_{i \geq 1}$  and  $(X_j^{(2)})_{j \geq 1}$  and if  $A, B, C$  and  $D$  are as defined in Section 2.2.2.2, then*

$$\mathbb{E} \left( \left| \sum_{1 \leq i, j \leq n} g_{i,j} \left( X_i^{(1)}, X_j^{(2)} \right) \right|^p \right) \leq K^p \left[ p^{p/2} C^p + p^p D^p + p^{3p/2} B^p + p^{2p} A^p \right],$$

for all  $p \geq 2$ .

Theorem 3.3 of [48] also provided an equivalent exponential inequality (obtained easily by integration of tail probabilities). We precise that Theorem 2.2 recovers the exponential bound established by Giné *et al.* [48], with estimates on the constants.

De la Peña and Montgomery-Smith [33] got the following decoupling inequality (see Theorem 1 of [33] for  $U$ -statistics of upper order): for all  $n \geq 2$ ,  $t > 0$ , there exists a numerical constant  $C_2$  such that

$$\mathbb{P} \left( \left| \sum_{I_n^2} g_{i,j} \left( X_i^{(1)}, X_j^{(1)} \right) \right| \geq t \right) \leq C_2 \mathbb{P} \left( C_2 \left| \sum_{I_n^2} g_{i,j} \left( X_i^{(1)}, X_j^{(2)} \right) \right| \geq t \right)$$

and reverse bound holds under a symmetry condition.

Consequently, by combining the results of Giné *et al.* [48] and de la Peña and Montgomery-Smith [33], we obtain a similar upper bound as the one established in Theorem 2.4, up to a constant for real non-decoupled  $U$ -statistics:

$$\mathbb{E} \left( \left| \sum_{1 \leq i, j \leq n} g_{i,j} \left( X_i^{(1)}, X_j^{(1)} \right) \right|^p \right) \leq C_2 K^p \left[ p^{p/2} C^p + p^p D^p + p^{3p/2} B^p + p^{2p} A^p \right],$$

with  $C_2$  an absolute constant.

# Chapitre 3

## Adaptive estimation in a Poissonian interactions model

This chapter deals with the study of dependencies between two given events modeled by point processes. In particular, we focus on the context of DNA to detect favored or avoided distances between two given motifs along a genome suggesting possible interactions at a molecular level. For this, we naturally introduce a so-called reproduction function  $h$  that allows to quantify the favored positions of the motifs and which is considered as the intensity of a Poisson process. Our first interest is the estimation of this function  $h$  assumed to be well localized. The estimator  $\tilde{h}$  based on random thresholds achieves an oracle inequality. Then, minimax properties of  $\tilde{h}$  on Besov balls  $\mathcal{B}_{2,\infty}^s(R)$  are established. Some simulations are provided, allowing the calibration of tuning parameters from a numerical point of view and proving the good practical behavior of our procedure. Finally, our method is applied to the analysis of the influence between gene occurrences along the *E. coli* genome and occurrences of a motif known to be part of the major promoter sites for this bacterium.

*This chapter is the full version (with detailed proofs) of the paper Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data. Accepted for publication in the Scandinavian Journal of Statistics, 2012.*

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>46</b>
<b>3.2</b>	<b>Estimation procedure</b>	<b>49</b>
3.2.1	Methodology	49
3.2.2	Results: oracle inequality and minimax rate	51
3.2.3	Comments on the estimator of $\beta_\lambda$ , on the threshold $\eta_\lambda$ and on the results	52
<b>3.3</b>	<b>Implementation</b>	<b>53</b>
3.3.1	Practical procedure	54
3.3.2	Cascade algorithm	55
3.3.3	Choice of the parameters	56

3.3.4	Simulation (with the robustness study)	57
3.3.5	Some extensions of our model	59
<b>3.4</b>	<b>Application to genomic data</b>	<b>60</b>
<b>3.5</b>	<b>Conclusion</b>	<b>63</b>
<b>3.6</b>	<b>Proofs</b>	<b>65</b>
3.6.1	Proof of Proposition 3.1	65
3.6.2	Proof of Theorem 3.1	70
<b>3.A</b>	<b>Supplement: Proof of Theorem 3.2</b>	<b>71</b>
3.A.1	A technical lemma	72
3.A.2	Proof of Assumption (A1) of Theorem 3.3	73
3.A.3	Proof of Assumption (A2) of Theorem 3.3	78
3.A.4	Proof of Assumption (A3) of Theorem 3.3	93
3.A.5	Completion of the proof of Theorem 3.2	93

## 3.1 Introduction

The goal of the present chapter is to study the dependence between two given events modeled by point processes. We propose a general statistical approach to analyze any type of interaction, for instance, interactions between neurons in neurosciences or the comprehension of bankruptcies by contagion in economics. In particular, we focus on a model to study favored or avoided distances between patterns on a strand of DNA, which is an important task in genomics.

We are first interested in the modeling of the influence between two given motifs, a motif being defined as a sequence of letters in the alphabet  $\{\mathbf{a}, \mathbf{c}, \mathbf{g}, \mathbf{t}\}$ . This alphabet represents the four nucleotides bases of DNA: adenine, cytosine, guanine and thymine. Our aim is to model the dependence between motifs in order to identify favored or avoided distances between them, suggesting possible interactions at a molecular level. Because genomes are long (some 1 million bases) and motifs of interest are short (3 up to 20 bases), motif occurrences can be viewed as points along genomes. For convenience, we work in a continuous framework and then, the occurrences of a motif along a genome are modeled by a point process lying in the interval  $[0; T]$ , where  $T$  is the normalized length of the studied genome and will drive the asymptotic. We add that our model focuses on only one direction of interactions, that is to say we investigate the way a first given motif influences a second one. To study the influence of the second motif on the first one, we just invert their roles in the model.

We observe the occurrences of both given motifs (we presuppose interactions between them) and we assume that their distributions are as follows. The locations of the first motif are modeled by a  $n$ -sample of uniform random variables on  $[0; T]$ , denoted  $U_1, \dots, U_n$  and named parents. As the parameter  $T$ , the number  $n$  of parents will also drive the asymptotic. Then, each  $U_i$  gives birth independently to a Poisson process  $N^i$  with intensity the function  $t \mapsto h(t - U_i)$  with respect to the Lebesgue measure on  $\mathbb{R}$  (for instance, see [67]), which models the locations of the second motif. We consequently observe the aggregated process

$$N = \sum_{i=1}^n N^i \quad \text{with intensity the function} \quad t \mapsto \sum_{i=1}^n h(t - U_i) \quad (3.1)$$

and the points of the process  $N$  are named children. But in this model, for any child we do not observe which parent gives birth to him. The unknown function  $h$  is so-called reproduction function. Our goal is then to estimate  $h$  with the observations of the  $U_i$ 's and realizations of  $N$ .

Such a modeling of locations of the first motif is linked to the work on the distribution of words in DNA sequences of Schbath and coauthors (for instance, see [88, 97, 102]). Indeed, the first motif of interest is a rare word and is modeled by a homogeneous Poisson process  $N^0$  on  $[0; T]$ . Thus, conditionally to the event "the number of points falling into  $[0; T]$  is  $n$ ", the points of the process  $N^0$  (i.e. the parents) obey the same law as a  $n$ -sample of uniform random variables on  $[0; T]$ . Moreover, with very high probability,  $n$  is proportional to  $T$  and this constitutes the asymptotic considered in genomics, to which we will refer as the "DNA case". With our model (considering a uniform law on the parents), we can also take into consideration the cases  $n \ll T$  (parents are far away with respect to each other and one can almost identify which points are the children of a given parent) and  $n \gg T$  (parents are too close to each other, which leads to hard statistical problems).

If  $n = 1$ , the purpose is to estimate the intensity of only one Poisson process. Many adaptive methods have been proposed to deal with Poisson intensity estimation. For instance, Rudemo [100] studied data-driven histogram and kernel estimates based on the cross-validation method. Donoho [35] fitted the universal thresholding procedure proposed by Donoho and Johnstone [37] by using the Anscombe's transform. Kolaczyk [68] refined this idea by investigating the tails of the distribution of the noisy wavelet coefficients of the intensity. By using model selection, other optimal estimators have been proposed by Reynaud-Bouret [90] or Willett and Nowak [114]. Reynaud-Bouret and Rivoirard [91] proposed a data-driven thresholding procedure that is near optimal under oracle and minimax points of view, with as few support assumptions as possible (the support of the intensity  $h$  may be unknown or not finite), unlike previous methods that need to assume that the intensity has a known bounded support.

We notice that the reproduction function  $h$  can be also viewed as the intensity of a Cox process (for instance, see [29]) where the covariates are the parents  $U_1, \dots, U_n$ . Comte *et al.* [26] proposed an original estimator of the conditional intensity of a Cox process (more generally, a marker-dependent counting process). Using model selection methods, they prove that their estimator satisfies an oracle inequality and has minimax properties. Note that we consider here point processes on the real line. Some aspects of similar spatial processes are studied in a parametric way [87], for instance.

Some work has been done to study the statistical dependence between motif occurrences, by using a Hawkes' model (see [59]). The Hawkes' process  $(N_t)_{t \in \mathbb{R}}$  is defined by its intensity, which satisfies

$$\lambda(t) = \left( \xi + \sum_{T \in N} g(t - T) + \sum_{U \in N^0} h(t - U) \right)_+, \quad (3.2)$$

where  $\xi$  is a positive parameter,  $(N_t^0)_{t \in \mathbb{R}}$  is a second point process and  $g$  and  $h$  are functions allowed to take negative values, which can model (self-)inhibition (see paragraph 3.3.5.2), with support on  $\mathbb{R}_+$  and  $\int g < 1$  (for instance, see [54] for more details). We have taken the positive part to ensure that the intensity remains positive. The first part of (3.2) describes the phenomenon of spontaneous apparition (a child can be an orphan), the second part the phenomenon of self-excitation (a child can give birth to another child) and the third part the interactions with another type of event (the parents here). In Gusto and Schbath's article [54], the framework consists in modeling the occurrences of two motifs by a Hawkes process: our framework can be viewed when the support of  $h$  is in  $\mathbb{R}_+$  as a very particular case of theirs. Their method, called FADO, uses maximum likelihood estimates of the coefficients of  $h$  on a Spline basis coupled



with an AIC criterion. However, even if the FADO procedure is quite effective and can manage interactions between two types of events, spontaneous apparition and self-excitation, there are several drawbacks. Indeed this procedure is actually a parametric estimation method coupled with a classical AIC criterion which behaves poorly for complex families of models. Moreover, FADO involves sparsity issues. Indeed, our feeling is that if interaction exists, say around the distance  $d$  bases, the function  $h$  to estimate should take large values around  $d$  and if there is no biological reason for any other interaction, then  $h$  should be null anywhere else. However, if the FADO estimate takes small values in this last situation, it does not vanish, which can result in misleading biological interpretations (see [93]). Finally, in this model, the occurrences of the first motif only depend on the past occurrences of first and second motifs. Reynaud-Bouret and Schbath [93] have proposed an alternative based on model selection principle for Hawkes processes that solves the sparsity problem. Their estimate satisfies an oracle inequality and has adaptive minimax properties with respect to certain classes of functions. But it manages only one motif whereas it is necessary to treat interaction with another type of events and the method has a high computational cost even for a small number of models. Note that Hawkes processes have a long story of parametric inference (see [29, 82, 83]). In particular, for genomic data, Carstensen *et al.* [21] recently deal with multivariate Hawkes process models in order to model the joint occurrences of multiple transcriptional regulatory elements (TREs) along the genome that are capable of providing new insights into dependencies among elements involved in transcriptional regulation.

In this chapter, the proposed model is simple. Each child comes from one parent (no orphan and no child who is a parent), that is to say we do not take into account the phenomenons of spontaneous apparition and self-excitation, contrary to Hawkes process models. But it brings novelties. To estimate the reproduction function  $h$ , we propose a nonparametric method, using a wavelet thresholding rule that will compensate sparsity issues of the FADO method. Furthermore, our model treats interaction between two types of events, with a possible influence of the past occurrences but also future occurrences. Then, there is the presence of a double asymptotic: the normalized length of the studied genome  $T$  and the number  $n$  of parents, which is not usual. In the biological context, it is not acceptable assuming to know each child's parent. Our model, via the reproduction function  $h$ , allows to quantify the favored locations of children in relation to their parent, even if one cannot attribute a child to a parent before the statistical inference. First we provide in this chapter theoretical results and we derive oracle inequalities and minimax rates showing that our method achieves good theoretical performances. The proofs of these results are essentially based on concentration inequalities and on exponential and moment inequalities for  $U$ -statistics (see [32, 48, 61]). Secondly some simulations are carried out to validate our procedure and an application on real data (*Escherichia coli* genome) is proposed. The procedure provides satisfying reconstructions, overcomes the problems raised by the FADO method and agrees with the knowledge of the considered biological mechanism. For these numerical aspects, we have used a low computational complexity cascade algorithm.

In Section 3.2, we describe the method based on wavelet thresholding procedures. Then we give the results and discuss the properties of our procedure for the oracle and minimax approaches. Section 3.3 is devoted to the implementation of our method and provides simulations. The cascade algorithm is presented in Subsection 3.3.2. Section 3.4 presents the application on the complete *Escherichia coli* genome. Proofs of the main results can be found in Section 3.6. An appendix is also provided for more detailed proofs (see Section 3.A).

## 3.2 Estimation procedure

### 3.2.1 Methodology

To estimate the reproduction function, we assume that  $h$  belongs to  $\mathbb{L}_1(\mathbb{R})$  and  $\mathbb{L}_\infty(\mathbb{R})$ . Consequently, we can consider the decomposition of  $h$  on a particular biorthogonal wavelet basis, built by Cohen *et al.* [23], that we can describe as follows. We set  $\phi = \mathbf{1}_{[0,1]}$  the analysis father wavelet and there exist three functions  $\psi$ ,  $\tilde{\phi}$  and  $\tilde{\psi}$  with the following properties:

- $\tilde{\phi}$  and  $\tilde{\psi}$  are compactly supported,
- $\psi$  is compactly supported and is a piecewise constant function,
- $\left\{ (\phi_k, \psi_{j,k})_{j \geq 0, k \in \mathbb{Z}}, (\tilde{\phi}_k, \tilde{\psi}_{j,k})_{j \geq 0, k \in \mathbb{Z}} \right\}$  is a biorthogonal family where for any  $j \geq 0$ , for any  $k \in \mathbb{Z}$  and for any  $x \in \mathbb{R}$ ,

$$\phi_k(x) = \phi(x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$$

and

$$\tilde{\phi}_k(x) = \tilde{\phi}(x - k), \quad \tilde{\psi}_{j,k}(x) = 2^{j/2} \tilde{\psi}(2^j x - k).$$

The Haar basis, used in practice, can be viewed as a particular biorthogonal wavelet basis, by setting  $\tilde{\phi} = \phi$  and  $\tilde{\psi} = \psi = \mathbf{1}_{[\frac{1}{2}, 1]} - \mathbf{1}_{[0, \frac{1}{2}]}$ . The interest of biorthogonal wavelet bases lies in the fact that they can have the following property: analysis wavelets  $\phi_k$  and  $\psi_{j,k}$  are piecewise constant functions and reconstruction wavelets  $\tilde{\phi}_k$  and  $\tilde{\psi}_{j,k}$  are typically smooth functions. The wavelet decomposition of  $h \in \mathbb{L}_2(\mathbb{R})$  has the following form:

$$h = \sum_{k \in \mathbb{Z}} \alpha_k \tilde{\phi}_k + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{j,k} \tilde{\psi}_{j,k}, \quad (3.3)$$

where for any  $j \geq 0$  and any  $k \in \mathbb{Z}$ ,

$$\alpha_k = \int_{\mathbb{R}} h(x) \phi_k(x) dx, \quad \beta_{j,k} = \int_{\mathbb{R}} h(x) \psi_{j,k}(x) dx.$$

This kind of decomposition has already been used in thresholding methods by Juditsky and Lambert-Lacroix [65], Reynaud-Bouret and Rivoirard [91], and Reynaud-Bouret *et al.* [92].

To shorten mathematical expressions, we set

$$\Lambda = \{ \lambda = (j, k) : j \geq -1, k \in \mathbb{Z} \},$$

and for any  $\lambda \in \Lambda$ ,

$$\varphi_\lambda = \begin{cases} \phi_k & \text{if } \lambda = (-1, k) \\ \psi_{j,k} & \text{if } \lambda = (j, k) \text{ with } j \geq 0 \end{cases}, \quad \tilde{\varphi}_\lambda = \begin{cases} \tilde{\phi}_k & \text{if } \lambda = (-1, k) \\ \tilde{\psi}_{j,k} & \text{if } \lambda = (j, k) \text{ with } j \geq 0 \end{cases}$$

and similarly

$$\beta_\lambda = \begin{cases} \alpha_k & \text{if } \lambda = (-1, k) \\ \beta_{j,k} & \text{if } \lambda = (j, k) \text{ with } j \geq 0 \end{cases}.$$

Then (3.3) can be rewritten as

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda \quad \text{with} \quad \beta_\lambda = \int_{\mathbb{R}} h(x) \varphi_\lambda(x) dx \quad (3.4)$$

and now, we have to estimate these wavelet coefficients.

For this purpose we define for any  $\lambda$  in  $\Lambda$ ,  $\hat{\beta}_\lambda$  the estimate of  $\beta_\lambda$  as follows

$$\hat{\beta}_\lambda = \frac{G(\varphi_\lambda)}{n}, \quad \text{with} \quad G(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] dN_t, \quad (3.5)$$

where  $\pi$  is the uniform distribution on  $[0; T]$  and  $\mathbb{E}_\pi(\varphi_\lambda(t - U))$  denotes the expectation of  $\varphi_\lambda(t - U)$  where  $U \sim \pi$  (an independent copy of  $U_1, \dots, U_n$ ).

If  $n = 1$ , we obtain the natural estimators of the  $\beta_\lambda$ 's in the case of only one Poisson process on the real line (see [91]). These estimates are unbiased and their variance can be controlled (see Subsection 3.2.3 for a more detailed discussion):

**Proposition 3.1.** *For all  $\lambda = (j, k)$  in  $\Lambda$ ,*

$$\mathbb{E}(G(\varphi_\lambda)) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx,$$

*i.e.  $\hat{\beta}_\lambda$  is an unbiased estimator for  $\beta_\lambda$ . Furthermore, its variance is upper bounded as follows:*

$$\text{Var}(\hat{\beta}_\lambda) \leq C \left\{ \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2} \right\}$$

and

$$\sup_{\lambda \in \Lambda} \text{Var}(\hat{\beta}_\lambda) \leq C' \left\{ \frac{1}{n} + \frac{n}{T^2} \right\}, \quad (3.6)$$

where  $C$  and  $C'$  depend on  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$  and  $\|\psi\|_2$ .

From now on, we assume that  $h$  is compactly supported in  $[-A; A]$ , with  $A$  a positive real number, not depending on  $n$  and  $T$ . This quantity  $A$  can denote the maximal memory along DNA sequences (this is chosen by the biologists (see [54]), depending on the underlying biological process they have in mind). Furthermore, the properties of the biorthogonal wavelet bases introduced previously allow us to assume that we know a positive real number  $M$  such that the support of  $\psi$  is contained in  $[-M; M]$ .

First, we introduce the following deterministic subset  $\Gamma$  of  $\Lambda$

$$\Gamma = \{ \lambda = (j, k) \in \Lambda : -1 \leq j \leq j_0, k \in \mathcal{K}_j \},$$

where  $j_0$  a positive integer that will be fixed later and at each resolution level  $j$ , we denote  $\mathcal{K}_j$  the set of integers such that the intersection of the supports of  $\varphi_\lambda$  and  $h$  is not empty, with  $\lambda = (j, k)$ . More precisely,

$$\mathcal{K}_{-1} = \{ k \in \mathbb{Z} : -A + 1 \leq k \leq A \}$$

and for any  $j \geq 0$ ,

$$\mathcal{K}_j = \{ k \in \mathbb{Z} : -(2^j A + M) \leq k \leq 2^j A + M \}.$$

Straightforward computations lead to a cardinal of  $\Gamma$  of order  $2^{j_0}$ .

Then, given some parameter  $\gamma > 0$ , we define for any  $\lambda \in \Gamma$ , the threshold

$$\eta_\lambda = \sqrt{2\gamma j_0 \tilde{V} \left( \frac{\varphi_\lambda}{n} \right)} + \frac{\gamma j_0}{3} B \left( \frac{\varphi_\lambda}{n} \right) + \Delta \frac{N_{\mathbb{R}}}{n} \quad (3.7)$$

where  $\Delta$  is a positive quantity and  $N_{\mathbb{R}}$  is the number of points of the aggregated process  $N$  lying in  $\mathbb{R}$ . For theoretical results,

$$\Delta = d \left\{ \frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T} \right\}, \quad (3.8)$$

where  $d$  is a constant depending on  $\gamma$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  (we can find the exact value of  $d$  in (3.34) in Section 3.A). In (3.7), we set

$$B\left(\frac{\varphi_\lambda}{n}\right) = \frac{1}{n} B(\varphi_\lambda) = \frac{1}{n} \left\| \sum_{i=1}^n \left[ \varphi_\lambda(\cdot - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(\cdot - U)) \right] \right\|_\infty \quad (3.9)$$

and

$$\tilde{V}\left(\frac{\varphi_\lambda}{n}\right) = \frac{1}{n^2} \tilde{V}(\varphi_\lambda) = \frac{1}{n^2} \left( \hat{V}(\varphi_\lambda) + \sqrt{2\gamma j_0 \hat{V}(\varphi_\lambda) B^2(\varphi_\lambda) + 3\gamma j_0 B^2(\varphi_\lambda)} \right) \quad (3.10)$$

where

$$\hat{V}(\varphi_\lambda) = \int_{\mathbb{R}} \left( \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right)^2 dN_t. \quad (3.11)$$

Since they only depend on the observations, the numerical values of  $B(\varphi_\lambda)$ ,  $\hat{V}(\varphi_\lambda)$  and so  $\tilde{V}(\varphi_\lambda)$  defined respectively by (3.9), (3.11) and (3.10) can be exactly computed.

We denote  $\tilde{\beta}$  the estimator of  $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$  associated with the previous thresholding rule:

$$\tilde{\beta} = \left( \hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma} \right)_{\lambda \in \Lambda} \quad (3.12)$$

and finally, we set

$$\tilde{h} = \sum_{\lambda \in \Lambda} \tilde{\beta}_\lambda \tilde{\varphi}_\lambda \quad (3.13)$$

an estimator of  $h$  that only depends on the choice of  $\gamma$  and  $j_0$  fixed later. We now give theoretical results of our procedure. They are discussed in Subsection 3.2.3.

### 3.2.2 Results: oracle inequality and minimax rate

Our main result is an oracle one. Given a collection of procedures (for example, penalization, projection or thresholding), the oracle represents the ideal "estimator" among the collection. In our setting the oracle gives, for our thresholding rule, the coefficients that have to be kept. In our framework (see [37] and [91]), the "oracle estimator" is

$$\bar{h} = \sum_{\lambda \in \Gamma} \bar{\beta}_\lambda \tilde{\varphi}_\lambda, \quad \text{with} \quad \bar{\beta}_\lambda = \hat{\beta}_\lambda \mathbf{1}_{\text{Var}(\hat{\beta}_\lambda) < \beta_\lambda^2}.$$

This "estimator" is not a true estimator, of course, since it depends on  $h$ . The approach of optimal adaptation is to derive true estimators which achieve the same performance as the "oracle estimator". Our goal is now to compare the risk of  $\tilde{h}$  defined in Subsection 3.2.1 to the oracle risk:

$$\mathbb{E}(\|\tilde{h} - h\|_2^2) = \sum_{\lambda \in \Gamma} \mathbb{E}[(\hat{\beta}_\lambda \mathbf{1}_{\text{Var}(\hat{\beta}_\lambda) < \beta_\lambda^2} - \beta_\lambda)^2] + \sum_{\lambda \notin \Gamma} \beta_\lambda^2 = \sum_{\lambda \in \Gamma} \min(\text{Var}(\hat{\beta}_\lambda), \beta_\lambda^2) + \sum_{\lambda \notin \Gamma} \beta_\lambda^2.$$

**Theorem 3.1.** *We assume that  $n \geq 2$ ,  $j_0 \in \mathbb{N}^*$  such that  $2^{j_0} \leq n < 2^{j_0+1}$ ,  $\gamma > 2 \log 2$ . Then the estimator  $\tilde{h}$  defined in Subsection 3.2.1 satisfies*

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \left[ (\log n)^4 \times \frac{1}{n} + (\log n)^2 \times \frac{n}{T^2} \right] |m| \right\} + C_2 \left[ \frac{1}{n} + \frac{n}{T^2} \right],$$

where  $|m|$  is the cardinal of the set  $m$ ,  $C_1$  is a positive constant depending on  $\gamma$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  and  $C_2$  is a positive constant depending on  $A$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ , the compact support of  $\psi$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$ .

The following corollary establishes an oracle type inequality in the "DNA case".

**Corollary 3.1.** *Assume that  $n$  is proportional to  $T$ , under the assumptions of Theorem 3.1. Then the estimator  $\tilde{h}$  defined in Subsection 3.2.1 satisfies*

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{(\log T)^4}{T} |m| \right\} + \frac{C_2}{T},$$

where  $C_1$  is a positive constant depending on  $\gamma$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  and  $C_2$  is a positive constant depending on  $A$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ , the compact support of  $\psi$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$ .

Then, we establish a minimax result on Besov balls still with  $n$  is proportional to  $T$ . For any  $R > 0$  and  $s \in \mathbb{R}$  such that  $0 < s < r + 1$  (where  $r > 0$  denotes the wavelet smoothness parameter introduced in the description of the biorthogonal wavelet bases at the beginning of the current section), we consider the following Besov ball of radius  $R$ :

$$\mathcal{B}_{2,\infty}^s(R) = \left\{ f \in \mathbb{L}_2(\mathbb{R}) : f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda, \forall j \geq -1, \sum_{k \in \mathcal{K}_j} \beta_{(j,k)}^2 \leq R^2 2^{-2js} \right\}.$$

Now, let us state the upper bound of the risk of  $\tilde{h}$  when  $h$  belongs to  $\mathcal{B}_{2,\infty}^s(R)$ .

**Corollary 3.2.** *Let  $R > 0$  and  $s \in \mathbb{R}$  such that  $0 < s < r + 1$ . Assume that  $h \in \mathcal{B}_{2,\infty}^s(R)$  and  $n$  is proportional to  $T$ , under the assumptions of Theorem 3.1. Then the estimator  $\tilde{h}$  defined in Subsection 3.2.1 satisfies*

$$\mathbb{E} \left( \|\tilde{h} - h\|_2^2 \right) \leq C \left( \frac{(\log T)^4}{T} \right)^{\frac{2s}{2s+1}},$$

where  $C$  is a positive constant depending on  $\gamma$ ,  $A$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ , the compact support of  $\psi$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$ ,  $\|\psi\|_\infty$  and  $R$ .

### 3.2.3 Comments on the estimator of $\beta_\lambda$ , on the threshold $\eta_\lambda$ and on the results

Proposition 3.1 shows that the estimator  $\hat{\beta}_\lambda$  of  $\beta_\lambda$  defined by (3.5) is unbiased. Furthermore, the behavior of its variance is not usual, because two parameters  $n$  and  $T$  are involved. Nevertheless, when  $n$  is proportional to  $T$  ("DNA case" as explained in Introduction), the variance is bounded by  $1/T$  up to a constant, as for the Hawkes process (see [93]). When  $n \ll T$ , the variance is bounded by  $1/n$  up to a constant, which means that the apparition's distance between two parents is large enough to make their interactions insignificant for the statistical analysis. So in this case, our framework can be viewed as the observation of a  $n$ -sample of a Poisson process

with common intensity  $h$  (see [91]). Finally, when  $n \gg T$ , the variance deteriorates and is only bounded by  $n/T^2$  up to a constant, and in this case, the small apparition's distance between two parents leads to rough statistical issues hard to overcome.

Our method is based on thresholding procedures that have been introduced by Donoho and Johnstone [37]. They derive from the sufficiency to keep a small amount of the coefficients to have a good estimation of the function  $h$ . The threshold  $\eta_\lambda$  seems to be defined in a rather complicated manner but the first term:  $\sqrt{2\gamma j_0 \tilde{V}(\frac{\varphi_\lambda}{n})}$  looks like the universal threshold proposed by [37] in the Gaussian regression framework, by choosing  $\gamma$  close to 1 and  $j_0$  of order  $\log n$ . The universal threshold of [37] is defined by  $\eta_\lambda = \sqrt{2\sigma^2 \log n}$ , where  $\sigma^2$  (assumed to be known) is the variance of each noisy wavelet coefficient. In our setting,  $\text{Var}(\hat{\beta}_\lambda)$  depends on  $h$ , so it is (over)estimated by  $\tilde{V}(\frac{\varphi_\lambda}{n})$ . The other terms of the threshold (3.7) are unavoidable remaining terms which allow to obtain sharp concentration inequalities.

Then we have established Theorem 3.1, an oracle type inequality. As the expression between brackets is of the same order as the upper bound of  $\text{Var}(\hat{\beta}_\lambda)$  established in Proposition 3.1 (up to a logarithmic term), this theorem proves that the estimator  $\tilde{h}$  achieves satisfying theoretical properties.

In particular with  $n$  proportional to  $T$  ("DNA case"), the oracle type inequality obtained in Corollary 3.1 is similar to the one obtained by Theorem 1 of [93] where the Hawkes process is considered. Since  $n$  is proportional to  $T$ , this inequality is typical of classical oracle inequalities obtained in model selection (for example, see Theorem 2.1 of [91] where only one Poisson process on the real line is considered or more generally, see [77] for density estimation). Corollary 3.2 illustrates one more time the optimality of the procedure  $\tilde{h}$  but in the minimax setting. Indeed, still with  $n$  is proportional to  $T$ , the rate of the risk of  $\tilde{h}$  corresponds to the minimax rate, up to the logarithmic term, for estimation of a compactly supported intensity of a Poisson process (see [90]) or for a compactly supported density when we have  $n$  i.i.d. observations (see [41]).

Actually Corollaries 3.1 and 3.2 remain true if  $n = O(T)$  because the analysis of (3.6) shows that the upper bound of the variance is of order  $1/n$  when  $n = O(T)$ . In this case we recover classical rates. This is not the case when  $n \gg T$  since in this case  $\frac{1}{n} + \frac{n}{T^2}$  is equivalent to  $n/T^2$ .

Therefore, this chapter reveals two regimes:

- the case  $n = O(T)$  where classical rates are obtained which shows that our procedure is optimal;
- the case  $n \gg T$  where the main term in the oracle inequality of Theorem 3.1 is  $n/T^2$  (up to logarithmic terms). In this case, the nature of the statistical problem is very different due to the overlaps of the different shifted versions of  $h$  in the observed intensity, which leads to an ill-posed inverse problem. Whether the rate  $n/T^2$  is optimal remains an open question. This challenging problem will not be investigated in this thesis.

### 3.3 Implementation

From now on we consider the context of DNA, i.e.  $n$  is proportional to  $T$ . As mentioned in Introduction, we can assume that the parents are the points of a homogeneous Poisson process  $N^0$  on  $[0; T]$  with constant intensity  $\mu$  which allows to write  $n \simeq \mu T$ .

In this section, we specify a procedure for the computation of the family of random thresholds  $(\eta_\lambda)_{\lambda \in \Gamma}$  to reconstruct the reproduction function  $h$ . We also provide some simulations in order to calibrate parameters from a numerical point of view and to show the robustness of our procedure.

The programs have been coded in `Scilab 5.2` and are available in Section A.1 of Appendix A.

### 3.3.1 Practical procedure

The wavelet set-up associated with biorthogonal wavelet bases is considered. More precisely, we focus on the Haar basis where

$$\phi = \tilde{\phi} = \mathbf{1}_{[0;1]} \quad \text{and} \quad \psi = \tilde{\psi} = \mathbf{1}_{] \frac{1}{2}; 1]} - \mathbf{1}_{[0; \frac{1}{2}]}$$

and on a special case of spline systems given in Figure 3.1.

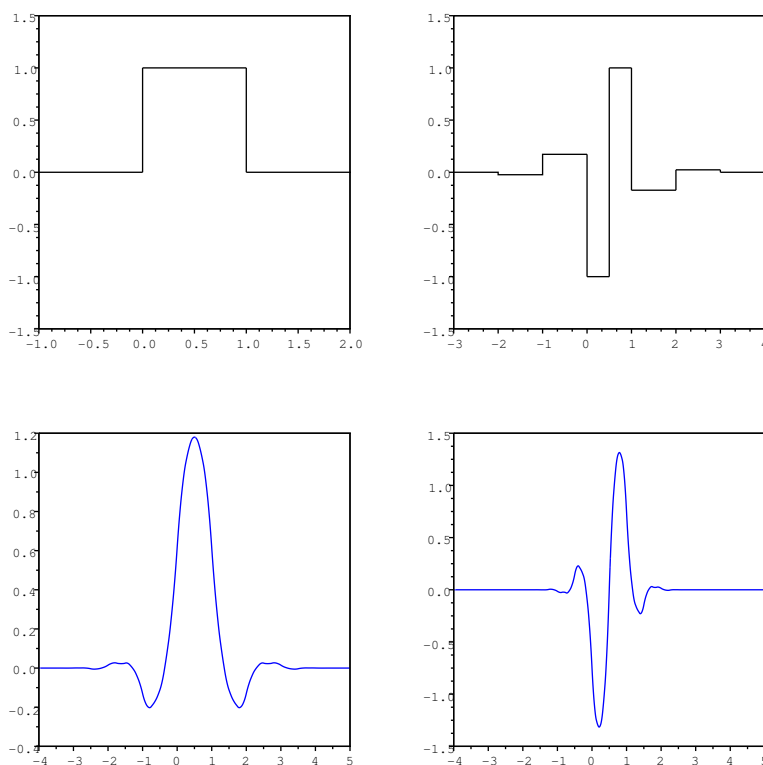


Figure 3.1: The Spline basis: top:  $\phi$  and  $\psi$ ; bottom:  $\tilde{\phi}$  and  $\tilde{\psi}$ .

The latter is called hereafter the Spline basis. We recall that  $h$  is assumed to be compactly supported in  $[-A; A]$ , with  $A$  a positive integer in practice.

We consider the thresholding rule  $\tilde{h}$  defined in Subsection 3.2.1 with

$$\Gamma = \{ \lambda = (j, k) \in \Lambda : -1 \leq j \leq j_0, k \in \mathcal{K}_j \},$$

and

$$\eta_\lambda = \sqrt{2\gamma j_0 \hat{V} \left( \frac{\varphi_\lambda}{n} \right) + \frac{\gamma j_0}{3} B \left( \frac{\varphi_\lambda}{n} \right) + \frac{\delta}{\sqrt{T}} \frac{N_{\mathbb{R}}}{n}}. \quad (3.14)$$

Observe that  $\eta_\lambda$  slightly differs from the threshold defined in (3.7). Indeed, by using equation (3.8) defining  $\Delta$ , with  $n$  proportional to  $T$ , we have replaced the quantity  $\Delta$  with  $\delta/\sqrt{T}$ . The

choice of the parameters  $\gamma$  and  $\delta$  from a practical point of view is discussed in Subsection 3.3.3. Moreover,  $\tilde{V}(\varphi_\lambda)$  is now replaced with  $\hat{V}(\varphi_\lambda)$ . There is no major difference in our simulations and this allows us to obtain computational simplifications. From a theoretical point of view, the ideal choice of the maximal resolution level  $j_0$  is given by Theorem 3.1, namely  $j_0$  is the positive integer such that  $2^{j_0} \leq n < 2^{j_0+1}$ . We will fix  $j_0 = 5$  in the sequel even if the large values of  $j_0$  give practically the same reconstruction, in particular to limit computation time.

Finally similarly to Subsection 3.2.1, we define  $\tilde{h}$  as (3.13) the estimator of  $h$  with the previous thresholding rule.

### 3.3.2 Cascade algorithm

A key point of the algorithm is the computation of the quantity

$$S(\varphi_\lambda)(t) = \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right], \quad \text{for all } t \in \mathbb{R} \quad (3.15)$$

that appears in  $\hat{\beta}_\lambda$ ,  $B(\varphi_\lambda)$  and  $\hat{V}(\varphi_\lambda)$ . We decompose it into two parts: a random "piecewise constant" part  $S_r(\varphi_\lambda) = \sum_{i=1}^n \varphi_\lambda(\cdot - U_i)$  and a deterministic (piecewise affine) part  $(n-1)\mathbb{E}_\pi(\varphi_\lambda(t - U))$ . Note that the deterministic part can easily be implemented with a low computational cost. This is not the case of the random "piecewise constant" part for which we have constructed a cascade algorithm, inspired by the pioneering work of Mallat [74]. We describe this algorithm in few words for the Haar basis. We use the following notations: for any  $j \geq 0$ , for any  $k \in \mathbb{Z}$ , for any  $x \in \mathbb{R}$ ,

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k),$$

where  $\phi_{j,k}$  are father wavelets and  $\psi_{j,k}$  mother wavelets. We have the following relationships between wavelets at level  $j$  and wavelets at level  $(j+1)$ :

$$\psi_{j,k} = \frac{\sqrt{2}}{2} (\phi_{j+1,2k+1} - \phi_{j+1,2k}) \quad \text{and} \quad \phi_{j,k} = \frac{\sqrt{2}}{2} (\phi_{j+1,2k} + \phi_{j+1,2k+1}). \quad (3.16)$$

We notice that only mother wavelets and the father wavelet of level  $j = 0$  (that corresponds to  $\varphi_\lambda$  with  $\lambda = (-1, k)$ ) are used to reconstruct the signal. The cascade algorithm is implemented as follows.

1. Compute  $S_r(\phi_{j_0,0})$ . Since  $S_r(\phi_{j_0,0})$  is a piecewise constant function, this computation gives a partition and the values of  $S_r(\phi_{j_0,0})$  on the intervals of the partition.
2. Shift by  $+2^{-j_0}k$  the intervals of the previous partition by keeping the same values on the partition to obtain  $S_r(\phi_{j_0,k})$  for any integer  $k$  in  $[-2^{j_0}A; 2^{j_0}A - 1]$ .
3. For any resolution level  $j$  going from  $(j_0 - 1)$  to  $0$ , in a decreasing way, compute  $S_r(\psi_{j,k})$  and  $S_r(\phi_{j,k})$  with expressions (3.16). The quantities  $S_r(\psi_{j,k})$  allow the reconstruction of the signal and the quantities  $S_r(\phi_{j,k})$  are transitional and will be used for the computations of the lower resolution level  $(j - 1)$ .
4. Also keep  $S_r(\phi_{0,k})$  because it is used for the reconstruction of the signal.

A similar algorithm has been implemented for the general Spline basis, using the filter-coefficients associated with the basis that can be found in [31].



### 3.3.3 Choice of the parameters

In this subsection, we deal with the tuning of the parameters  $\gamma$  and  $\delta$  in our procedure from a practical point of view. The question is: how to choose the optimal parameters? In order to calibrate the precise values of the parameters in the practical procedure, we consider the Haar wavelets basis and we work with two testing functions denoted 'Signal1' and 'Signal2' whose definitions are given in the following table:

'Signal1'	'Signal2'
$\nu \times \mathbf{1}_{[0;1]}$	$\nu \times \frac{8}{3} (\mathbf{1}_{[0.5;0.625]} + \mathbf{1}_{[1;1.25]})$

with  $\nu$ , the children's intensity, set to 4. We fix willfully  $A = 10$ . Such a choice of  $A$  (remember that  $[-A; A]$  is the support of  $h$ ) assumes that we do not know the support of functions. We recall that  $j_0 = 5$ .

Given  $T$ ,  $\mu$  the parents' intensity and a testing function, we denote  $R(\gamma, \delta)$  the quadratic risk of our procedure  $\tilde{h}$  (depending on  $(\gamma, \delta)$ ) defined in Subsection 3.3.1. Of course, we aim at finding values of  $(\gamma, \delta)$  such that this quadratic risk is minimal. The average over 100 simulations of  $R(\gamma, \delta)$  is computed providing an estimation of  $\mathbb{E}(R(\gamma, \delta))$ . This average risk, denoted  $\bar{R}(\gamma, \delta)$  and viewed as a function of the parameters  $(\gamma, \delta)$ , is plotted for  $(T, \mu) \in \{(10000, 0.1), (2000, 0.1), (2000, 0.5)\}$  and for the two signals considered previously: 'Signal1' and 'Signal2'.

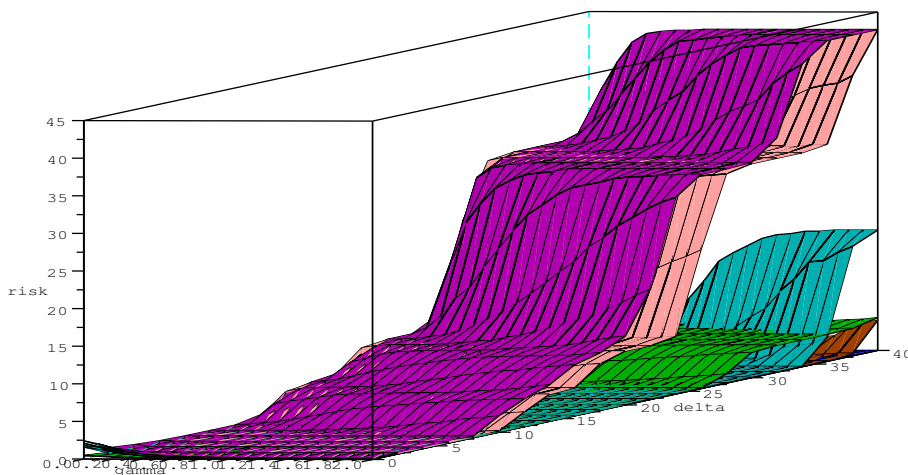


Figure 3.2: The function  $(\gamma, \delta) \mapsto \bar{R}(\gamma, \delta)$  for 'Signal1' and 'Signal2' for different values of  $T$  and  $\mu$ : 'Signal1' in ■ and 'Signal2' in ■ with  $(T, \mu) = (10000, 0.1)$ ; 'Signal1' in ■ and 'Signal2' in ■ with  $(T, \mu) = (2000, 0.1)$ ; 'Signal1' in ■ and 'Signal2' in ■ with  $(T, \mu) = (2000, 0.5)$ .

Figure 3.2 displays  $\bar{R}$  for 'Signal1' and 'Signal2' decomposed on the Haar basis. This figure allows to draw the following conclusion: for any  $(T, \mu) \in \{(10000, 0.1), (2000, 0.1), (2000, 0.5)\}$

and for 'Signal1' or 'Signal2',

$$\bar{R}(\gamma, \delta) \approx 0$$

for many values of  $(\gamma, \delta)$ . So, we observe a kind of "plateau phenomenon" and we choose  $(\gamma, \delta) = (0.18, 2.4)$ , a common value of several plateaus.

### 3.3.4 Simulation (with the robustness study)

#### 3.3.4.1 Some reconstructions

Reconstructions of the intensities of 'Signal1' and 'Signal2', in the Haar basis, are respectively given in Figure 3.3 and Figure 3.4 with the choice  $(\gamma, \delta) = (0.18, 2.4)$ . Note the good performance of our thresholding rule, in particular for  $T = 10000$  and  $\mu = 0.1$  (we have  $\mu T = 1000$  parents and  $\mu\nu T = 4000$  children in average), which corresponds to the real case treated in Section 3.4. Thus, we propose to take systematically  $(\gamma, \delta) = (0.18, 2.4)$  in our procedure  $\tilde{h}$  defined in Subsection 3.3.1.

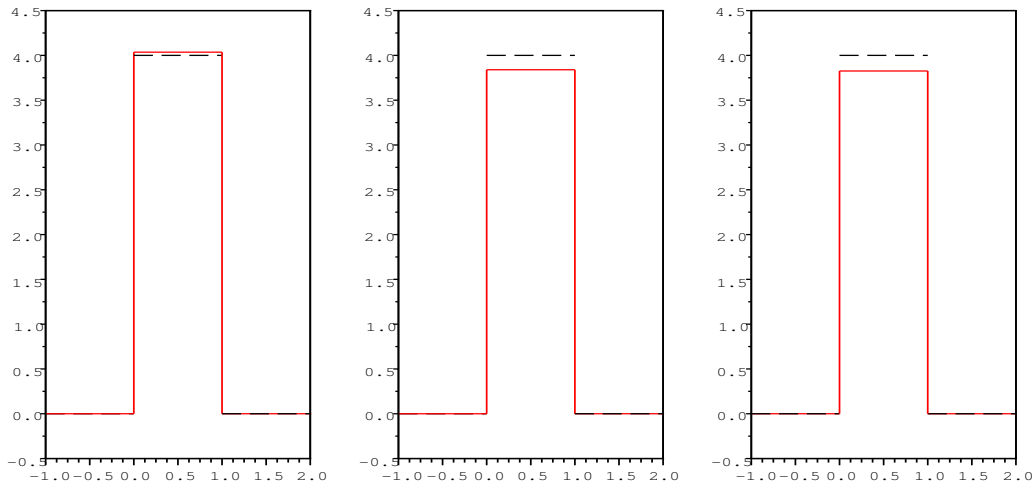


Figure 3.3: Reconstructions of 'Signal1' in the Haar basis (true: dotted line, estimate: solid line): left:  $(T, \mu) = (10000, 0.1)$ ; middle:  $(T, \mu) = (2000, 0.1)$ ; right:  $(T, \mu) = (2000, 0.5)$ .

Reconstructions of the intensities of 'Signal1' and 'Signal2', in the Spline basis, are given in Figure 3.5 with the choice  $(\gamma, \delta) = (0.18, 2.4)$ .

Figure 3.6 displays reconstructions of 'SignalN' whose definition is given in the following table:

<p>'SignalN'</p> $\nu \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
--

with  $\nu$ , the children's intensity, set to 4, in the two bases.

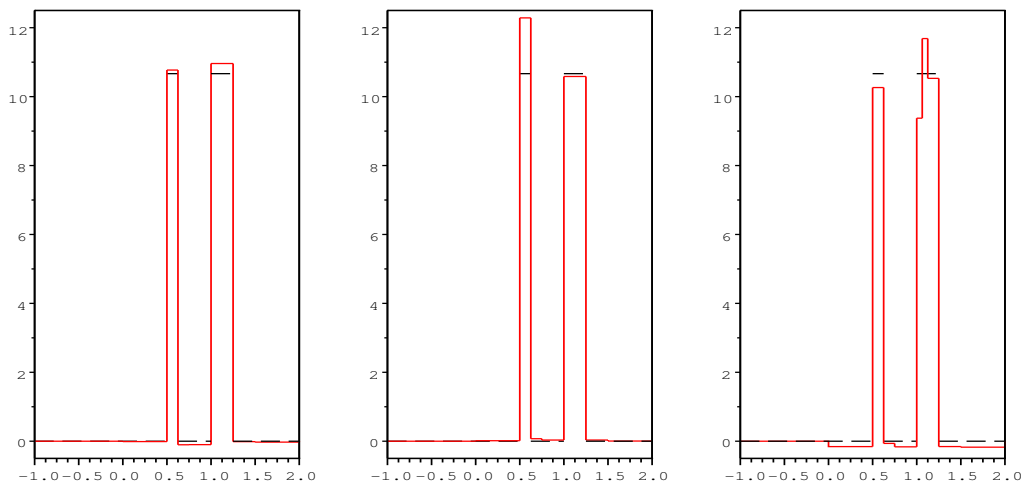


Figure 3.4: Reconstructions of 'Signal2' in the Haar basis (true: dotted line, estimate: solid line): left:  $(T, \mu) = (10000, 0.1)$ ; middle:  $(T, \mu) = (2000, 0.1)$ ; right:  $(T, \mu) = (2000, 0.5)$ .

### 3.3.4.2 About the support of $h$

We are interested in the robustness of our procedure with respect to the support issue from a numerical point of view. What happens if we are wrong about the support of the function that we want to estimate? For instance, we consider the testing function denoted 'Signal3' whose definition is given in the following table:

<p>'Signal3'</p> $\nu \times \frac{1}{4} (\mathbf{1}_{[-0.75; -0.5]} + \mathbf{1}_{[4.25; 8]})$
---

with  $\nu$ , the children's intensity, set to 4.

Figure 3.7 displays reconstructions of 'Signal3' in the Haar basis with different supports of  $h$ :  $[-A; A]$ , with  $A \in \{1, 5, 10\}$ . This figure shows that when we take a not large enough support ( $A = 1$  or  $5$ ), we do not make large errors of approximation on  $[-A; A]$ . So, the procedure seems to take into account what happens beyond the chosen support. And for  $A = 10$ , we have a good complete reconstruction of 'Signal3'.

Finally, even if the support of the reproduction function is unknown, our method estimates correctly the signal on the chosen support, which explains the robustness of our procedure with respect to the support issue.

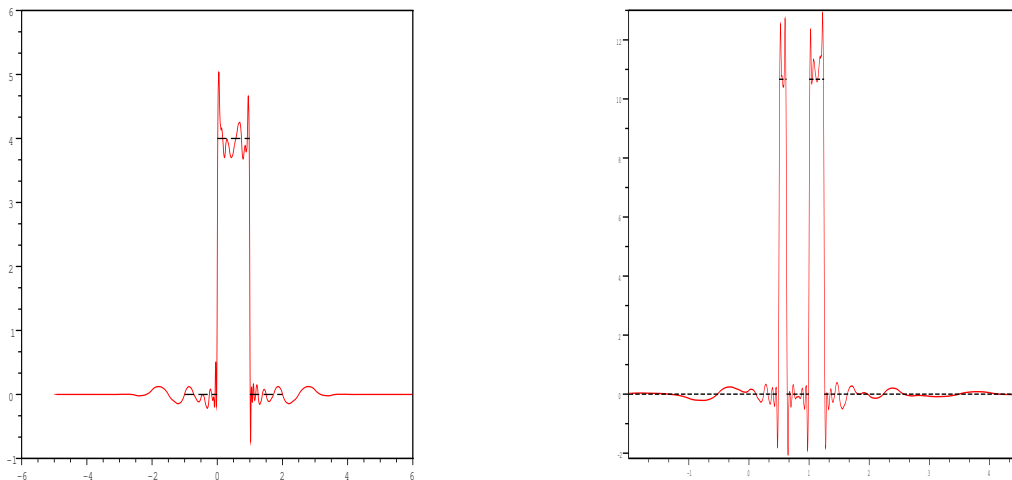


Figure 3.5: Reconstructions of 'Signal1' (left) and 'Signal2' (right) in the Spline basis (true: dotted line, estimate: solid line) with  $(T, \mu) = (10000, 0.1)$ .

### 3.3.5 Some extensions of our model

#### 3.3.5.1 The case of spontaneous apparition

Here, we investigate the case of spontaneous apparition. Even if our model does not take into account the spontaneous apparition (i.e. children can not be orphans), we are interested by the performance of our procedure if there is a presence of orphans. On the one hand, let us give two processes: a process of intensity 'Signal1' with  $\nu = 3$ ,  $T = 10000$  and  $\mu = 0.1$ , to which is added a homogeneous Poisson process on  $[0; T + 1]$  with intensity  $\mu(4 - \nu) = 0.1$  (the orphans are viewed as a Poissonian noise). Thus, we have in average 1000 parents, 3000 children having a parent and 1000 children being orphans. On the other hand, let us give two other processes: a process of intensity 'Signal1' with  $\nu = 1$  this time,  $T = 10000$  and  $\mu = 0.1$ , to which is added a homogeneous Poisson process on  $[0; T + 1]$  with intensity  $\mu(4 - \nu) = 0.3$ . Thus, we have in average 1000 parents, 1000 children having a parent and 3000 children being orphans.

Reconstructions of 'Signal1' with  $\nu = 3$  and  $\nu = 1$  in the Haar basis are given in Figure 3.8.

When there is a small proportion of children being orphans, the reconstruction is still acceptable; the procedure can manage few orphans. But, when there are too many orphans, our procedure makes approximation errors, which are due to the fact that our model consists in associating any child with a parent.

We mention that the case of spontaneous apparition is only numerical. For a more precise study of this phenomenon, we should extend our model by adding a positive constant to the intensity function  $t \mapsto \sum_{i=1}^n h(t - U_i)$ , that would represent the orphans. This is outside the scope of this thesis.

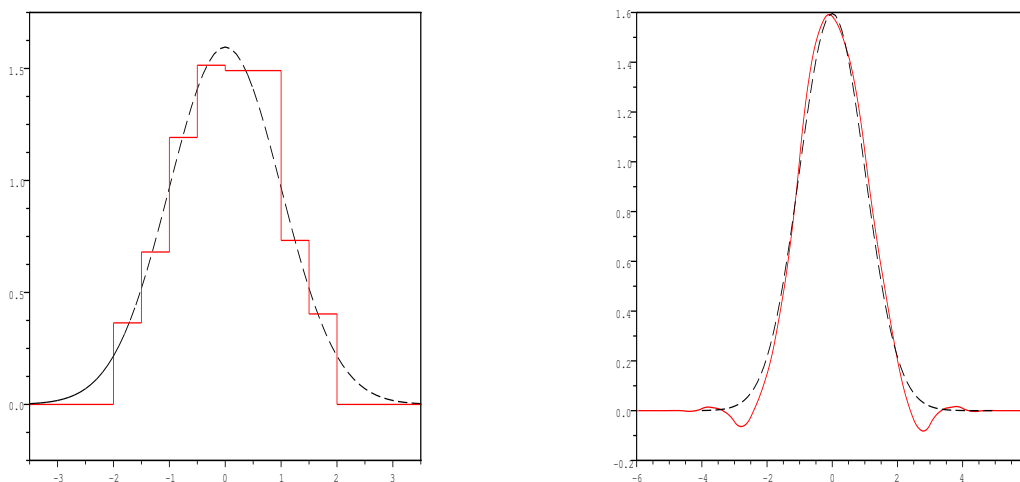


Figure 3.6: Reconstructions of 'SignalN' (true: dotted line, estimate: solid line): left: in the Haar basis; right: in the Spline basis.

### 3.3.5.2 The case of inhibition

This case explained in Introduction is not studied in this thesis from a theoretical point of view. Figure 3.9 shows a reconstruction for the testing function denoted 'Signal4' whose definition is given in the following table:

<p>'Signal4'</p> $(-\mathbf{1}_{[0;0.5]} + 4 \times \mathbf{1}_{[0.5;2]})$
--

Because the intensity of the process is  $(\sum_{i=1}^n h(t - U_i))_+$  with  $h$  equals to 'Signal4', this figure shows a good reconstruction of the positive part of  $h$  and a well detection of the negative part of  $h$  even if the level is underestimated (as for the study of the possibly negative function  $f_2$  in [93], with Figure 8 of this article).

## 3.4 Application to genomic data

As application, we are interested in the *Escherichia coli* genome. *E. coli* is an intestinal bacterium in mammals and very common in humans which is widely studied and used in genetics. More precisely, we are interested in the study of the dependence between promoter sites and genes along the complete genome of the bacterium. In particular, promoters are usually structured motifs located before the genes and not too far from them. Here, we have considered the major promoter of the bacterium *E. coli* and more precisely the word **tataat**. Most of the genes of *E. coli* should be preceded by this word at a very short distance apart. In order to validate our

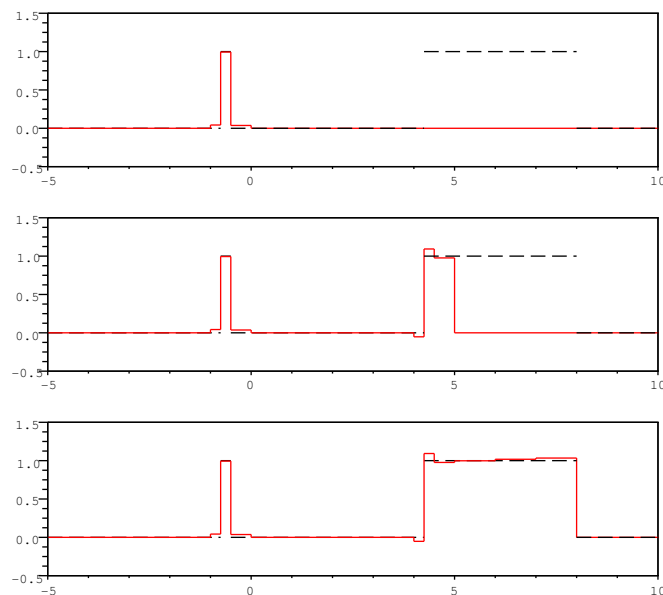


Figure 3.7: Reconstructions of 'Signal3' in the Haar basis (true: dotted line, estimate: solid line) with different supports: top:  $A = 1$ ; middle:  $A = 5$ ; bottom:  $A = 10$ .

thresholding estimation procedure (proposed at Section 3.3), we hope to detect short favored distances between genes and previous occurrences of **tataat**.

For this, as in [54] we have analyzed the sequence composed of both strands of *E. coli* genome (4639221 bases); each strand being separated by 10000 artificial bases to avoid artificial dependencies between occurrences on one strand and occurrences on the other strand; we took 10000 bases for the maximal memory. It then represents a sequence of length 9288442; there are 4290 genes (we took the positions of the first base of coding sequences) and 1036 occurrences of **tataat**. For convenience, we set  $T = 9289$  and so  $A = 10$  (we work on a scale of 1 : 1000). We recall that we have fixed  $j_0 = 5$  and taken  $(\gamma, \delta) = (0.18, 2.4)$ .

First, we investigate the way the DNA motif **tataat** influences genes and so, in our model, the parents are the occurrences of **tataat** and children are the occurrences of genes. To give general insight on  $h$ , Figure 3.10 gives the estimator  $\tilde{h}$  defined in Subsection 3.3.1 without thresholding, i.e. we have kept all the estimated coefficients. We observe a peak around 0 which corresponds to what we thought about the fact that most of the genes of *E. coli* should be preceded by the word **tataat** at a very short distance apart. We also observe other peaks, for instance around 1200 bases. The biological significance of these peaks remains an open question.

We apply the complete procedure proposed in Subsection 3.3.1 (with thresholding) and we obtain Figure 3.11. The shape of this estimator explains how occurrences of genes are influenced by occurrences of **tataat**. We can draw following conclusions, that coincide with the ones we could expect:

- The estimator  $\tilde{h}(t) = 0$  if  $t \leq 0$  and  $t \geq 500$ . It means that for such  $t$ 's, gene occurrences seem to be uncorrelated of **tataat** occurrences.
- Conversely, if  $t \in [0; 500]$ ,  $\tilde{h}(t) > 0$ , meaning that short distances are favored; smaller the distance, higher is the influence.

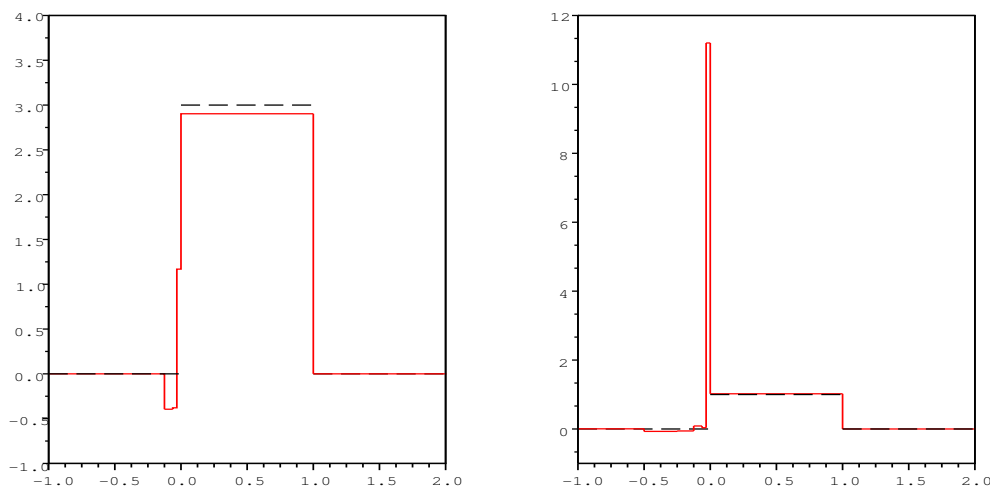


Figure 3.8: Reconstructions of 'Signal1' in the Haar basis (true: dotted line, estimate: solid line) with different values of  $\nu$ : left:  $\nu = 3$ ; right:  $\nu = 1$ .

Then, we investigate the way genes influences the DNA motif **tataat** and so, in our model, the parents are the occurrences of genes and children are the occurrences of **tataat**. Figure 3.12 gives the estimator  $\tilde{h}$  defined in Subsection 3.3.1 (with  $(\gamma, \delta) = (0.72, 2.4)$ ). The shape of this estimator explains how occurrences of **tataat** are influenced by occurrences of genes. We can draw following conclusions, that is completely coherent with biological observations:

- When  $t \leq -500$  and  $t \geq 1000$ ,  $\tilde{h}(t) = 0$ . It means that for such  $t$ 's, **tataat** occurrences seem to be uncorrelated of gene occurrences.
- When  $t \in [-500; 0]$ ,  $\tilde{h}(t) > 0$ , meaning that there is a preference having a word **tataat** just before the occurrence of a gene. It corresponds to the same conclusions drawn from Figure 3.11 (second point). The motif **tataat** is part of the most common promoter sites of *E. coli* meaning that it should occur in front of the majority of the genes.
- When  $t \in [0; 1000]$ ,  $\tilde{h}(t) < 0$ ; occurrences of **tataat** are avoided for such distances  $t$ . Genes on the same strand do not usually overlap and they are about 1000 bases long in average: this fact can explain this conclusion.

Finally, Figure 3.13 presents the results of the FADO procedure [54] and Figure 3.14 presents the results of the Islands procedure of [93]. For the FADO procedure, we have forced the estimators to be piecewise constant to make the comparison easier. Our results agree with the ones obtained by FADO and Islands. But our method has advantage to point out that nothing significant happens after a certain distance (contrary to the FADO procedure), has advantage to treat interaction with another type of events (contrary to the Islands procedure) and has advantage to deal with the dependence on the past occurrences but also on the future occurrences (the function  $h$  is supported in  $\mathbb{R}_+$  for the two other procedures). For algorithmic reason, a practical limitation of our method is that we only consider piecewise constant estimators (as for the Islands procedure), but it is enough to get a general trend on favored or avoided distances within a point process.

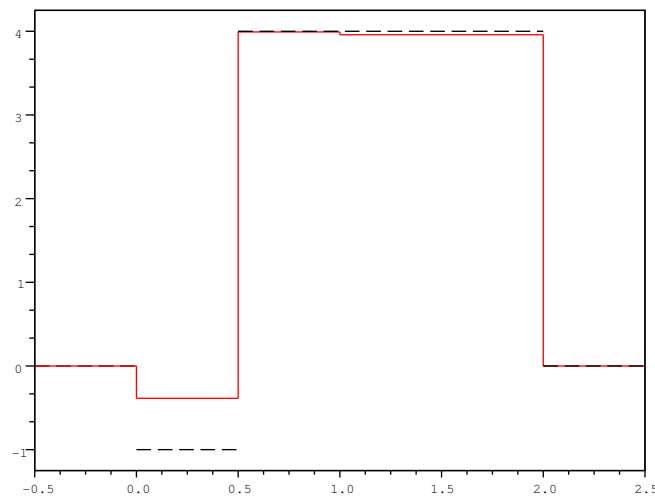


Figure 3.9: Reconstructions of 'Signal4' in the Haar basis (true: dotted line, estimate: solid line).

### 3.5 Conclusion

In this chapter, we have investigated the dependencies between two given motifs. A random thresholding procedure has been proposed in Subsection 3.2.1. The general results of Section 3.2.2 have revealed the optimality of the procedure in the oracle and minimax setting. Our theoretical results have been strengthened by simulations illustrating the robustness of our procedure, despite a calibration of parameters from a practical point of view that differs from the theoretical choice. Section 3.4 has validated the procedure with a good detection of favored or avoided distances between occurrences of `tataat` and genes along the *E. coli* genome.

Further extensions of our model could be investigated. First, we could consider a more sophisticated model that takes into account the phenomena of spontaneous apparition and self-excitation (as for the complete Hawkes model). But this model raises serious difficulties from the theoretical point of view. This is an exciting challenge to overcome them. Secondly, we could extend our cascade algorithm to general wavelet bases and not only to Haar bases or the Spline basis. Finally, it is also relevant to study similar processes in the spatial framework and to connect them, for instance, to the Neymann-Scott process (see Section 6.3 of [29]), which is a stimulating topic we wish to consider.



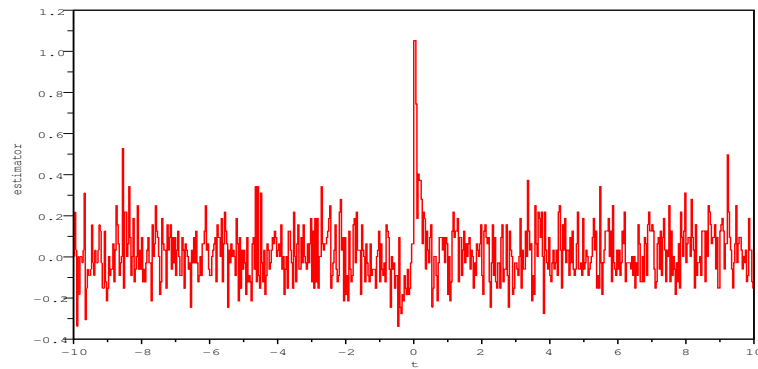


Figure 3.10: Estimator, no thresholding, for *E. coli* data at the scale 1 : 1000 (i.e. 1 corresponds to 1000 bases), with parents=*tataat* and children=*genes*.

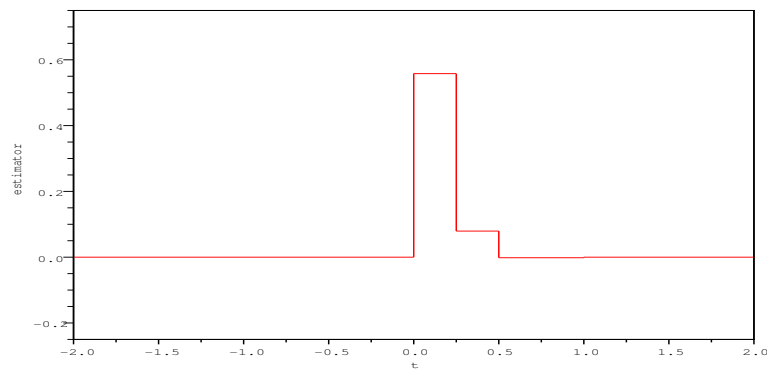


Figure 3.11: Estimator  $\tilde{h}$  defined in Subsection 3.3.1 for *E. coli* data at the scale 1 : 1000, with parents=*tataat* and children=*genes*.

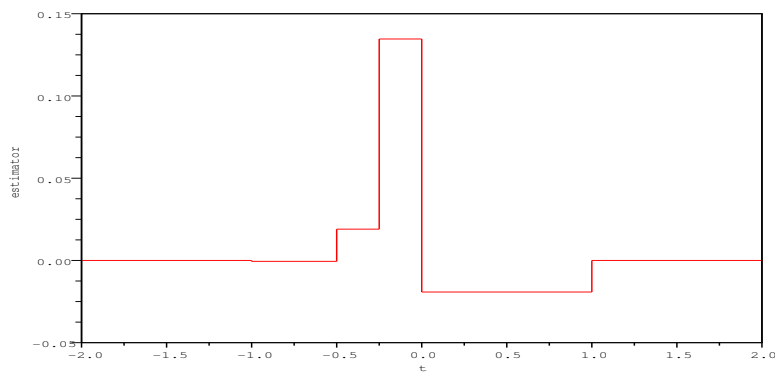


Figure 3.12: Estimator  $\tilde{h}$  defined in Subsection 3.3.1 for *E. coli* data at the scale 1 : 1000, with parents=*genes* and children=*tataat*.

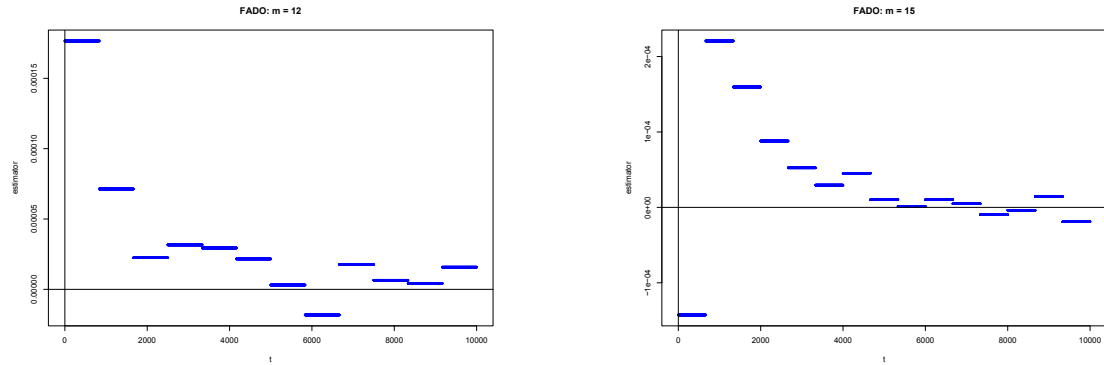


Figure 3.13: *FADO* estimators for both *E. coli* datasets: left: *tataat*; right: *genes*.

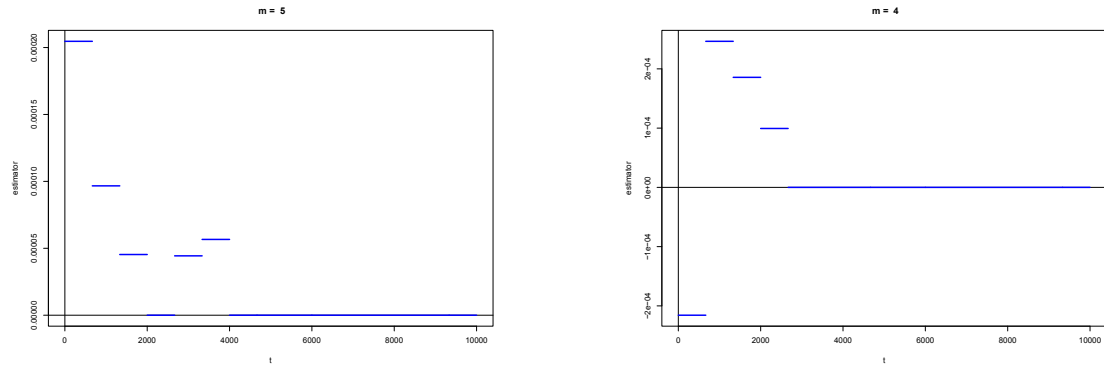


Figure 3.14: *Islands* estimators for both *E. coli* datasets: left: *tataat*; right: *genes*.

## 3.6 Proofs

### 3.6.1 Proof of Proposition 3.1

#### 3.6.1.1 Computation of $\mathbb{E}(\hat{\beta}_\lambda)$

Let  $\lambda \in \Lambda$  be fixed.  $G(\varphi_\lambda)$ , defined by (3.5), is a measurable function of the observations and by considering the aggregated process (3.1), we can write

$$\begin{aligned}
 G(\varphi_\lambda) &= \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] dN_t \\
 &= \int_{\mathbb{R}} \sum_{i=1}^n \varphi_\lambda(t - U_i) dN_t - (n-1) \int_{\mathbb{R}} \mathbb{E}_\pi(\varphi_\lambda(t - U)) dN_t \\
 &= \sum_{1 \leq i, j \leq n} \int_{\mathbb{R}} \varphi_\lambda(t - U_i) dN_t^j - \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} \mathbb{E}_\pi(\varphi_\lambda(t - U)) dN_t^j \\
 &= \sum_{i=1}^n \left[ \int_{\mathbb{R}} \varphi_\lambda(t - U_i) dN_t^i + \sum_{j \neq i} \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] dN_t^j \right].
 \end{aligned}$$

We have

$$\begin{aligned} & \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n) \\ &= \sum_{i=1}^n \left[ \int_{\mathbb{R}} \varphi_\lambda(t - U_i) h(t - U_i) dt + \sum_{j \neq i} \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] h(t - U_j) dt \right]. \end{aligned}$$

Write  $x = t - U_i$  in the first integral. Therefore,

$$\mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx + W(\varphi_\lambda), \quad (3.17)$$

where

$$W(\varphi_\lambda) = \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] h(t - U_j) dt. \quad (3.18)$$

Moreover,

$$\begin{aligned} \mathbb{E}(W(\varphi_\lambda)) &= \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} \mathbb{E} \left( [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] h(t - U_j) \right) dt \\ &= \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} \mathbb{E} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}(h(t - U_j)) dt \\ &= 0. \end{aligned}$$

Finally,

$$\mathbb{E}(G(\varphi_\lambda)) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx,$$

i.e.  $\hat{\beta}_\lambda$  is an unbiased estimator for  $\beta_\lambda$ :

$$\mathbb{E}(\hat{\beta}_\lambda) = \mathbb{E} \left( \frac{G(\varphi_\lambda)}{n} \right) = \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx = \beta_\lambda,$$

that proves the first part of Proposition 3.1.

### 3.6.1.2 Control of $\text{Var}(\hat{\beta}_\lambda)$

Remember that  $\hat{\beta}_\lambda = \frac{G(\varphi_\lambda)}{n}$ . We have the following decomposition of  $\text{Var}(G(\varphi_\lambda))$  into two terms:

$$\text{Var}(G(\varphi_\lambda)) = \mathbb{E}(\text{Var}(G(\varphi_\lambda)|U_1, \dots, U_n)) + \text{Var}(\mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)). \quad (3.19)$$

We denote  $V(\varphi_\lambda) = \text{Var}(G(\varphi_\lambda)|U_1, \dots, U_n)$  and by using (3.17) and (3.18),

$$\text{Var}(\mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)) = \mathbb{E}(W(\varphi_\lambda)^2), \quad (3.20)$$

We start by dealing with the first term of (3.19) by using techniques for Poisson processes. We have

$$\begin{aligned} & V(\varphi_\lambda) \\ &= \int_{\mathbb{R}} \left( \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right)^2 \sum_{j=1}^n h(t - U_j) dt \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbb{R}} \sum_{j=1}^n \left( \varphi_{\lambda}(t - U_j) + \sum_{i \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \right)^2 h(t - U_j) dt \\
&= \sum_{j=1}^n \int_{\mathbb{R}} \varphi_{\lambda}^2(t - U_j) h(t - U_j) dt \\
&\quad + 2 \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \varphi_{\lambda}(t - U_j) h(t - U_j) dt \\
&\quad + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \sum_{k \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] h(t - U_j) dt.
\end{aligned}$$

In the first integral, write  $x = t - U_j$ . So,

$$\begin{aligned}
V(\varphi_{\lambda}) &= n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + 2 \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \varphi_{\lambda}(t - U_j) h(t - U_j) dt \\
&\quad + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \sum_{k \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] h(t - U_j) dt.
\end{aligned} \tag{3.21}$$

Each term can be computed by taking the expectation conditionally to  $U_j$  (in each sum) and we obtain

$$\begin{aligned}
\mathbb{E}(V(\varphi_{\lambda})) &= n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \mathbb{E} \left( [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 \right) \mathbb{E}(h(t - U_j)) dt \\
&= n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + n(n-1) \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t - U)) \mathbb{E}_{\pi}(h(t - U)) dt \\
&\leq n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + \frac{n(n-1)}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \int_{\mathbb{R}} h(x) dx,
\end{aligned} \tag{3.22}$$

by using the two following lemmas.

**Lemma 3.1.** For any function  $f$  in  $\mathbb{L}_2(\mathbb{R})$  and for all  $t \in \mathbb{R}$ ,

$$\text{Var}_{\pi}(f(t - U)) \leq \frac{1}{T} \int_{\mathbb{R}} f^2(x) dx,$$

where  $\text{Var}_{\pi}(f(t - U))$  denotes the variance of  $f(t - U)$  where  $U \sim \pi$ .

*Proof.* Let  $f \in \mathbb{L}_2(\mathbb{R})$  and  $t \in \mathbb{R}$ . Then,

$$\text{Var}_{\pi}(f(t - U)) \leq \mathbb{E}_{\pi}(f^2(t - U)) = \frac{1}{T} \int_0^T f^2(t - u) du \leq \frac{1}{T} \int_{\mathbb{R}} f^2(x) dx.$$

□

**Lemma 3.2.** For any function  $f$  in  $\mathbb{L}_1(\mathbb{R})$  and for all  $t \in \mathbb{R}$ ,

$$\int_{\mathbb{R}} \mathbb{E}_{\pi}(f(t - U)) dt = \int_{\mathbb{R}} f(x) dx.$$

*Proof.* Let  $f \in \mathbb{L}_1(\mathbb{R})$  and  $t \in \mathbb{R}$ . Then,

$$\int_{\mathbb{R}} \mathbb{E}_{\pi}(f(t - U)) dt = \mathbb{E}_{\pi} \left( \int_{\mathbb{R}} f(t - U) dt \right) = \mathbb{E} \left( \int_{\mathbb{R}} f(x) dx \right) = \int_{\mathbb{R}} f(x) dx.$$

□

Now, we deal with the second term of (3.19) expressed in (3.20) by using the  $U$ -statistics techniques. Note that  $W(\varphi_{\lambda})$  is a  $U$ -statistics of order 2 but it is not degenerate. So we write

$$W(\varphi_{\lambda}) = W_1(\varphi_{\lambda}) + W_2(\varphi_{\lambda}),$$

with

$$\begin{aligned} W_1(\varphi_{\lambda}) &= \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \\ &= (n - 1) \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \end{aligned}$$

and

$$W_2(\varphi_{\lambda}) = \sum_{1 \leq i \neq j \leq n} g(U_i, U_j),$$

where

$$g(U_i, U_j) = \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt.$$

$W_2(\varphi_{\lambda})$  is a degenerate  $U$ -statistics. It is easy to verify that

$$\mathbb{E}(W(\varphi_{\lambda})^2) = \mathbb{E}(W_1(\varphi_{\lambda})^2) + \mathbb{E}(W_2(\varphi_{\lambda})^2).$$

First we compute  $\mathbb{E}(W_1(\varphi_{\lambda})^2)$ .

$$\begin{aligned} \mathbb{E}(W_1(\varphi_{\lambda})^2) &= \text{Var}(W_1(\varphi_{\lambda})) \\ &= n(n - 1)^2 \text{Var} \left( \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_1) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \right) \\ &\leq n(n - 1)^2 \mathbb{E} \left[ \left( \int_{\mathbb{R}} |\varphi_{\lambda}(t - U_1)| \mathbb{E}_{\pi}(h(t - U)) dt \right)^2 \right] \\ &\leq \frac{n(n - 1)^2}{T^2} \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2, \end{aligned} \tag{3.23}$$

by applying the following lemma.

**Lemma 3.3.** For any nonnegative function  $f$  in  $\mathbb{L}_1(\mathbb{R})$  and for all  $t \in \mathbb{R}$ ,

$$\mathbb{E}_{\pi}(f(t - U)) \leq \frac{1}{T} \int_{\mathbb{R}} f(x) dx.$$

*Proof.* Let  $f \in \mathbb{L}_1(\mathbb{R})$  such that  $f \geq 0$  and  $t \in \mathbb{R}$ . Then,

$$\mathbb{E}_{\pi}(f(t - U)) = \frac{1}{T} \int_0^T f(t - u) du \leq \frac{1}{T} \int_{\mathbb{R}} f(x) dx.$$

□

It remains to compute  $\mathbb{E}(W_2(\varphi_\lambda)^2)$ . It is easy to see that

$$\begin{aligned} & \mathbb{E}(W_2(\varphi_\lambda)^2) \\ &= \sum_{1 \leq i \neq j \leq n} \mathbb{E}[g(U_i, U_j)(g(U_i, U_j) + g(U_j, U_i))] \leq 2 \sum_{1 \leq i \neq j \leq n} \mathbb{E}[g(U_i, U_j)^2] \\ &\leq 2n(n-1) \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [h(t - U_2) - \mathbb{E}_\pi(h(t - U))] dt \right)^2 \right]. \end{aligned}$$

We denote  $\mathbb{E}_{(U,V) \sim \pi \otimes \pi}(g(U, V))$  the expectation of  $g(U, V)$  where  $U \sim \pi$  and  $V \sim \pi$  are independent and  $f^X(t) = f(t - X)$ . Hence,

$$\begin{aligned} & \mathbb{E}(W_2(\varphi_\lambda)^2) \\ &\leq 2n(n-1) \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda^U(t) - \mathbb{E}_\pi(\varphi_\lambda^U(t))] [h^V(t) - \mathbb{E}_\pi(h^V(t))] dt \right)^2 \right] \\ &\leq 2n(n-1) \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left[ \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt - \mathbb{E}_{V \sim \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right. \right. \\ &\quad \left. \left. - \mathbb{E}_{U \sim \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) + \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right)^2 \right] \\ &\leq 2n(n-1) \left\{ \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left[ \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right)^2 \right] - \mathbb{E}_{U \sim \pi} \left[ \left( \mathbb{E}_{V \sim \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right)^2 \right] \right. \\ &\quad \left. - \mathbb{E}_{V \sim \pi} \left[ \left( \mathbb{E}_{U \sim \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right)^2 \right] + \left( \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right)^2 \right\} \\ &\leq 2n(n-1) \left\{ \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left[ \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right)^2 \right] + \left( \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right)^2 \right\}. \end{aligned}$$

But,

$$\begin{aligned} \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left[ \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right)^2 \right] &\leq \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} (\varphi_\lambda^U)^2(t) h^V(t) dt \int_{\mathbb{R}} h^V(t) dt \right) \\ &= \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} (\varphi_\lambda^U)^2(t) h^V(t) dt \right) \int_{\mathbb{R}} h(x) dx \\ &\leq \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2 \end{aligned}$$

and

$$\left| \mathbb{E}_{(U,V) \sim \pi \otimes \pi} \left( \int_{\mathbb{R}} \varphi_\lambda^U(t) h^V(t) dt \right) \right| = \left| \int_{\mathbb{R}} \mathbb{E}_\pi(\varphi_\lambda^U(t)) \mathbb{E}_\pi(h^V(t)) dt \right| \leq \frac{1}{T} \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \int_{\mathbb{R}} h(x) dx,$$

by using Lemma 3.2 and Lemma 3.3. So,

$$\mathbb{E}(W_2(\varphi_\lambda)^2) \leq 2n(n-1) \left\{ \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2 + \frac{1}{T^2} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2 \right\}. \quad (3.24)$$

Finally, by combining inequalities (3.19), (3.22), (3.23) and (3.24), we obtain the following control of the variance of the estimator  $\hat{\beta}_\lambda$ :

$$\begin{aligned} \text{Var}(\hat{\beta}_\lambda) &= \text{Var}\left(\frac{G(\varphi_\lambda)}{n}\right) \\ &\leq \frac{1}{n} \int_{\mathbb{R}} \varphi_\lambda^2(x) h(x) dx + \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \int_{\mathbb{R}} h(x) dx \\ &\quad + \frac{n}{T^2} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2 + \frac{2}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2 \\ &\leq \frac{1}{n} \int_{\mathbb{R}} \varphi_\lambda^2(x) h(x) dx + \frac{1}{T} \|\varphi_\lambda\|_2^2 \|h\|_1 + \frac{n}{T^2} \|\varphi_\lambda\|_1^2 \|h\|_1^2 + \frac{2}{T} \|\varphi_\lambda\|_2^2 \|h\|_1^2. \end{aligned}$$

By using the properties of the biorthogonal wavelet bases considered in this chapter, for any  $\lambda = (j, k)$  in  $\Lambda$ , we have:  $\|\varphi_\lambda\|_1 \leq 2^{-j/2} \max(\sqrt{2}/2, \|\psi\|_1)$  and  $\|\varphi_\lambda\|_2 \leq \max(1, \|\psi\|_2)$ , which allows us to get the upper bound of Proposition 3.1.

### 3.6.2 Proof of Theorem 3.1

Theorem 3.1 is a consequence of the following result. The detailed proof of Theorem 3.2 is available in the next section.

**Theorem 3.2.** *Let  $n \geq 1$ ,  $j_0 \in \mathbb{N}^*$  and  $\gamma > 0$ . Then the estimator  $\tilde{\beta}$  defined by (3.12) in Subsection 3.2.1 satisfies*

$$\mathbb{E} \left( \|\tilde{\beta} - \beta\|_{\ell_2}^2 \right) \leq C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + F(j_0, n, T) |m| \right\} + C_2 R (e^{-\kappa_1 j_0 \gamma / 2} + e^{-\kappa_2 n \|h\|_1 / 2}) 2^{j_0},$$

where  $C_1$  is a positive constant depending on  $\gamma$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$ ,  $C_2$  is a positive constant depending on  $A$  and the compact support of  $\psi$ ,

$$\begin{aligned} F(j_0, n, T) &= \frac{j_0^3}{n} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0 n}{T^2} + \frac{j_0^2}{n^2 T^2}, \\ R &= C_R \left\{ \frac{1}{n} + \frac{2^{j_0/2}}{n^{3/2}} + \frac{2^{j_0/2}}{n T^{1/2}} + \frac{n}{T^2} \right\}, \end{aligned}$$

with  $C_R$  a positive constant depending on  $\|h\|_1$ ,  $\|h\|_\infty$ , the compact support of  $\psi$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  and  $\kappa_1$  and  $\kappa_2$  are absolute constants in  $]0; 1[$ .

To obtain Theorem 3.1, we consider  $n \geq 2$ , we take  $j_0$  the positive integer such that  $2^{j_0} \leq n < 2^{j_0+1}$  and  $\gamma > 2 \log 2$  in Theorem 3.2. Therefore, we note that

$$\begin{aligned} F(j_0, n, T) &= \frac{j_0^3}{n} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0 n}{T^2} + \frac{j_0^2}{n^2 T^2} \\ &\leq K \left\{ \frac{(\log n)^4}{n} + \frac{(\log n)^2 n}{T^2} \right\}, \end{aligned}$$

by comparing all the terms of the right-hand side between them (for this, we distinguish the cases  $n \leq T$  and  $n \geq T$ ), with  $K$  an absolute positive constant (that changes from line to line) and

$$R \leq K' C_R \left\{ \frac{1}{n} + \frac{n}{T^2} \right\},$$

with  $K'$  an absolute positive constant. Moreover,  $e^{-\kappa_1 j_0 \gamma / 2} 2^{j_0}$  is bounded thanks to the choice of  $\gamma$ .

Finally, since

$$\|\tilde{h} - h\|_2^2 \leq K_0 \|\tilde{\beta} - \beta\|_{\ell_2}^2,$$

with  $K_0$  a positive constant depending only on the functions that generate the biorthogonal wavelet basis, we establish Theorem 3.1.

### 3.A Supplement: Proof of Theorem 3.2

In this supplement, we will consider:  $n \geq 1$ ,  $T \geq 1$  and  $j_0 = O(n)$  and we will use following notations:  $M_{h,1} = \max(\|h\|_1, 1)$ ,  $M_{h,\infty} = \max(\|h\|_\infty, 1)$ ,  $M_{\psi,1} = \max(\|\psi\|_1, \sqrt{2}/2)$ ,  $M_{\psi,2} = \max(\|\psi\|_2, 1)$  and  $M_{\psi,\infty} = \max(\|\psi\|_\infty, \sqrt{2})$  (so that, for any  $\lambda = (j, k) \in \Lambda$ , we have:  $\|\varphi_\lambda\|_1 \leq 2^{-j/2} M_{\psi,1}$ ,  $\|\varphi_\lambda\|_2 \leq M_{\psi,2}$  and  $\|\varphi_\lambda\|_\infty \leq 2^{j/2} M_{\psi,\infty}$ ). We recall that  $A$  and  $M$  are positive real numbers such that  $h$  and  $\psi$  are compactly supported in  $[-A; A]$  and in  $[-M; M]$  respectively.

We first give a general result stated and proved in [91].

**Theorem 3.3** (Theorem 2.2 of [91]). *To estimate a countable family  $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ , such that  $\|\beta\|_{\ell_2} < \infty$ , we assume that a family of coefficient estimators  $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$ , where  $\Gamma$  is a known deterministic subset of  $\Lambda$ , and a family of possibly random thresholds  $(\eta_\lambda)_{\lambda \in \Gamma}$  are available and we consider the thresholding rule*

$$\tilde{\beta} = \left( \hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma} \right)_{\lambda \in \Lambda}.$$

Let  $\varepsilon > 0$  be fixed. Assume that there exist a deterministic family  $(H_\lambda)_{\lambda \in \Gamma}$  and three constants  $\kappa \in [0; 1[$ ,  $\omega \in [0; 1]$  and  $\zeta > 0$  (that may depend on  $\varepsilon$  but not on  $\lambda$ ) with the following properties:

(A1) For all  $\lambda$  in  $\Gamma$ ,

$$\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \leq \omega.$$

(A2) There exist  $1 < p, q < \infty$  with  $\frac{1}{p} + \frac{1}{q} = 1$  and a constant  $R > 0$  such that for all  $\lambda$  in  $\Gamma$ ,

$$\left[ \mathbb{E} \left( |\hat{\beta}_\lambda - \beta_\lambda|^{2p} \right) \right]^{\frac{1}{p}} \leq R \max \left\{ H_\lambda, H_\lambda^{\frac{1}{p}} \varepsilon^{\frac{1}{q}} \right\}.$$

(A3) There exists a constant  $\theta$  such that for all  $\lambda$  in  $\Gamma$  such that  $H_\lambda < \theta \varepsilon$ ,

$$\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda \right) \leq H_\lambda \zeta.$$

Then the estimator  $\tilde{\beta}$  satisfies

$$\begin{aligned} & \frac{1 - \kappa^2}{1 + \kappa^2} \mathbb{E} \left( \|\tilde{\beta} - \beta\|_{\ell_2}^2 \right) \\ & \leq \mathbb{E} \left( \inf_{m \subset \Gamma} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda \in m} \eta_\lambda^2 \right\} \right) + LD \sum_{\lambda \in \Gamma} H_\lambda, \end{aligned}$$

with  $LD = \frac{R}{\kappa^2} \left( (1 + \theta^{-1/q}) \omega^{1/q} + (1 + \theta^{1/q}) \varepsilon^{1/q} \zeta^{1/q} \right)$ .

In order to prove Theorem 3.2, we apply this previous theorem and for this purpose we have to verify Assumptions: (A1), (A2) and (A3). In the sequel, the values of the constants  $K, K', K_0, K_1, K_2, K_3, \dots$  may change from line to line. For the sake of clarity, the proof is fully detailed in this supplement.



### 3.A.1 A technical lemma

Before proving Theorem 3.2, we establish a lemma which we will use throughout the proof. This result is a Rosenthal type inequality for any Poisson process, that extends Lemma 6.2 of [91].

**Lemma 3.4.** *Let  $p \geq 1$ . Consider a Poisson process  $N$  on  $(\mathbb{X}, \mathcal{X})$  a measurable space, with a finite mean measure  $\nu : \mathcal{X} \mapsto \mathbb{R}_+$  and a function  $\varphi : \mathbb{X} \mapsto \mathbb{R}$  which belongs to  $\mathbb{L}_{2p}(\nu)$ . We denote  $\hat{\beta} = \int_{\mathbb{X}} \varphi(x) dN_x$  a natural estimator of  $\beta = \int_{\mathbb{X}} \varphi(x) d\nu(x)$  that satisfies  $\mathbb{E}(\hat{\beta}) = \beta$ . Then, there exists a positive constant  $C(p)$  only depending on  $p$  such that*

$$\mathbb{E}(|\hat{\beta} - \beta|^{2p}) \leq C(p) \left( \int_{\mathbb{X}} |\varphi(x)|^{2p} d\nu(x) + (\text{Var}(\hat{\beta}))^p \right),$$

where  $\text{Var}(\hat{\beta}) = \int_{\mathbb{X}} \varphi^2(x) d\nu(x)$ .

*Proof.* Let  $p \geq 1$ . Suppose  $\|\varphi\|_{\infty} < +\infty$  first. As a Poisson process is infinitely divisible, we can write: for any positive integer  $k$ ,

$$dN = \sum_{i=1}^k dN^i,$$

where the  $N^i$ 's are mutually independent Poisson processes on  $\mathbb{X}$  with mean measure  $\nu/k$ . Hence,

$$\hat{\beta} - \beta = \sum_{i=1}^k \int_{\mathbb{X}} \varphi(x) (dN_x^i - k^{-1} d\nu(x)) = \sum_{i=1}^k Y_i,$$

where for any  $i$ ,

$$Y_i = \int_{\mathbb{X}} \varphi(x) (dN_x^i - k^{-1} d\nu(x)).$$

So the  $Y_i$ 's are i.i.d. centered variables, each of them has moments of order  $2p$  and  $2$ . We apply the classical Rosenthal's inequality (for instance, see Proposition 10.2 of [57]): there exists a positive constant  $C(p)$  only depending on  $p$  such that

$$\mathbb{E} \left( \left| \sum_{i=1}^k Y_i \right|^{2p} \right) \leq C(p) \left( \sum_{i=1}^k \mathbb{E}(|Y_i|^{2p}) + \left( \sum_{i=1}^k \mathbb{E}(Y_i^2) \right)^p \right).$$

Now, we give an upper bound of the limit of  $\mathbb{E} \left( \sum_{i=1}^k |Y_i|^{\ell} \right)$  for  $\ell \in \{2p, 2\}$  when  $k \rightarrow \infty$ . Let us introduce

$$\Omega_k = \{ \forall i \in \{1, \dots, k\}, N_{\mathbb{X}}^i \leq 1 \},$$

where  $N_{\mathbb{X}}^i$  is the number of points of  $N^i$  lying in  $\mathbb{X}$ . Then,

$$\begin{aligned} \mathbb{P}(\Omega_k^c) &= \mathbb{P}(\exists i \in \{1, \dots, k\}, N_{\mathbb{X}}^i \geq 2) \\ &\leq \sum_{i=1}^k \mathbb{P}(N_{\mathbb{X}}^i \geq 2) = k \sum_{j \geq 2} \frac{(\nu(\mathbb{X})/k)^j}{j!} e^{-\nu(\mathbb{X})/k} \\ &\leq k(\nu(\mathbb{X})/k)^2 = k^{-1} \nu(\mathbb{X})^2. \end{aligned}$$

On  $\Omega_k$ , if  $N_{\mathbb{X}}^i = 0$  (so  $\int_{\mathbb{X}} \varphi(x) dN_x^i = 0$ ),

$$|Y_i|^\ell = O_k(k^{-\ell})$$

and if  $N_{\mathbb{X}}^i = 1$  (so  $\int_{\mathbb{X}} \varphi(x) dN_x^i = \varphi(T)$ , where  $T$  is the point of the process  $N^i$ ),

$$|Y_i|^\ell = |\varphi(T)|^\ell + O_k(k^{-1}|\varphi(T)|^{\ell-1}).$$

Consequently,

$$\begin{aligned} & \mathbb{E} \left( \sum_{i=1}^k |Y_i|^\ell \right) \\ & \leq \mathbb{E} \left[ \mathbf{1}_{\Omega_k} \left( k O_k(k^{-\ell}) + \sum_{T \in N} [|\varphi(T)|^\ell + O_k(k^{-1}|\varphi(T)|^{\ell-1})] \right) \right] + \sqrt{\mathbb{P}(\Omega_k^c)} \sqrt{\mathbb{E} \left[ \left( \sum_{i=1}^k |Y_i|^\ell \right)^2 \right]}. \end{aligned} \quad (3.25)$$

But we have

$$\begin{aligned} \sum_{i=1}^k |Y_i|^\ell & \leq 2^{\ell-1} \sum_{i=1}^k \left[ \left| \int_{\mathbb{X}} \varphi(x) dN_x^i \right|^\ell + \left( k^{-1} \int_{\mathbb{X}} |\varphi(x)| d\nu(x) \right)^\ell \right] \\ & \leq 2^{\ell-1} \left( \sum_{i=1}^k \|\varphi\|_\infty^\ell (N_{\mathbb{X}}^i)^\ell + k \left( k^{-1} \int_{\mathbb{X}} |\varphi(x)| d\nu(x) \right)^\ell \right) \\ & \leq 2^{\ell-1} \left( \|\varphi\|_\infty^\ell N_{\mathbb{X}}^\ell + k \left( k^{-1} \int_{\mathbb{X}} |\varphi(x)| d\nu(x) \right)^\ell \right). \end{aligned}$$

Thus, when  $k \rightarrow \infty$ , the last term in (3.25) converges to 0 since a Poisson variable has moments of every order and

$$\limsup_{k \rightarrow \infty} \mathbb{E} \left( \sum_{i=1}^k |Y_i|^\ell \right) \leq \mathbb{E} \left( \int_{\mathbb{X}} |\varphi(x)|^\ell dN_x \right) = \int_{\mathbb{X}} |\varphi(x)|^\ell d\nu(x),$$

which concludes the proof in the bounded case.

But for any function  $\varphi$  such that  $\int_{\mathbb{X}} |\varphi(x)|^{2p} d\nu(x) < +\infty$ , the desired upper bound is finite and we get it by approximating  $\varphi$  by, for instance, piecewise constant functions.  $\square$

### 3.A.2 Proof of Assumption (A1) of Theorem 3.3

Let  $\lambda \in \Gamma$  be fixed. Remember that conditionally to the  $U_i$ 's, the expression given in (3.1) is a Poisson process. We apply Lemma 6.1 of [91]: for any  $\alpha > 0$ , with probability larger than  $1 - 2e^{-\alpha}$ , conditionally to the  $U_i$ 's, we have

$$\left| G(\varphi_\lambda) - n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx - W(\varphi_\lambda) \right| \leq \sqrt{2\alpha V(\varphi_\lambda)} + \frac{\alpha}{3} B(\varphi_\lambda),$$

where  $W(\varphi_\lambda)$  is defined by (6.5),  $V(\varphi_\lambda) = \text{Var}(G(\varphi_\lambda)|U_1, \dots, U_n)$  and

$$B(\varphi_\lambda) = \left\| \sum_{i=1}^n \left[ \varphi_\lambda(\cdot - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(\cdot - U)) \right] \right\|_\infty.$$

Unlike  $B(\varphi_\lambda)$ ,  $V(\varphi_\lambda)$  is non-observable (it depends on the unknown function  $h$ ). This is the reason why, by fixing  $\alpha > 0$ , we estimate  $V(\varphi_\lambda)$  by

$$\tilde{V}(\varphi_\lambda) = \hat{V}(\varphi_\lambda) + \sqrt{2\alpha\hat{V}(\varphi_\lambda)B^2(\varphi_\lambda) + 3\alpha B^2(\varphi_\lambda)}$$

where

$$\hat{V}(\varphi_\lambda) = \int_{\mathbb{R}} \left( \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right)^2 dN_t.$$

Moreover, by Lemma 6.1 of [91], we have also:  $\mathbb{P}(V(\varphi_\lambda) \geq \tilde{V}(\varphi_\lambda)) \leq e^{-\alpha}$ . So, with probability larger than  $1 - 3e^{-\alpha}$ ,

$$\left| G(\varphi_\lambda) - n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx \right| \leq \sqrt{2\alpha\tilde{V}(\varphi_\lambda)} + \frac{\alpha}{3} B(\varphi_\lambda) + |W(\varphi_\lambda)|. \quad (3.26)$$

We provide a control in probability of  $W_1(\varphi_\lambda)$ .

$$W_1(\varphi_\lambda) = (n-1) \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt.$$

This is a sum of i.i.d. random variables. We apply Bernstein's inequality (for instance, see Proposition 2.9 of [77]) to get that with probability larger than  $1 - 2e^{-\alpha}$ ,

$$|W_1(\varphi_\lambda)| \leq \sqrt{2\alpha v(\varphi_\lambda)} + \frac{\alpha}{3} b(\varphi_\lambda),$$

with

$$v(\varphi_\lambda) = \text{Var}(W_1(\varphi_\lambda)) \leq \frac{n(n-1)^2}{T^2} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2$$

(see inequality (6.6)) and

$$\begin{aligned} b(\varphi_\lambda) &= (n-1) \sup_{u \in [0; T]} \left| \int_{\mathbb{R}} [\varphi_\lambda(t - u) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right| \\ &\leq \frac{2(n-1)}{T} \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \int_{\mathbb{R}} h(x) dx, \end{aligned}$$

using Lemma 3.2 and Lemma 3.3. Then, with probability larger than  $1 - 2e^{-\alpha}$ ,

$$|W_1(\varphi_\lambda)| \leq \frac{\sqrt{2\alpha n}(n-1)}{T} \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \int_{\mathbb{R}} h(x) dx + \frac{2\alpha(n-1)}{3T} \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \int_{\mathbb{R}} h(x) dx. \quad (3.27)$$

Now it remains to control  $W_2(\varphi_\lambda)$ , with

$$W_2(\varphi_\lambda) = \sum_{1 \leq i \neq j \leq n} g(U_i, U_j),$$

where

$$g(U_i, U_j) = \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [h(t - U_j) - \mathbb{E}_\pi(h(t - U))] dt.$$

This is a degenerate  $U$ -statistics of order 2, we can rewrite it as

$$W_2(\varphi_\lambda) = \sum_{1 \leq j < i \leq n} \mathcal{G}(U_i, U_j),$$

where

$$\mathcal{G}(U_i, U_j) = g(U_i, U_j) + g(U_j, U_i).$$

We apply Theorem 3.4 of [61] to  $W_2$  and  $-W_2$  (keeping the same notations of [61]): for all  $\varepsilon_0 > 0$  ( $\varepsilon_0 = 1$  for instance), with probability larger than  $1 - 2 \times 2.77e^{-\alpha}$ ,

$$|W_2(\varphi_\lambda)| \leq 2(1 + \varepsilon_0)^{3/2} C \sqrt{\alpha} + \eta(\varepsilon_0) D \alpha + \beta(\varepsilon_0) B \alpha^{3/2} + \gamma(\varepsilon_0) A \alpha^2, \quad (3.28)$$

where

- $A = \|\mathcal{G}\|_\infty$  and by applying Lemma 3.2 with  $f = h$ , we easily have

$$A \leq 8 \|\varphi_\lambda\|_\infty \int_{\mathbb{R}} h(x) dx, \quad (3.29)$$

- $C^2 = \mathbb{E}(W_2(\varphi_\lambda)^2)$  and with (6.7), we have

$$C^2 \leq 2n(n-1) \left\{ \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2 + \frac{1}{T^2} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2 \right\}, \quad (3.30)$$

- $D = \sup \left\{ \mathbb{E} \left( \sum_{j < i} \mathcal{G}(U_i, U_j) a_i(U_i) b_j(U_j) \right) : \mathbb{E} \left( \sum_{i=2}^n a_i(U_i)^2 \right), \mathbb{E} \left( \sum_{j=1}^{n-1} b_j(U_j)^2 \right) \leq 1 \right\},$

$$D = \sup \left\{ \mathbb{E} \left( \sum_{i=2}^n \sum_{j=1}^{i-1} g(U_i, U_j) a_i(U_i) b_j(U_j) + \sum_{j=1}^{n-1} \sum_{i=j+1}^n g(U_j, U_i) a_i(U_i) b_j(U_j) \right) : \mathbb{E} \left( \sum_{i=2}^n a_i(U_i)^2 \right) \leq 1, \mathbb{E} \left( \sum_{j=1}^{n-1} b_j(U_j)^2 \right) \leq 1 \right\}.$$

But, with the conditions on the  $a_i$ 's and the  $b_j$ 's, we have:

$$\begin{aligned} & \mathbb{E} \left( \sum_{i=2}^n \sum_{j=1}^{i-1} g(U_i, U_j) a_i(U_i) b_j(U_j) \right) \\ & \leq \int_{\mathbb{R}} \mathbb{E} \left( \sum_{i=2}^n |\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))| |a_i(U_i)| \right) \\ & \quad \times \mathbb{E} \left( \sum_{j=1}^{n-1} |h(t - U_j) - \mathbb{E}_\pi(h(t - U))| |b_j(U_j)| \right) dt \\ & \leq \int_{\mathbb{R}} \sqrt{(n-1) \text{Var}_\pi(\varphi_\lambda(t - U))} \mathbb{E} \left( \sum_{j=1}^{n-1} |h(t - U_j) - \mathbb{E}_\pi(h(t - U))| |b_j(U_j)| \right) dt \\ & \leq \sqrt{\frac{n-1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx} \mathbb{E} \left( \sum_{j=1}^{n-1} \int_{\mathbb{R}} |h(t - U_j) - \mathbb{E}_\pi(h(t - U))| |b_j(U_j)| dt \right) \\ & \leq \sqrt{\frac{n-1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx} \mathbb{E} \left( 2 \sum_{j=1}^{n-1} |b_j(U_j)| \int_{\mathbb{R}} h(x) dx \right) \\ & \leq 2(n-1) \sqrt{\frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx} \int_{\mathbb{R}} h(x) dx, \end{aligned}$$

using Lemma 3.1 and Lemma 3.2. Inverting the  $a_i$ 's and the  $b_j$ 's, the same computations apply to the second term and we obtain

$$D \leq 4(n-1) \sqrt{\frac{1}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx} \int_{\mathbb{R}} h(x) dx, \quad (3.31)$$

- $B^2 = \max \left\{ \sup_{u,i} \left( \sum_{j=1}^{i-1} \mathbb{E}(\mathcal{G}(u, U_j)^2) \right), \sup_{u,j} \left( \sum_{i=j+1}^n \mathbb{E}(\mathcal{G}(U_i, u)^2) \right) \right\}$  and since  $\mathcal{G}$  is symmetric, we have:

$$\begin{aligned} B^2 &= \sup_{u,i} \left( \sum_{j=1}^{i-1} \mathbb{E}(\mathcal{G}(u, U_j)^2) \right) \\ &= \sup_u \left( \sum_{j=1}^{n-1} \mathbb{E}(\mathcal{G}(u, U_j)^2) \right) \\ &\leq 2 \sup_u \left( \sum_{j=1}^{n-1} [\mathbb{E}(g(u, U_j)^2) + \mathbb{E}(g(U_j, u)^2)] \right). \end{aligned}$$

But,

$$\begin{aligned} &\mathbb{E}(g(u, U_j)^2) \\ &= \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_{\lambda}(t-u) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))] [h(t-U_j) - \mathbb{E}_{\pi}(h(t-U))] dt \right)^2 \right] \\ &\leq \mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t-u) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 |h(t-U_j) - \mathbb{E}_{\pi}(h(t-U))| dt \right. \\ &\quad \left. \times \int_{\mathbb{R}} |h(t-U_j) - \mathbb{E}_{\pi}(h(t-U))| dt \right] \\ &\leq 2\mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t-u) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 |h(t-U_j) - \mathbb{E}_{\pi}(h(t-U))| dt \right] \int_{\mathbb{R}} h(x) dx \\ &\leq \frac{4}{T} \int_{\mathbb{R}} [\varphi_{\lambda}(t-u) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 dt \left( \int_{\mathbb{R}} h(x) dx \right)^2 \\ &\leq \frac{16}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2 \end{aligned}$$

and in the same way

$$\begin{aligned} &\mathbb{E}(g(U_j, u)^2) \\ &= \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_{\lambda}(t-U_j) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))] [h(t-u) - \mathbb{E}_{\pi}(h(t-U))] dt \right)^2 \right] \\ &\leq \mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t-U_j) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 |h(t-u) - \mathbb{E}_{\pi}(h(t-U))| dt \right. \\ &\quad \left. \times \int_{\mathbb{R}} |h(t-u) - \mathbb{E}_{\pi}(h(t-U))| dt \right] \\ &\leq 2\mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t-U_j) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 |h(t-u) - \mathbb{E}_{\pi}(h(t-U))| dt \right] \int_{\mathbb{R}} h(x) dx \end{aligned}$$

$$\begin{aligned} &\leq \frac{2}{T} \int_{\mathbb{R}} |h(t-u) - \mathbb{E}_{\pi}(h(t-U))| dt \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \int_{\mathbb{R}} h(x) dx \\ &\leq \frac{4}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2, \end{aligned}$$

by using Lemma 3.1, Lemma 3.2 and Lemma 3.3. Hence,

$$B^2 \leq \frac{40(n-1)}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2. \quad (3.32)$$

Finally, by inequalities (3.26), (3.27) and (3.28) combined with (3.29), (3.32), (3.30) and (3.31), we obtain: for any  $\varepsilon_0 > 0$ , with probability larger than  $1 - (5 + 2 \times 2.77)e^{-\alpha}$ ,

$$\begin{aligned} |\hat{\beta}_{\lambda} - \beta_{\lambda}| &\leq \sqrt{2\alpha \tilde{V}\left(\frac{\varphi_{\lambda}}{n}\right)} + \frac{\alpha}{3} B\left(\frac{\varphi_{\lambda}}{n}\right) \\ &\quad + \left\{ \frac{\sqrt{2\alpha n}}{T} \|\varphi_{\lambda}\|_1 + \frac{2\alpha}{3T} \|\varphi_{\lambda}\|_1 + 2(1+\varepsilon_0)^{3/2} \sqrt{2\alpha} \sqrt{\frac{1}{T} \|\varphi_{\lambda}\|_2^2 + \frac{1}{T^2} \|\varphi_{\lambda}\|_1^2} \right. \\ &\quad \left. + 4\eta(\varepsilon_0)\alpha \sqrt{\frac{1}{T} \|\varphi_{\lambda}\|_2^2} + \beta(\varepsilon_0)\alpha^{3/2} \sqrt{\frac{40}{nT} \|\varphi_{\lambda}\|_2^2} + \frac{8}{n} \gamma(\varepsilon_0)\alpha^2 \|\varphi_{\lambda}\|_{\infty} \right\} \|h\|_1 \\ &\leq \sqrt{2\alpha \tilde{V}\left(\frac{\varphi_{\lambda}}{n}\right)} + \frac{\alpha}{3} B\left(\frac{\varphi_{\lambda}}{n}\right) \\ &\quad + \left\{ 2M_{\psi,1} \frac{\sqrt{\alpha n}}{T} + 2\sqrt{2}M_{\psi,1} \frac{\alpha}{3T} + 2\sqrt{2}(1+\varepsilon_0)^{3/2} M_{\psi,2} \sqrt{\frac{\alpha}{T}} + 4(1+\varepsilon_0)^{3/2} M_{\psi,1} \frac{\sqrt{\alpha}}{T} \right. \\ &\quad \left. + 4\eta(\varepsilon_0)M_{\psi,2} \frac{\alpha}{\sqrt{T}} + \sqrt{40}\beta(\varepsilon_0)M_{\psi,2} \frac{\alpha^{3/2}}{\sqrt{nT}} + 8\gamma(\varepsilon_0)M_{\psi,\infty} \frac{2^{j_0/2}\alpha^2}{n} \right\} \|h\|_1, \end{aligned}$$

because  $2^{-j/2} \leq \sqrt{2}$  if  $-1 \leq j \leq j_0$ . We denote  $b$  the quantity between braces above.

This upper bound depends on  $h$  (via  $\|h\|_1$ ) and this potential threshold could not be used for applications because  $h$  is unknown. So we overestimate  $\|h\|_1$  by  $\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}$  and we have a threshold that does not depend on  $h$ . So, for any value of  $\kappa \in ]0; 1[$ , by fixing  $\alpha = \kappa^2 j_0 \gamma$  with  $\gamma > 0$ , we define for all  $\lambda$  in  $\Gamma$ ,

$$\eta_{\lambda}(\gamma, \Delta) = \sqrt{2j_0\gamma \tilde{V}\left(\frac{\varphi_{\lambda}}{n}\right)} + \frac{j_0\gamma}{3} B\left(\frac{\varphi_{\lambda}}{n}\right) + \Delta \frac{N_{\mathbb{R}}}{n},$$

where

$$\begin{aligned} \Delta = (1+\varepsilon_0) \left\{ 2M_{\psi,1} \frac{\sqrt{j_0\gamma n}}{T} + 2\sqrt{2}M_{\psi,1} \frac{j_0\gamma}{3T} + 2\sqrt{2}(1+\varepsilon_0)^{3/2} M_{\psi,2} \sqrt{\frac{j_0\gamma}{T}} + 4\eta(\varepsilon_0)M_{\psi,2} \frac{j_0\gamma}{\sqrt{T}} \right. \\ \left. + 4(1+\varepsilon_0)^{3/2} M_{\psi,1} \frac{\sqrt{j_0\gamma}}{T} + \sqrt{40}\beta(\varepsilon_0)M_{\psi,2} \frac{j_0^{3/2}\gamma^{3/2}}{\sqrt{nT}} + 8\gamma(\varepsilon_0)M_{\psi,\infty} \frac{2^{j_0/2}j_0^2\gamma^2}{n} \right\}. \end{aligned}$$

Thus, for all  $\lambda$  in  $\Gamma$ ,

$$\begin{aligned} &\mathbb{P}(|\hat{\beta}_{\lambda} - \beta_{\lambda}| > \kappa \eta_{\lambda}(\gamma, \Delta)) \\ &\leq \mathbb{P}\left(|\hat{\beta}_{\lambda} - \beta_{\lambda}| > \sqrt{2\alpha \tilde{V}\left(\frac{\varphi_{\lambda}}{n}\right)} + \frac{\alpha}{3} B\left(\frac{\varphi_{\lambda}}{n}\right) + b \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}, \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} > \|h\|_1\right) \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \sqrt{2\alpha\tilde{V} \left( \frac{\varphi_\lambda}{n} \right)} + \frac{\alpha}{3} B \left( \frac{\varphi_\lambda}{n} \right) + b \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}, \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leq \|h\|_1 \right) \\
& \leq \mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \sqrt{2\alpha\tilde{V} \left( \frac{\varphi_\lambda}{n} \right)} + \frac{\alpha}{3} B \left( \frac{\varphi_\lambda}{n} \right) + b\|h\|_1 \right) + \mathbb{P} \left( \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leq \|h\|_1 \right) \\
& \leq (5 + 2 \times 2.77)e^{-\alpha} + \mathbb{P} \left( \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leq \|h\|_1 \right),
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{P} \left( \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leq \|h\|_1 \right) & = \mathbb{P} \left( N_{\mathbb{R}} - n\|h\|_1 \leq -\frac{\varepsilon_0 n \|h\|_1}{1+\varepsilon_0} \right) \\
& \leq \exp(-g(\varepsilon_0)n\|h\|_1),
\end{aligned}$$

using Proposition 7 of [90] with  $g(\varepsilon_0) = \frac{1}{1+\varepsilon_0} \left( \log \frac{1}{1+\varepsilon_0} - 1 \right) + 1$ .

Therefore, Assumption (A1) is true if we take  $\omega = (5 + 2 \times 2.77)e^{-\kappa^2 j_0 \gamma} + \exp(-g(\varepsilon_0)n\|h\|_1)$ , with  $\gamma > 0$  and  $\varepsilon_0 > 0$ . Furthermore, the threshold (3.7) that lies at the heart of the chapter is achieved by rewriting  $\Delta$  by grouping the constants into one:

$$\Delta = d(\gamma, \|\psi\|_1, \|\psi\|_2, \|\psi\|_\infty) \left\{ \frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T} \right\} \quad (3.33)$$

with

$$\begin{aligned}
& d(\gamma, \|\psi\|_1, \|\psi\|_2, \|\psi\|_\infty) \\
& = (1+\varepsilon_0) \left\{ 2\sqrt{\gamma}M_{\psi,1} + \frac{2\sqrt{2}}{3}\gamma M_{\psi,1} + 2\sqrt{2}(1+\varepsilon_0)^{3/2}\sqrt{\gamma}M_{\psi,2} + 4(1+\varepsilon_0)^{3/2}\sqrt{\gamma}M_{\psi,1} \right. \\
& \quad \left. + 4\eta(\varepsilon_0)\gamma M_{\psi,2} + \sqrt{40}\beta(\varepsilon_0)\gamma^{3/2}M_{\psi,2} + 8\gamma(\varepsilon_0)\gamma^2 M_{\psi,\infty} \right\}, \quad (3.34)
\end{aligned}$$

where  $\beta(\varepsilon_0)$ ,  $\gamma(\varepsilon_0)$  and  $\eta(\varepsilon_0)$  are defined in [61] with  $\varepsilon_0 = 1$ .

### 3.A.3 Proof of Assumption (A2) of Theorem 3.3

Let  $\lambda \in \Gamma$  be fixed. For any  $p \geq 1$ ,

$$\begin{aligned}
\mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) & = \mathbb{E} \left( \left| G \left( \frac{\varphi_\lambda}{n} \right) - \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx \right|^{2p} \right) \\
& = \frac{1}{n^{2p}} \mathbb{E} \left( |G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n) + W(\varphi_\lambda)|^{2p} \right) \\
& \leq \frac{2^{2p-1}}{n^{2p}} \left[ \mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)|^{2p}) + \mathbb{E}(|W(\varphi_\lambda)|^{2p}) \right]. \quad (3.35)
\end{aligned}$$

Now, let us give an upper bound of each term of the right-hand side of the previous inequality.

We first study the first term of (3.35). We have:

$$\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)|^{2p}) = \mathbb{E}[\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)|^{2p}) | U_1, \dots, U_n)]$$

and conditionally to the  $U_i$ 's,  $N$  is a Poisson process. We apply Lemma 3.4: for any  $p \geq 1$ , there exists a positive constant  $C(p)$  only depending on  $p$  such that

$$\begin{aligned} & \mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)|^{2p} | U_1, \dots, U_n) \\ & \leq C(p) \left( \int_{\mathbb{R}} \left| \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right|^{2p} \sum_{j=1}^n h(t - U_j) dt + V(\varphi_\lambda)^p \right). \end{aligned} \quad (3.36)$$

On the one hand, we provide a control in expectation of the first term of (3.36). We have:

$$\begin{aligned} & \mathbb{E} \left( \int_{\mathbb{R}} \left| \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right|^{2p} \sum_{j=1}^n h(t - U_j) dt \right) \\ & = \mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n \left| \varphi_\lambda(t - U_j) + \sum_{i \neq j} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} h(t - U_j) dt \right) \\ & \leq 2^{2p-1} \left[ \mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n |\varphi_\lambda(t - U_j)|^{2p} h(t - U_j) dt \right) \right. \\ & \quad \left. + \mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n \left| \sum_{i \neq j} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} h(t - U_j) dt \right) \right], \end{aligned}$$

with

$$\mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n |\varphi_\lambda(t - U_j)|^{2p} h(t - U_j) dt \right) = n \int_{\mathbb{R}} |\varphi_\lambda(x)|^{2p} h(x) dx \leq n \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_\infty$$

and

$$\begin{aligned} & \mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n \left| \sum_{i \neq j} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} h(t - U_j) dt \right) \\ & = \sum_{j=1}^n \int_{\mathbb{R}} \mathbb{E} \left( \left| \sum_{i \neq j} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} \right) \mathbb{E}(h(t - U_j)) dt \\ & = n \int_{\mathbb{R}} \mathbb{E} \left( \left| \sum_{i=1}^{n-1} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} \right) \mathbb{E}_\pi(h(t - U)) dt. \end{aligned}$$

By applying Rosenthal's inequality, there exists a positive constant  $C(p)$  only depending on  $p$  such that

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{i=1}^{n-1} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|^{2p} \right) \\ & \leq C(p) \left( (n-1) \mathbb{E} \left( |\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))|^{2p} \right) + (n-1)^p [\text{Var}_\pi(\varphi_\lambda(t - U))]^p \right). \end{aligned}$$

But,

$$\mathbb{E} \left( |\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))|^{2p} \right)$$



$$\begin{aligned}
&\leq K \left( \mathbb{E}_\pi(|\varphi_\lambda(t-U)|^{2p}) + |\mathbb{E}_\pi(\varphi_\lambda(t-U))|^{2p} \right) \\
&\leq K \left( \frac{1}{T} \int_0^T |\varphi_\lambda(t-u)|^{2p} du + \left| \frac{1}{T} \int_0^T \varphi_\lambda(t-u) du \right|^{2p} \right) \\
&\leq K \left( \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \|\varphi_\lambda\|_\infty^{2p-2} + \frac{1}{T^{2p}} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^{2p} \right),
\end{aligned}$$

with  $K$  a positive constant only depending on  $p$  and using Lemma 3.1,

$$\text{Var}_\pi(\varphi_\lambda(t-U)) \leq \frac{1}{T} \int_{\mathbb{R}} \varphi_\lambda^2(x) dx.$$

Thus,

$$\begin{aligned}
&\mathbb{E} \left( \int_{\mathbb{R}} \sum_{j=1}^n \left| \sum_{i \neq j} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \right|^{2p} h(t-U_j) dt \right) \\
&\leq n K \left( \frac{n-1}{T} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} + \frac{n-1}{T^{2p}} \|\varphi_\lambda\|_1^{2p} + \frac{(n-1)^p}{T^p} \|\varphi_\lambda\|_2^{2p} \right) \int_{\mathbb{R}} \mathbb{E}_\pi(h(t-U)) dt \\
&\leq K \left( \frac{n^2}{T} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} + \frac{n^2}{T^{2p}} \|\varphi_\lambda\|_1^{2p} + \frac{n^{p+1}}{T^p} \|\varphi_\lambda\|_2^{2p} \right) \|h\|_1,
\end{aligned}$$

using Lemma 3.2.

Therefore, we have the following control of the first term of (3.36)

$$\begin{aligned}
&\mathbb{E} \left( \int_{\mathbb{R}} \left| \sum_{i=1}^n \left[ \varphi_\lambda(t-U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t-U)) \right] \right|^{2p} \sum_{j=1}^n h(t-U_j) dt \right) \\
&\leq K \left[ n \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_\infty + \left( \frac{n^2}{T} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} + \frac{n^2}{T^{2p}} \|\varphi_\lambda\|_1^{2p} + \frac{n^{p+1}}{T^p} \|\varphi_\lambda\|_2^{2p} \right) \|h\|_1 \right],
\end{aligned} \tag{3.37}$$

Now let us provide a control in expectation of the second term of (3.36), i.e  $V(\varphi_\lambda)^p$ . First, we recall that  $V(\varphi_\lambda) = \text{Var}(G(\varphi_\lambda)|U_1, \dots, U_n)$  and we remark that  $\mathbb{E}(V(\varphi_\lambda)^p) \leq [\mathbb{E}(V(\varphi_\lambda)^{2p})]^{1/2}$  (using the Cauchy-Schwarz inequality). So, we focus on the moments of  $V(\varphi_\lambda)$  of any order  $m \geq 2$ .

Let  $m \geq 2$ . According to the expression (6.2) of  $V(\varphi_\lambda)$ , we have:

$$\begin{aligned}
&V(\varphi_\lambda) \\
&= n \int_{\mathbb{R}} \varphi_\lambda^2(x) h(x) dx + 2 \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \varphi_\lambda(t-U_j) h(t-U_j) dt \\
&\quad + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \sum_{k \neq j} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] [\varphi_\lambda(t-U_k) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] h(t-U_j) dt \\
&= n \int_{\mathbb{R}} \varphi_\lambda^2(x) h(x) dx + 2 \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U) h(t-U)) dt \\
&\quad + 2 \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] [\varphi_\lambda(t-U_j) h(t-U_j) - \mathbb{E}_\pi(\varphi_\lambda(t-U) h(t-U))] dt
\end{aligned}$$

$$\begin{aligned}
& + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \sum_{k \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \\
& + \sum_{j=1}^n \int_{\mathbb{R}} \sum_{i \neq j} \sum_{k \neq j} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \\
& \quad \times [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt.
\end{aligned}$$

This formula provides a decomposition of  $V(\varphi_{\lambda})$  in a sum of degenerate  $U$ -statistics of order 0, 1, 2 and 3. Indeed

$$V(\varphi_{\lambda}) = \mathcal{W}_0(\varphi_{\lambda}) + \mathcal{W}_1(\varphi_{\lambda}) + \mathcal{W}_2(\varphi_{\lambda}) + \mathcal{W}_3(\varphi_{\lambda}),$$

with  $\mathcal{W}_i(\varphi_{\lambda})$  is a degenerate  $U$ -statistic of order  $i$  defined as follows:

$$\begin{aligned}
\mathcal{W}_3(\varphi_{\lambda}) = & \sum_{1 \leq i \neq j \neq k \leq n} \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \\
& \times [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt,
\end{aligned}$$

$\mathcal{W}_2(\varphi_{\lambda})$

$$\begin{aligned}
& = 2 \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_j) h(t - U_j) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U) h(t - U))] dt \\
& + (n - 2) \sum_{i \neq k} \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [\varphi_{\lambda}(t - U_k) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \\
& + \sum_{i \neq j} \int_{\mathbb{R}} \left[ [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 - \text{Var}_{\pi}(\varphi_{\lambda}(t - U)) \right] [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt,
\end{aligned}$$

$$\begin{aligned}
\mathcal{W}_1(\varphi_{\lambda}) = & 2(n - 1) \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U) h(t - U)) dt \\
& + (n - 1) \sum_{i=1}^n \int_{\mathbb{R}} \left[ [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 - \text{Var}_{\pi}(\varphi_{\lambda}(t - U)) \right] \mathbb{E}_{\pi}(h(t - U)) dt \\
& + (n - 1) \sum_{j=1}^n \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t - U)) [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{W}_0(\varphi_{\lambda}) & = n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + n(n - 1) \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t - U)) \mathbb{E}_{\pi}(h(t - U)) dt \\
& = \mathbb{E}(V(\varphi_{\lambda})) \leq n \int_{\mathbb{R}} \varphi_{\lambda}^2(x) h(x) dx + \frac{n(n - 1)}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \int_{\mathbb{R}} h(x) dx, \tag{3.38}
\end{aligned}$$

by using (6.3) and (6.4).

First, we are interested in the moments of  $\mathcal{W}_1(\varphi_{\lambda})$  that we write:

$$\mathcal{W}_1(\varphi_{\lambda}) = \mathcal{W}_{1,1}(\varphi_{\lambda}) + \mathcal{W}_{1,2}(\varphi_{\lambda}) + \mathcal{W}_{1,3}(\varphi_{\lambda}),$$

with:

- $\mathcal{W}_{1,1}(\varphi_\lambda) = 2(n-1) \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt.$

We have:

$$\begin{aligned} & \mathbb{E}(|\mathcal{W}_{1,1}(\varphi_\lambda)|^m) \\ &= 2^m(n-1)^m \mathbb{E} \left( \left| \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right|^m \right) \\ &\leq 2^m(n-1)^m \times C(m) \left( n \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right|^m \right) \right. \\ &\quad \left. + n^{m/2} \left[ \text{Var} \left( \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right) \right]^{m/2} \right), \end{aligned}$$

using Rosenthal's inequality, where  $C(m)$  is a positive constant only depending on  $m$ . But, applying Lemma 3.1, Lemma 3.2 and Lemma 3.3,

$$\begin{aligned} & \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right|^m \right) \\ &\leq \frac{2^m}{T^m} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^m \left( \int_{\mathbb{R}} |\varphi_\lambda(x)|h(x) dx \right)^m \end{aligned}$$

and

$$\begin{aligned} & \text{Var} \left( \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))] \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right) \\ &= \text{Var} \left( \int_{\mathbb{R}} \varphi_\lambda(t-U_1) \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)) dt \right) \\ &\leq \mathbb{E} \left[ \left( \int_{\mathbb{R}} |\varphi_\lambda(t-U_1)| \mathbb{E}_\pi(|\varphi_\lambda(t-U)|h(t-U)) dt \right)^2 \right] \\ &\leq \frac{1}{T^2} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^2 \left( \int_{\mathbb{R}} |\varphi_\lambda(x)|h(x) dx \right)^2. \end{aligned}$$

So,

$$\begin{aligned} \mathbb{E}(|\mathcal{W}_{1,1}(\varphi_\lambda)|^m) &\leq 2^m(n-1)^m \times C(m) \left( \frac{2^m n}{T^m} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^m \left( \int_{\mathbb{R}} |\varphi_\lambda(x)|h(x) dx \right)^m \right. \\ &\quad \left. + \frac{n^{m/2}}{T^m} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^m \left( \int_{\mathbb{R}} |\varphi_\lambda(x)|h(x) dx \right)^m \right) \\ &\leq K_{1,1} \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^m \left( \int_{\mathbb{R}} |\varphi_\lambda(x)|h(x) dx \right)^m \\ &\leq K_{1,1} \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^{2m} \|h\|_\infty^m, \end{aligned} \tag{3.39}$$

with  $K_{1,1}$  a positive constant only depending on  $m$ .

- $\mathcal{W}_{1,2}(\varphi_\lambda) = (n-1) \sum_{i=1}^n \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt.$

We have:

$$\begin{aligned} & \mathbb{E}(|\mathcal{W}_{1,2}(\varphi_\lambda)|^m) \\ &= (n-1)^m \mathbb{E} \left( \left| \sum_{i=1}^n \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt \right|^m \right) \\ &\leq n^m C(m) \left( n \mathbb{E} \left( \left| \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt \right|^m \right) \right. \\ &\quad \left. + n^{m/2} \left[ \text{Var} \left( \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt \right) \right]^{m/2} \right), \end{aligned}$$

using Rosenthal's inequality, where  $C(m)$  is a positive constant only depending on  $m$ . But, applying Lemma 3.1, Lemma 3.2 and Lemma 3.3,

$$\begin{aligned} & \mathbb{E} \left( \left| \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt \right|^m \right) \\ &\leq K_{1,2} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m, \end{aligned}$$

with  $K_{1,2}$  a positive constant only depending on  $m$  and

$$\begin{aligned} & \text{Var} \left( \int_{\mathbb{R}} \left[ [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 - \text{Var}_\pi(\varphi_\lambda(t-U)) \right] \mathbb{E}_\pi(h(t-U)) dt \right) \\ &= \text{Var} \left( \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 \mathbb{E}_\pi(h(t-U)) dt \right) \\ &\leq \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 \mathbb{E}_\pi(h(t-U)) dt \right)^2 \right] \\ &\leq \frac{1}{T^2} \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]^2 dt \right)^2 \right] \left( \int_{\mathbb{R}} h(x) dx \right)^2 \\ &\leq K_{1,2} \frac{1}{T^2} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2 \end{aligned}$$

So,

$$\begin{aligned} & \mathbb{E}(|\mathcal{W}_{1,2}(\varphi_\lambda)|^m) \\ &\leq K_{1,2} n^m \left( \frac{n}{T^m} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m + \frac{n^{m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m \right) \\ &\leq K_{1,2} \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m. \end{aligned} \tag{3.40}$$

$$\bullet \mathcal{W}_{1,3}(\varphi_\lambda) = (n-1) \sum_{j=1}^n \int_{\mathbb{R}} \text{Var}_\pi(\varphi_\lambda(t-U)) [h(t-U_j) - \mathbb{E}_\pi(h(t-U))] dt.$$

We have:

$$\begin{aligned} & \mathbb{E}(|\mathcal{W}_{1,3}(\varphi_\lambda)|^m) \\ &= (n-1)^m \mathbb{E} \left( \left| \sum_{j=1}^n \int_{\mathbb{R}} \text{Var}_\pi(\varphi_\lambda(t-U)) [h(t-U_j) - \mathbb{E}_\pi(h(t-U))] dt \right|^m \right) \end{aligned}$$

$$\begin{aligned} &\leq (n-1)^m \times C(m) \left( n \mathbb{E} \left( \left| \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) [h(t-U_1) - \mathbb{E}_{\pi}(h(t-U))] dt \right|^m \right) \right. \\ &\quad \left. + n^{m/2} \left[ \text{Var} \left( \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) [h(t-U_1) - \mathbb{E}_{\pi}(h(t-U))] dt \right) \right]^{m/2} \right), \end{aligned}$$

using Rosenthal's inequality, where  $C(m)$  is a positive constant only depending on  $m$ . But, applying Lemma 3.1, Lemma 3.2 and Lemma 3.3,

$$\mathbb{E} \left( \left| \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) [h(t-U_1) - \mathbb{E}_{\pi}(h(t-U))] dt \right|^m \right) \leq \frac{2^m}{T^m} \|\varphi_{\lambda}\|^{2m} \left( \int_{\mathbb{R}} h(x) dx \right)^m,$$

and

$$\begin{aligned} &\text{Var} \left( \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) [h(t-U_1) - \mathbb{E}_{\pi}(h(t-U))] dt \right) \\ &= \text{Var} \left( \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) h(t-U_1) dt \right) \\ &\leq \mathbb{E} \left[ \left( \int_{\mathbb{R}} \text{Var}_{\pi}(\varphi_{\lambda}(t-U)) h(t-U_1) dt \right)^2 \right] \\ &\leq \frac{1}{T^2} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^2 \mathbb{E} \left[ \left( \int_{\mathbb{R}} h(t-U_1) dt \right)^2 \right] \\ &\leq \frac{1}{T^2} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2. \end{aligned}$$

So,

$$\begin{aligned} \mathbb{E}(|\mathcal{W}_{1,3}(\varphi_{\lambda})|^m) &\leq (n-1)^m \times C(m) \left( \frac{2^m n}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m \right. \\ &\quad \left. + \frac{n^{m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m \right) \\ &\leq K_{1,3} \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m, \end{aligned} \quad (3.41)$$

with  $K_{1,3}$  a positive constant only depending on  $m$ .

Next we deal with the moments of  $\mathcal{W}_2(\varphi_{\lambda})$  that we write:

$$\mathcal{W}_2(\varphi_{\lambda}) = \mathcal{W}_{2,1}(\varphi_{\lambda}) + \mathcal{W}_{2,2}(\varphi_{\lambda}) + \mathcal{W}_{2,3}(\varphi_{\lambda}),$$

with:

$$\bullet \mathcal{W}_{2,1}(\varphi_{\lambda}) = 2 \sum_{1 \leq i \neq j \leq n} \int_{\mathbb{R}} [\varphi_{\lambda}(t-U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))] [\varphi_{\lambda}(t-U_j) h(t-U_j) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t-U)) h(t-U)] dt.$$

We want to use Theorem 8.1.6 of [32] (a moment inequality for  $U$ -statistics using decoupling) so we write:

$$\mathcal{W}_{2,1}(\varphi_{\lambda}) = 2 \sum_{1 \leq i \neq j \leq n} f(U_i, U_j),$$

where

$$f(U_i, U_j) = \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_j)h(t - U_j) - \mathbb{E}_\pi(\varphi_\lambda(t - U)h(t - U))] dt.$$

There exists a positive constant  $C_{2,m}$  depending on  $m$  only such that

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{1 \leq i \neq j \leq n} f(U_i, U_j) \right|^m \right) \\ & \leq C_{2,m} n^m \mathbb{E}(|f(U_1, U_2)|^m) \\ & \leq C_{2,m} n^m \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [(\varphi_\lambda h)(t - U_2) - \mathbb{E}_\pi((\varphi_\lambda h)(t - U))] dt \right|^m \right) \\ & \leq C_{2,m} n^m \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 dt \right)^{m/2} \right] \\ & \quad \times \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t - U_2)h(t - U_2) - \mathbb{E}_\pi(\varphi_\lambda(t - U)h(t - U))]^2 dt \right)^{m/2} \right] \\ & \leq K_{2,1} n^m \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^{m/2} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x)h^2(x) dx \right)^{m/2} \\ & \leq K_{2,1} n^m \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \|h\|_\infty^m, \end{aligned}$$

by applying Lemma 3.1, Lemma 3.2 and Lemma 3.3 and setting  $K_{2,1}$  a positive constant only depending on  $m$ . So,

$$\mathbb{E}(|\mathcal{W}_{2,1}(f)|^m) \leq K_{2,1} n^m \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \|h\|_\infty^m. \quad (3.42)$$

- We may write:  $\mathcal{W}_{2,2}(\varphi_\lambda) = (n-2) \sum_{1 \leq i \neq k \leq n} f(U_i, U_k)$ , where

$$f(U_i, U_k) = \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_k) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt.$$

We use Theorem 8.1.6 of [32]: there exists a positive constant  $C_{2,m}$  depending on  $m$  only such that

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{1 \leq i \neq k \leq n} f(U_i, U_k) \right|^m \right) \\ & \leq C_{2,m} n^m \mathbb{E}(|f(U_1, U_2)|^m) \\ & \leq C_{2,m} n^m \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_2) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right|^m \right) \\ & \leq K_{2,2} \frac{n^m}{T^m} \left\{ \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 dt \right)^{m/2} \right]^2 \left( \int_{\mathbb{R}} h(x) dx \right)^m \right\} \\ & \leq K_{2,2} \frac{n^m}{T^m} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m, \end{aligned}$$

by applying Lemma 3.1, Lemma 3.2 and Lemma 3.3 and setting  $K_{2,2}$  a positive constant only depending on  $m$ . So,

$$\mathbb{E}(|\mathcal{W}_{2,2}(f)|^m) \leq K_{2,2} \frac{n^{2m}}{T^m} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m. \quad (3.43)$$

- We may write:  $\mathcal{W}_{2,3}(\varphi_\lambda) = \sum_{1 \leq i \neq j \leq n} f(U_i, U_j)$ , where

$$f(U_i, U_j) = \int_{\mathbb{R}} \left[ [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 - \text{Var}_\pi(\varphi_\lambda(t - U)) \right] [h(t - U_j) - \mathbb{E}_\pi(h(t - U))] dt.$$

We use Theorem 8.1.6 of [32]: there exists a positive constant  $C_{2,m}$  depending on  $m$  only such that

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{1 \leq i \neq j \leq n} f(U_i, U_j) \right|^m \right) \\ & \leq C_{2,m} n^m \mathbb{E}(|f(U_1, U_2)|^m) \\ & \leq C_{2,m} n^m \mathbb{E} \left( \left| \int_{\mathbb{R}} [[\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 - \text{Var}_\pi(\varphi_\lambda(t - U))] \right. \right. \\ & \quad \left. \left. \times [h(t - U_2) - \mathbb{E}_\pi(h(t - U))] dt \right|^m \right) \\ & \leq K_{2,3} n^m \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 - \text{Var}_\pi(\varphi_\lambda(t - U)) \right| dt \right)^m \right] \|h\|_\infty^m \\ & \leq K_{2,3} n^m \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \|h\|_\infty^m, \end{aligned}$$

by applying Lemma 3.1, Lemma 3.2 and Lemma 3.3 and setting  $K_{2,3}$  a positive constant only depending on  $m$ . So,

$$\mathbb{E}(|\mathcal{W}_{2,3}(\varphi_\lambda)|^m) \leq K_{2,3} n^m \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \|h\|_\infty^m. \quad (3.44)$$

And finally, we focus on the moments of  $\mathcal{W}_3(\varphi_\lambda)$  that we write:

$$\mathcal{W}_3(\varphi_\lambda) = \sum_{1 \leq i \neq j \neq k \leq n} f(U_i, U_j, U_k),$$

where

$$\begin{aligned} & f(U_i, U_j, U_k) \\ & = \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_k) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [h(t - U_j) - \mathbb{E}_\pi(h(t - U))] dt. \end{aligned}$$

We use Theorem 8.1.6 of [32]: there exists a positive constant  $C_{3,m}$  depending on  $m$  only such that

$$\mathbb{E} \left( \left| \sum_{1 \leq i \neq j \neq k \leq n} f(U_i, U_j, U_k) \right|^m \right)$$

$$\begin{aligned}
&\leq C_{3,m} n^{3m/2} \mathbb{E}(|f(U_1, U_2, U_3)|^m) \\
&\leq C_{3,m} n^{3m/2} \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_2) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right. \right. \\
&\quad \left. \left. \times [h(t - U_3) - \mathbb{E}_\pi(h(t - U))] dt \right|^m \right) \\
&\leq K_3 n^{3m/2} \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_2) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right| dt \right)^m \right] \|h\|_\infty^m,
\end{aligned}$$

by applying Lemma 3.1, Lemma 3.2 and Lemma 3.3 and setting  $K_3$  a positive constant only depending on  $m$ . Furthermore, using the support properties of the biorthogonal wavelet bases considered in this chapter, we have

$$\begin{aligned}
&\mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [\varphi_\lambda(t - U_2) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right| dt \right)^m \right] \\
&= \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \varphi_\lambda(t - U_2) - \varphi_\lambda(t - U_1) \mathbb{E}_\pi(\varphi_\lambda(t - U)) - \varphi_\lambda(t - U_2) \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right. \right. \right. \\
&\quad \left. \left. \left. + [\mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 \right| dt \right)^m \right] \\
&\leq K_3 \left\{ \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \varphi_\lambda(t - U_2) \right| dt \right)^m \right] + \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right| dt \right)^m \right] \right. \\
&\quad \left. + \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\mathbb{E}_\pi(\varphi_\lambda(t - U))]^2 dt \right)^m \right] \right\},
\end{aligned}$$

with:

$$\begin{aligned}
\mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \varphi_\lambda(t - U_2) \right| dt \right)^m \right] &= \frac{1}{T^2} \int_0^T du_1 \int_0^T du_2 \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - u_1) \varphi_\lambda(t - u_2) \right| dt \right)^m \\
&\leq \frac{1}{T^2} \int_0^T du_1 \int_{u_1-2M}^{u_1+2M} du_2 \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m \\
&\leq \frac{4M}{T} \left( \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right| dt \right)^m \right] \\
&= \mathbb{E} \left[ \left( \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \left( \frac{1}{T} \int_0^T \varphi_\lambda(t - u) du \right) \right| dt \right)^m \right] \\
&\leq \frac{1}{T^m} \mathbb{E} \left[ \left( \int_0^T du \int_{\mathbb{R}} \left| \varphi_\lambda(t - U_1) \varphi_\lambda(t - u) \right| dt \right)^m \right] \\
&\leq \frac{1}{T^{m+1}} \int_0^T du_1 \left( \int_0^T du \int_{\mathbb{R}} \left| \varphi_\lambda(t - u_1) \varphi_\lambda(t - u) \right| dt \right)^m \\
&\leq \frac{1}{T^{m+1}} \int_0^T du_1 \left( \int_{u_1-2M}^{u_1+2M} du \int_{\mathbb{R}} \varphi_\lambda^2(x) dx \right)^m
\end{aligned}$$



$$\leq \frac{(4M)^m}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m$$

and

$$\begin{aligned} \int_{\mathbb{R}} [\mathbb{E}_{\pi}(\varphi_{\lambda}(t-U))]^2 dt &= \frac{1}{T^2} \int_{\mathbb{R}} dt \int_0^T du \varphi_{\lambda}(t-u) \int_0^T dv \varphi_{\lambda}(t-v) \\ &= \frac{1}{T^2} \int_0^T du \int_0^T dv \int_{\mathbb{R}} \varphi_{\lambda}(t-u) \varphi_{\lambda}(t-v) dt \\ &\leq \frac{1}{T^2} \int_0^T du \int_{u-2M}^{u+2M} dv \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \\ &\leq \frac{4M}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx. \end{aligned}$$

So,

$$\mathbb{E}(|\mathcal{W}_3(\varphi_{\lambda})|^m) \leq K'_3 n^{3m/2} \left[ \frac{1}{T} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m + \frac{1}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \right] \|h\|_{\infty}^m, \quad (3.45)$$

with  $K'_3$  a positive constant only depending on  $m$  and the compact support of  $\psi$ . Note that if we had used the same method as for the control of the moments of  $W_{2,2}(\varphi_{\lambda})$ , we would not get the correct rate of convergence. We obtain a better rate of convergence thanks to the properties of the biorthogonal wavelet bases used here.

Thus, combining inequalities (3.38), (3.39), (3.40), (3.41), (3.42), (3.43), (3.44) and (3.45) yields

$$\begin{aligned} &\mathbb{E}(V(\varphi_{\lambda})^m) \\ &\leq K \left\{ n^m \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \|h\|_{\infty}^m + \frac{n^{2m}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m \right. \\ &\quad + \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^{2m} \|h\|_{\infty}^m + \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m \\ &\quad + \frac{n^{3m/2}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m + n^m \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \|h\|_{\infty}^m \\ &\quad + \frac{n^{2m}}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \left( \int_{\mathbb{R}} h(x) dx \right)^m + n^m \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \|h\|_{\infty}^m \\ &\quad \left. + n^{3m/2} \left[ \frac{1}{T} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m + \frac{1}{T^m} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^m \right] \|h\|_{\infty}^m \right\} \\ &\leq K \left\{ \frac{n^{3m/2}}{T^m} \|\varphi_{\lambda}\|_1^{2m} \|h\|_{\infty}^m + \frac{n^{2m}}{T^m} \|\varphi_{\lambda}\|_2^{2m} \|h\|_1^m + \left[ n^m + \frac{n^{3m/2}}{T} \right] \|\varphi_{\lambda}\|_2^{2m} \|h\|_{\infty}^m \right\}, \end{aligned}$$

with  $K$  a positive constant only depending on  $m$  and the compact support of  $\psi$ . So, we obtain

$$\mathbb{E}(V(\varphi_{\lambda})^p) \leq K \left\{ \frac{n^{3p/2}}{T^p} \|\varphi_{\lambda}\|_1^{2p} \|h\|_{\infty}^p + \frac{n^{2p}}{T^p} \|\varphi_{\lambda}\|_2^{2p} \|h\|_1^p + \left[ n^p + \frac{n^{3p/2}}{T} \right] \|\varphi_{\lambda}\|_2^{2p} \|h\|_{\infty}^p \right\}, \quad (3.46)$$

with  $K$  a positive constant only depending on  $p$  and the compact support of  $\psi$ .

To conclude for the first term of (3.35), using inequalities (3.37) and (3.46) in (3.36), we have

$$\begin{aligned} & \mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1, \dots, U_n)|^{2p}) \\ & \leq K \left\{ \frac{n^2}{T^{2p}} \|\varphi_\lambda\|_1^{2p} \|h\|_1 + \frac{n^{p+1}}{T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1 + \frac{n^2}{T} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_1 + n \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_\infty \right. \\ & \quad \left. + \frac{n^{3p/2}}{T^p} \|\varphi_\lambda\|_1^{2p} \|h\|_\infty^p + \frac{n^{2p}}{T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1^p + \left[ n^p + \frac{n^{3p/2}}{T} \right] \|\varphi_\lambda\|_2^{2p} \|h\|_\infty^p \right\}. \end{aligned} \quad (3.47)$$

Now, we have to focus on the second term of (3.35). Recall the definition (6.5) of  $W(\varphi_\lambda)$

$$W(\varphi_\lambda) = W_1(\varphi_\lambda) + W_2(\varphi_\lambda),$$

with

$$W_1(\varphi_\lambda) = (n-1) \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt,$$

and

$$W_2(\varphi_\lambda) = \sum_{1 \leq i \neq j \leq n} g(U_i, U_j),$$

where

$$g(U_i, U_j) = \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] [h(t - U_j) - \mathbb{E}_\pi(h(t - U))] dt.$$

So,

$$\mathbb{E}(|W(\varphi_\lambda)|^{2p}) \leq 2^{2p-1} [\mathbb{E}(|W_1(\varphi_\lambda)|^{2p}) + \mathbb{E}(|W_2(\varphi_\lambda)|^{2p})].$$

On the one hand, we have to control  $\mathbb{E}(|W_1(\varphi_\lambda)|^{2p})$ . We use Rosenthal's inequality: there exists a positive constant  $C(p)$  only depending on  $p$  such that

$$\begin{aligned} & \mathbb{E}(|W_1(\varphi_\lambda)|^{2p}) \\ & = (n-1)^{2p} \mathbb{E} \left( \left| \sum_{i=1}^n \int_{\mathbb{R}} [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right|^{2p} \right) \\ & \leq (n-1)^{2p} \times C(p) \left( n \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right|^{2p} \right) \right. \\ & \quad \left. + n^p \left[ \text{Var} \left( \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right) \right]^p \right). \end{aligned}$$

But, applying Lemma 3.1, Lemma 3.2 and Lemma 3.3,

$$\begin{aligned} & \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \mathbb{E}_\pi(h(t - U)) dt \right|^{2p} \right) \\ & \leq \frac{1}{T^{2p}} \mathbb{E} \left( \left| \int_{\mathbb{R}} [\varphi_\lambda(t - U_1) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] dt \right|^{2p} \right) \left( \int_{\mathbb{R}} h(x) dx \right)^{2p} \\ & \leq \frac{2^{2p}}{T^{2p}} \left( \int_{\mathbb{R}} |\varphi_\lambda(x)| dx \right)^{2p} \left( \int_{\mathbb{R}} h(x) dx \right)^{2p} \end{aligned}$$

and

$$\begin{aligned}
& \text{Var} \left( \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_1) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] \mathbb{E}_{\pi}(h(t - U)) dt \right) \\
&= \text{Var} \left( \int_{\mathbb{R}} \varphi_{\lambda}(t - U_1) \mathbb{E}_{\pi}(h(t - U)) dt \right) \\
&\leq \mathbb{E} \left[ \left( \int_{\mathbb{R}} |\varphi_{\lambda}(t - U_1)| \mathbb{E}_{\pi}(h(t - U)) dt \right)^2 \right] \\
&\leq \frac{1}{T^2} \mathbb{E} \left[ \left( \int_{\mathbb{R}} |\varphi_{\lambda}(t - U_1)| dt \right)^2 \right] \left( \int_{\mathbb{R}} h(x) dx \right)^2 \\
&\leq \frac{1}{T^2} \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^2 \left( \int_{\mathbb{R}} h(x) dx \right)^2.
\end{aligned}$$

So,

$$\mathbb{E}(|W_1(\varphi_{\lambda})|^{2p}) \leq K_1 n^{2p} \left[ \frac{n}{T^{2p}} \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^{2p} + \frac{n^p}{T^{2p}} \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^{2p} \right] \left( \int_{\mathbb{R}} h(x) dx \right)^{2p}, \quad (3.48)$$

with  $K_1$  a positive constant only depending on  $p$ .

And on the other hand, we have to control  $\mathbb{E}(|W_2(\varphi_{\lambda})|^{2p})$ . We have:

$$W_2(\varphi_{\lambda}) = \sum_{1 \leq i \neq j \leq n} g(U_i, U_j).$$

We use Theorem 3.3 of [48] associated with Theorem 1 of [33] (we keep the same notations of [48]).

We set  $h_{i,j} = \begin{cases} 0 & \text{if } i = j \\ g & \text{otherwise} \end{cases}$  and we consider  $(U_i^{(1)}, i = 1 \dots n)$  and  $(U_i^{(2)}, i = 1 \dots n)$  two independent copies of  $(U_i, i = 1 \dots n)$ . With Theorem 3.3 of [48], there exists an universal constant  $K$  such that

$$\mathbb{E} \left( \left| \sum_{1 \leq i, j \leq n} h_{i,j}(U_i^{(1)}, U_j^{(2)}) \right|^{2p} \right) \leq K^{2p} [(2p)^p C^{2p} + (2p)^{2p} D^{2p} + (2p)^{3p} B^{2p} + (2p)^{4p} A^{2p}],$$

where

- $A = \max_{i,j} \|h_{i,j}\|_{\infty} = \|g\|_{\infty}$ . But, for all  $(x, y) \in \mathbb{R}^2$ ,

$$\begin{aligned}
|g(x, y)| &= \left| \int_{\mathbb{R}} [\varphi_{\lambda}(t - x) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [h(t - y) - \mathbb{E}_{\pi}(h(t - U))] dt \right| \\
&\leq 4 \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \|h\|_{\infty},
\end{aligned}$$

using Lemma 3.2 with  $f = \varphi_{\lambda}$ . So,

$$A \leq 4 \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \|h\|_{\infty}, \quad (3.49)$$

- $C^2 = \sum_{i,j} \mathbb{E}(h_{i,j}^2(U_i^{(1)}, U_j^{(2)})) = \sum_{i \neq j} \mathbb{E}(g^2(U_i, U_j))$ . But, for all  $i \neq j$ ,

$$\mathbb{E}(g^2(U_i, U_j))$$

$$\begin{aligned}
&\leq \mathbb{E} \left[ \left( \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))] [h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))] dt \right)^2 \right] \\
&\leq \mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 |h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))| dt \right. \\
&\quad \left. \times \int_{\mathbb{R}} |h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))| dt \right] \\
&\leq 2\mathbb{E} \left[ \int_{\mathbb{R}} [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 |h(t - U_j) - \mathbb{E}_{\pi}(h(t - U))| dt \right] \int_{\mathbb{R}} h(x) dx \\
&\leq \frac{4}{T} \int_{\mathbb{R}} \mathbb{E} \left( [\varphi_{\lambda}(t - U_i) - \mathbb{E}_{\pi}(\varphi_{\lambda}(t - U))]^2 \right) dt \left( \int_{\mathbb{R}} h(x) dx \right)^2 \\
&\leq \frac{4}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2,
\end{aligned}$$

using Lemma 3.1, Lemma 3.2 and Lemma 3.3. So,

$$C^2 \leq \frac{4n(n-1)}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2, \quad (3.50)$$

$$\bullet B^2 = \max \left[ \left\| \sum_i \mathbb{E}_1(h_{i,j}^2(U_i, y)) \right\|_{\infty}, \left\| \sum_j \mathbb{E}_2(h_{i,j}^2(x, U_j)) \right\|_{\infty} \right], \text{ with}$$

$$\mathbb{E}_1(h_{i,j}^2(U_i, y)) = \begin{cases} \mathbb{E}_{\pi}(g^2(U, y)) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \leq \frac{4}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2$$

and

$$\mathbb{E}_1(h_{i,j}^2(x, U_j)) = \begin{cases} \mathbb{E}_{\pi}(g^2(x, U)) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \leq \frac{16}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2,$$

using established inequalities to get equation (3.32) in the proof of the assumption (A1). So,

$$B^2 \leq \frac{16n}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \left( \int_{\mathbb{R}} h(x) dx \right)^2, \quad (3.51)$$

$$\bullet D = \sup \left\{ \mathbb{E}(\sum_{i,j} h_{i,j}(U_i^{(1)}, U_j^{(2)}) a_i(U_i^{(1)}) b_j(U_j^{(2)})) : \mathbb{E}(\sum_i a_i^2(U_i)), \mathbb{E}(\sum_j b_j^2(U_j)) \leq 1 \right\}.$$

By using established inequalities to get equation (3.31) in the proof of the assumption (A1), we obtain

$$D \leq 2n \sqrt{\frac{1}{T} \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx} \int_{\mathbb{R}} h(x) dx. \quad (3.52)$$

Moreover, we use the equivalence of Theorem 3.3 of [48] and the decoupling inequality provided in Theorem 1 of [33] to obtain the following upper bound of  $\mathbb{E}(|W_2(\varphi_{\lambda})|^{2p})$ :

$$\begin{aligned}
\mathbb{E}(|W_2(\varphi_{\lambda})|^{2p}) &\leq K_2 \left[ \frac{n^{2p}}{T^p} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^p \left( \int_{\mathbb{R}} h(x) dx \right)^{2p} \right. \\
&\quad \left. + \frac{n^p}{T^p} \left( \int_{\mathbb{R}} \varphi_{\lambda}^2(x) dx \right)^p \left( \int_{\mathbb{R}} h(x) dx \right)^{2p} + \left( \int_{\mathbb{R}} |\varphi_{\lambda}(x)| dx \right)^{2p} \|h\|_{\infty}^{2p} \right],
\end{aligned} \quad (3.53)$$

with  $K_2$  a positive constant only depending on  $p$ .

Finally, by using inequalities (3.47), (3.48) and (3.53) in (3.35), we obtain:

$$\begin{aligned} & \mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) \\ & \leq K \left\{ \frac{1}{n^{2p-2}T^{2p}} \|\varphi_\lambda\|_1^{2p} \|h\|_1 + \frac{1}{n^{p-1}T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1 + \frac{1}{n^{2p-2}T} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_1 \right. \\ & \quad + \frac{1}{n^{2p-1}} \|\varphi_\lambda\|_2^2 \|\varphi_\lambda\|_\infty^{2p-2} \|h\|_\infty + \frac{1}{n^{p/2}T^p} \|\varphi_\lambda\|_1^{2p} \|h\|_\infty^p + \frac{1}{T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1^p \\ & \quad + \frac{1}{n^p} \|\varphi_\lambda\|_2^{2p} \|h\|_\infty^p + \frac{1}{n^{p/2}T} \|\varphi_\lambda\|_2^{2p} \|h\|_\infty^p + \frac{n}{T^{2p}} \|\varphi_\lambda\|_1^{2p} \|h\|_1^{2p} + \frac{n^p}{T^{2p}} \|\varphi_\lambda\|_1^{2p} \|h\|_1^{2p} \\ & \quad \left. + \frac{1}{T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1^{2p} + \frac{1}{n^p T^p} \|\varphi_\lambda\|_2^{2p} \|h\|_1^{2p} + \frac{1}{n^{2p}} \|\varphi_\lambda\|_1^{2p} \|h\|_\infty^{2p} \right\} \end{aligned}$$

and so,

$$\begin{aligned} & \left[ \mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) \right]^{\frac{1}{p}} \\ & \leq K \left\{ \frac{1}{n^{2-2/p}T^2} \|\varphi_\lambda\|_1^2 \|h\|_1^{1/p} + \frac{n}{T^2} \|\varphi_\lambda\|_1^2 \|h\|_1^2 + \frac{1}{n^{1/2}T} \|\varphi_\lambda\|_1^2 \|h\|_\infty + \frac{1}{n^2} \|\varphi_\lambda\|_1^2 \|h\|_\infty^2 \right. \\ & \quad + \frac{1}{n^{1-1/p}T} \|\varphi_\lambda\|_2^2 \|h\|_1^{1/p} + \frac{1}{T} \|\varphi_\lambda\|_2^2 \|h\|_1 + \frac{1}{T} \|\varphi_\lambda\|_2^2 \|h\|_1^2 + \left[ \frac{1}{n} + \frac{1}{n^{1/2}T^{1/p}} \right] \|\varphi_\lambda\|_2^2 \|h\|_\infty \\ & \quad \left. + \left[ \frac{1}{n^{2-2/p}T^{1/p}} + \frac{1}{n^{2-1/p}} \right] \|\varphi_\lambda\|_2^{2/p} \|\varphi_\lambda\|_\infty^{2-2/p} \|h\|_\infty^{1/p} \right\}, \end{aligned}$$

with  $K$  a positive constant only depending on  $p$  and the compact support of  $\psi$ .

Recall that for any  $\lambda = (j, k) \in \Lambda$ , we have:

$$\|\varphi_\lambda\|_1 \leq 2^{-j/2} M_{\psi,1}, \quad \|\varphi_\lambda\|_2 \leq M_{\psi,2} \quad \text{and} \quad \|\varphi_\lambda\|_\infty \leq 2^{j/2} M_{\psi,\infty}$$

We consider  $1 < p < \infty$  and we fix  $1 < q < \infty$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ , so that

$$\begin{aligned} \left[ \mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) \right]^{\frac{1}{p}} & \leq K \left\{ \frac{n}{T^2} \|h\|_1^2 + \frac{1}{n^2} \|h\|_\infty^2 + \frac{1}{n^{1/q}T} \|h\|_1^{1/p} + \frac{1}{T} \|h\|_1 + \frac{1}{T} \|h\|_1^2 \right. \\ & \quad \left. + \left[ \frac{1}{n} + \frac{1}{n^{1/2}T^{1/p}} \right] \|h\|_\infty + \left[ \frac{1}{n^{2/q}T^{1/p}} + \frac{1}{n^{1+1/q}} \right] 2^{j_0/q} \|h\|_\infty^{1/p} \right\}, \end{aligned}$$

with  $K$  a positive constant depending on  $p$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$ ,  $\|\psi\|_\infty$  and the compact support of  $\psi$ .

Finally, choosing  $p = 2$ , Assumption (A2) is fulfilled with

$$R = C_R \left\{ \frac{1}{n} + \frac{2^{j_0/2}}{n^{3/2}} + \frac{2^{j_0/2}}{nT^{1/2}} + \frac{n}{T^2} \right\},$$

where  $C_R$  is a positive constant depending on  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$ ,  $\|\psi\|_\infty$  and the compact support of  $\psi$ ,

$$H_\lambda = \mathbf{1}_{\lambda \in \Gamma}$$

and  $\varepsilon = 1$ .

### 3.A.4 Proof of Assumption (A3) of Theorem 3.3

To shorten mathematical expressions, we denote  $\eta_\lambda = \eta_\lambda(\gamma, \Delta)$  in the sequel. The following inequality:

$$\mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa\eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda\right) \leq H_\lambda \zeta$$

is obvious with  $\zeta = \omega$ , which proves Assumption (A3) choosing  $\theta = \frac{1+\varepsilon}{\varepsilon}$ .

### 3.A.5 Completion of the proof of Theorem 3.2

Therefore we can apply [91]: the estimator  $\tilde{\beta} = \left(\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma}\right)_{\lambda \in \Lambda}$  satisfies

$$\begin{aligned} & \frac{1 - \kappa^2}{1 + \kappa^2} \mathbb{E}\left(\|\tilde{\beta} - \beta\|_{\ell_2}^2\right) \\ & \leq \mathbb{E}\left(\inf_{m \subset \Gamma} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda \in m} \eta_\lambda^2 \right\}\right) + LD \sum_{\lambda \in \Gamma} H_\lambda \\ & \leq \inf_{m \subset \Gamma} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} \mathbb{E}((\hat{\beta}_\lambda - \beta_\lambda)^2) + \sum_{\lambda \in m} \mathbb{E}(\eta_\lambda^2) \right\} + LD \sum_{\lambda \in \Gamma} H_\lambda, \end{aligned}$$

with

- for all  $\lambda = (j, k)$  in  $\Gamma$ ,

$$\mathbb{E}((\hat{\beta}_\lambda - \beta_\lambda)^2) = \text{Var}(\hat{\beta}_\lambda) \leq K \left\{ \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2} \right\},$$

where  $K$  is a positive constant depending on  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$  and  $\|\psi\|_2$  (see Proposition 1);

- for all  $\lambda = (j, k)$  in  $\Gamma$ ,

$$\eta_\lambda \leq K \left( \sqrt{j_0 \tilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + j_0 B\left(\frac{\varphi_\lambda}{n}\right) + \tilde{\Delta} \frac{N_{\mathbb{R}}}{n} \right)$$

where  $K$  depends on  $\varepsilon$ ,  $\kappa$ ,  $\gamma$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  and

$$\tilde{\Delta} = \frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T};$$

- $LD = \frac{R}{\kappa^2} ((1 + \theta^{-1/2})\omega^{1/2} + (1 + \theta^{1/2})\varepsilon^{1/2}\zeta^{1/2}) \leq K R(e^{-\kappa^2 j_0 \gamma/2} + \exp(-g(\varepsilon_0)n\|h\|_1/2))$ , where  $K$  is a positive constant depending only on  $\varepsilon$  and  $\kappa$ ,
- $\sum_{\lambda \in \Gamma} H_\lambda = |\Gamma|$ , where  $|\Gamma|$  is the cardinal of the set  $\Gamma$ . So, we can upper bound this quantity by  $K 2^{j_0}$ , where  $K$  is a positive constant depending only on the compact support of  $h$  and the compact support of  $\psi$ .

Recall that  $\varepsilon = 1$ ,  $\kappa \in ]0; 1[$  will be fixed in the sequel and  $\gamma > 0$ , according Assumption (A1).

It remains to compute  $\mathbb{E}(\eta_\lambda^2)$ . Let  $\lambda \in \Gamma$ . We have:

$$\mathbb{E}(\eta_\lambda^2) \leq K \left( j_0 \mathbb{E}\left(\tilde{V}\left(\frac{\varphi_\lambda}{n}\right)\right) + j_0^2 \mathbb{E}\left(B^2\left(\frac{\varphi_\lambda}{n}\right)\right) + \left\{ \frac{j_0 n}{T^2} + \frac{j_0^2}{T} + \frac{2^{j_0} j_0^4}{n^2} \right\} \frac{\mathbb{E}(N_{\mathbb{R}}^2)}{n^2} \right),$$

with  $K$  depending on  $\varepsilon$ ,  $\kappa$ ,  $\gamma$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$  and  $\mathbb{E}(N_{\mathbb{R}}^2) = n\|h\|_1 + n^2\|h\|_1^2 \leq 2n^2M_{h,1}^2$ . We control  $\tilde{V}(\frac{\varphi_\lambda}{n})$  in expectation and we recall that  $\alpha = \kappa^2 j_0 \gamma$ .

$$\mathbb{E}(\tilde{V}(\frac{\varphi_\lambda}{n})) \leq \mathbb{E}(\hat{V}(\frac{\varphi_\lambda}{n})) + \sqrt{2\alpha\mathbb{E}(\hat{V}(\frac{\varphi_\lambda}{n}))\mathbb{E}(B^2(\frac{\varphi_\lambda}{n}))} + 3\alpha\mathbb{E}(B^2(\frac{\varphi_\lambda}{n})), \quad (3.54)$$

with using inequality (6.7),

$$\mathbb{E}(\hat{V}(\frac{\varphi_\lambda}{n})) = \mathbb{E}(V(\frac{\varphi_\lambda}{n})) \leq K \left\{ \frac{1}{n}\|h\|_\infty + \frac{1}{T}\|h\|_1 \right\},$$

where  $K$  is a positive constant depending only on  $\|\psi\|_2$ .

Now, we focus on  $\mathbb{E}(B^2(\frac{\varphi_\lambda}{n}))$  where

$$\begin{aligned} B(\frac{\varphi_\lambda}{n}) &= \frac{1}{n} \left\| \sum_{i=1}^n \left[ \varphi_\lambda(\cdot - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(\cdot - U)) \right] \right\|_\infty \\ &= \frac{1}{n} \sup_{t \in \mathbb{R}} \left| \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right| \\ &\leq \tilde{B}(\frac{\varphi_\lambda}{n}) + \frac{1}{nT} \|\varphi_\lambda\|_1, \end{aligned}$$

$$\text{with } \tilde{B}(\frac{\varphi_\lambda}{n}) = \frac{1}{n} \sup_{t \in \mathbb{R}} \left| \sum_{i=1}^n [\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))] \right|.$$

Then, we have to control  $\mathbb{E}(\tilde{B}(\frac{\varphi_\lambda}{n}))$ . Since it is a decomposition biorthogonal wavelet,  $\varphi_\lambda$  is a piecewise constant function and we can write:

$$\varphi_\lambda = 2^{j/2} \sum_{l=1}^N c_l \mathbf{1}_{[a_l 2^{-j}; b_l 2^{-j}]},$$

where  $N \in \mathbb{N}^*$  and for any  $l \in \{1, \dots, N\}$ ,  $a_l, b_l, c_l \in \mathbb{R}$ ,  $a_l < b_l$  and  $c_l$  an absolute constant. It is easy to see that

$$\tilde{B}(\frac{\varphi_\lambda}{n}) \leq 2^{j/2} \sum_{l=1}^N \tilde{B}(c_l \mathbf{1}_{[a_l 2^{-j}; b_l 2^{-j}]}/n) = 2^{j/2} \sum_{l=1}^N |c_l| \tilde{B}(\mathbf{1}_{[a_l 2^{-j}; b_l 2^{-j}]}/n).$$

It remains to compute  $\mathbb{E}(\tilde{B}(\mathbf{1}_{[a 2^{-j}; b 2^{-j}]}/n))$  for some interval  $[a; b]$ .

$$\begin{aligned} \tilde{B}(\mathbf{1}_{[a 2^{-j}; b 2^{-j}]}/n) &= \frac{1}{n} \sup_{t \in \mathbb{R}} \left| \sum_{i=1}^n [\mathbf{1}_{[a 2^{-j}; b 2^{-j}]}(t - U_i) - \mathbb{E}_\pi(\mathbf{1}_{[a 2^{-j}; b 2^{-j}]}(t - U))] \right| \\ &\leq \frac{1}{n} \sup_{B_t, t \in \mathbb{R}} \left| \sum_{i=1}^n [\mathbf{1}_{B_t}(U_i) - \mathbb{E}_\pi(\mathbf{1}_{B_t}(U))] \right|, \end{aligned}$$

where for any  $t \in \mathbb{R}$ ,  $B_t = [t - b 2^{-j}; t - a 2^{-j}]$ . We set  $\mathcal{B} = \{B_t, t \in \mathbb{R}\}$  and for every integer  $n$ ,  $m_n(\mathcal{B}) = \sup_{A \subset \mathbb{R}, |A|=n} |\{A \cap B_t, t \in \mathbb{R}\}|$ . It is easy to see that

$$m_n(\mathcal{B}) \leq 1 + \frac{n(n+1)}{2}$$

and so, the VC-dimension  $\mathcal{V}$  of  $\mathcal{B}$  defined by  $\sup\{n \geq 0, m_n(\mathcal{B}) = 2^n\}$  is bounded by 2 (see Definition 6.2 of [77]). Let us define  $\sigma^2 = \max\{c^2 2^{-j}, \xi^2 \mathcal{V}(1 + |\log(c2^{-j/2})|)/n\}$  with  $c^2 = b - a$  and  $\xi$  the absolute constant given by Lemma 6.4 of [77]. The quantity  $\sigma^2$  satisfies in particular the two following assertions:

$$\forall B \in \mathcal{B}, \mathbb{P}_\pi[U \in B] \leq \sigma^2 \quad \text{and} \quad \sigma \geq \xi \sqrt{\mathcal{V}(1 + \log(\sigma^{-1} \vee 1))/n}.$$

Indeed, on the one hand, if  $\sigma^2 = c^2 2^{-j}$ , in particular we have:  $\xi^2 \mathcal{V}(1 + \log(\sigma^{-1} \vee 1))/n \leq \xi^2 \mathcal{V}(1 + |\log(c2^{-j/2})|)/n \leq \sigma^2$ , and on the other hand, if  $\sigma^2 = \xi^2 \mathcal{V}(1 + |\log(c2^{-j/2})|)/n$ , we have:

$$\xi^2 \mathcal{V}(1 + \log(\sigma^{-1} \vee 1))/n \leq \xi^2 \mathcal{V}(1 + |\log(c2^{-j/2})|)/n = \sigma^2.$$

By applying Lemma 6.4 of [77], we obtain:

$$\begin{aligned} & \sqrt{n} \mathbb{E}(\tilde{B}(\mathbf{1}_{[a2^{-j}; b2^{-j}]}/n)) \\ & \leq \frac{\xi}{2} \sigma \sqrt{\mathcal{V}(1 + |\log \sigma|)} \\ & \leq \frac{\xi}{2} c 2^{-j/2} \sqrt{\mathcal{V}(1 + |\log(c2^{-j/2})|)} \\ & \quad + \frac{\xi}{2} \xi \sqrt{\mathcal{V}(1 + |\log(c2^{-j/2})|)/n} \sqrt{\mathcal{V}(1 + |\log(\xi \sqrt{\mathcal{V}(1 + |\log(c2^{-j/2})|)/n})|)} \\ & \leq K \left\{ 2^{-j/2} \sqrt{j} + \frac{j}{\sqrt{n}} \right\}, \end{aligned}$$

with  $K$  an absolute positive constant. Consequently, for any  $\lambda$  in  $\Gamma$ ,

$$\mathbb{E}(\tilde{B}(\varphi_\lambda/n)) \leq K \left\{ \frac{\sqrt{j}}{\sqrt{n}} + 2^{j/2} \frac{j}{n} \right\}.$$

But, we want an upper bound of  $\mathbb{E}(\tilde{B}^2(\varphi_\lambda/n))$ . For this, we use Theorem 11 of [12]:

$$[\mathbb{E}(\tilde{B}^2(\varphi_\lambda/n))]^{1/2} \leq K \left\{ \mathbb{E}(\tilde{B}(\varphi_\lambda/n)) + (\mathbb{E}[\mathcal{M}^2])^{1/2} \right\},$$

where

$$\mathcal{M} = \frac{1}{n} \max_{1 \leq i \leq n} \sup_{t \in \mathbb{R}} |\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U))|.$$

Hence,

$$\mathbb{E}[\mathcal{M}^2] \leq \frac{4}{n^2} \|\varphi_\lambda\|_\infty^2 \leq \frac{K}{n^2} 2^j,$$

with  $K$  a constant only depending on  $\|\psi\|_\infty$ . Finally,

$$\begin{aligned} \mathbb{E}(B^2(\varphi_\lambda/n)) & \leq K \left\{ \mathbb{E}(\tilde{B}^2(\varphi_\lambda/n)) + \frac{2^{-j}}{n^2 T^2} \right\} \\ & \leq K \left\{ [\mathbb{E}(\tilde{B}(\varphi_\lambda/n))]^2 + \frac{2^j}{n^2} + \frac{2^{-j}}{n^2 T^2} \right\} \\ & \leq K \left\{ \frac{j}{n} + 2^j \frac{j^2}{n^2} + \frac{2^j}{n^2} + \frac{2^{-j}}{n^2 T^2} \right\}, \end{aligned} \tag{3.55}$$

with  $K$  a constant only depending on  $\|\psi\|_1$  and  $\|\psi\|_\infty$ .



Then combining (3.54) and (3.55) yields

$$\begin{aligned} \mathbb{E}(n_\lambda^2) &\leq K \left( \frac{j_0}{n} + \frac{j_0}{T} + \frac{j_0^3}{n} + 2^{j_0} \frac{j_0^4}{n^2} + \frac{j_0^2 2^{j_0}}{n^2} + \frac{j_0^2}{n^2 T^2} + \frac{j_0 n}{T^2} + \frac{j_0^2}{T} + \frac{2^{j_0} j_0^4}{n^2} \right) \\ &\leq K \left( \frac{j_0^3}{n} + \frac{j_0^2}{T} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{n^2 T^2} + \frac{j_0 n}{T^2} \right) \end{aligned}$$

where  $K$  is a constant depending on  $\gamma$ ,  $\|h\|_1$ ,  $\|h\|_\infty$ ,  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$ , which concludes the proof of Theorem 3.2 by setting

$$F(j_0, n, T) = \frac{j_0^3}{n} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0 n}{T^2} + \frac{j_0^2}{n^2 T^2}.$$

# Chapitre 4

## Detection of dependence in a Poissonian interactions model

This chapter proposes a model of interactions between two point processes, ruled by a reproduction function  $h$ , which is considered as the intensity of a Poisson process. In particular, we focus on the context of neurosciences to detect possible interactions in the cerebral activity associated with two neurons. To provide a mathematical answer to this specific problem of neurobiologists, we address so the question of testing the nullity of the intensity  $h$ . We construct a multiple testing procedure obtained by the aggregation of single tests based on a wavelet thresholding method. This test has good theoretical properties: it is possible to guarantee the level but also the power under some assumptions and its uniform separation rate over weak Besov bodies is adaptive minimax. Then, some simulations are provided showing the good practical behavior of our testing procedure.

*This chapter is the fruit of a collaboration with Christine Tuleau-Malot (Université Nice Sophia-Antipolis) and constitutes the paper A model of Poissonian interactions and detection of dependence. Submitted, 2013.*

### Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>98</b>
4.1.1	Context of neurosciences	98
4.1.2	Description of our model	98
4.1.3	Detection of dependence and tests in the Poisson setting	100
4.1.4	The general results	101
4.1.5	Overview of the chapter and notations	101
<b>4.2</b>	<b>Description of our testing procedure</b>	<b>102</b>
4.2.1	The single testing procedures	103
4.2.2	The multiple testing procedure	103
<b>4.3</b>	<b>Main theoretical results</b>	<b>104</b>
4.3.1	Probability of first kind error	104
4.3.2	Probability of second kind error	104
4.3.3	Uniform separation rates	106

<b>4.4</b>	<b>Simulation study</b>	<b>107</b>
4.4.1	Description of the data	107
4.4.2	The Kolmogorov-Smirnov test	108
4.4.3	The GAUE method adapted to our context	109
4.4.4	Our procedure in practice	109
4.4.5	Results	110
<b>4.5</b>	<b>Conclusion</b>	<b>111</b>
<b>4.6</b>	<b>Proofs</b>	<b>111</b>
4.6.1	Proof of Proposition 4.1	112
4.6.2	Proof of Proposition 4.2	113
4.6.3	Proof of Theorem 4.1	113
4.6.4	Proof of Theorem 4.2	115
4.6.5	Proof of Theorem 4.3	116
4.6.6	Proof of lemmas	117

## 4.1 Introduction

### 4.1.1 Context of neurosciences

In neurosciences, an important issue lies in a better understanding of the dynamics of cerebral activity in the cortex. In practice it is possible to measure, in vivo and for a specific task, the cerebral activity through the emission of action potentials by several neurons, and the specific interest of the neurobiologists is to understand how these action potentials appear. During a task, the recording of all arrival times of these action potentials (or spikes) on a neuron forms a spike train. From this point of view, the spike train can be modeled by a point process.

Several years ago it was thought that activities of different neurons during a task were independent (for example, see Barlow [6]); this explains why in the studies, the spike trains were usually modeled by independent Poisson processes. Today, thanks to technological advances in terms of recording brain activity, various studies show that this belief is false (for instance, see Gerstein [47] and Lestienne [73]). Thus the recent studies consider neuronal assemblies instead of the separate neuronal activities. For example activities of pairs of neurons, that have been recorded simultaneously, show that there exists a phenomenon called synchronization (for example, see Grammont and Riehle [50] and Grün *et al.* [52]): the presence of a spike on one of the two spike trains can affect the apparition of a spike, with a delay, on the second spike train. From a biological point of view, such a phenomenon reflects a reality. Indeed, an action potential appears if the neuron is sufficiently excited. However, to obtain a sufficient excitation, two strategies exist: either the frequency of spikes received by a single neuron increases, or the receiving neuron receives less spikes but at the same time from different neurons. This second strategy is precisely the synchronization. From a biological point of view, it is less energy consuming and the reaction is faster. Therefore, the neurobiologists are interested in detecting the synchronization phenomenon. More generally, they want to detect whether or not neurons evolve independently of each other, a dependence being a hint of a functional connection during a task.

### 4.1.2 Description of our model

To mathematically answer this question, we need a model taking into account the possible interactions between two neurons. In neurosciences, a possible model is the Hawkes process (for

instance, see [59] for theoretical aspects and [13, 69, 84, 85] for its introduction in neurosciences). However, the Hawkes process is, theoretically speaking, a very complicated model, thus we consider a simpler version of Hawkes process which is realistic for the possible applications (in neurosciences, or in genomics, ...) and for which it is possible to carry out computations. Our model is the following one. Let  $N_p$  and  $N_c$  be two point processes with respective conditional intensity

$$\tilde{\lambda}_p : t \mapsto \mu_p \quad \text{and} \quad \tilde{\lambda}_c : t \mapsto \mu_c + \int_{-\infty}^t h(t-u) dN_p(u), \quad (4.1)$$

where  $\mu_p$  and  $\mu_c$  are positive parameters describing the spontaneous part (in the context of neurosciences, the spontaneous apparition of spikes) and  $h$  is a function which reflects the influence of  $N_p$  on  $N_c$ . In this model, we have to assume that  $\text{supp}(h) \subset \mathbb{R}_+^*$ , where  $\text{supp}(h)$  is the support of the function  $h$ . Moreover,  $N_p$  is a homogeneous Poisson process (for example, see [67]) and  $N_c$  is a special case of Hawkes process. The biological problem which consists in knowing whether  $N_p$  influences  $N_c$  is equivalent to test the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " against the alternative  $\mathcal{H}_1$ : " $h \neq 0$ ".

The above formulation of  $\tilde{\lambda}_c$  is an integral form. However it is possible conditionally on the points of  $N_p$  to have a vision in terms of descendants and no more in terms of intensity conditionally on the only past observations. Indeed, given  $T$  a positive real number representing the time of record of the neuronal activity and given  $n$  an integer, conditionally on the event "the number of points of  $N_p$  lying in  $[0; T]$  is  $n$ ", the points of the process  $N_p$  obey the same law as a  $n$ -sample of uniform random variables on  $[0; T]$ , denoted  $U_1, \dots, U_n$  and named parents. Thus, conditionally on  $U_1, \dots, U_n$ , we can write  $\tilde{\lambda}_c(t) = \mu_c + \sum_{i=1}^n h(t - U_i)$ . This new expression of  $\tilde{\lambda}_c$  can be interpreted as follows. Each  $U_i$  gives birth independently to a Poisson process  $N_c^i$  with intensity the function  $t \mapsto h(t - U_i)$  with respect to the Lebesgue measure on  $\mathbb{R}$ , to which is added a homogeneous Poisson process  $N_c^0$  with constant intensity  $\mu_c$ , representing the orphans. We consequently consider the aggregated process

$$N_c = \sum_{i=0}^n N_c^i \quad \text{whose intensity is given by the function} \quad t \mapsto \mu_c + \sum_{i=1}^n h(t - U_i) \quad (4.2)$$

and the points of the process  $N_c$  are named children. With this interpretation, the goal of the present chapter is to test the "influence or not" of the parents on the children, via the reproduction function  $h$ . This second writing contains many benefits. First, the assumption  $\text{supp}(h) \subset \mathbb{R}_+^*$  is not mandatory. With respect to the first formulation, this may appear like a minor difference, but in practice the impact is considerable. Indeed, if we refer to the context of neurosciences, assuming that the support of  $h$  is in  $\mathbb{R}_+$  means that one favors a sense of interactions, namely  $N_p$  affects  $N_c$ . However in practice, we do not have this information a priori. Therefore, when the test does not reject  $\mathcal{H}_0$ , it means that  $N_p$  does not seem to influence  $N_c$ , but this may be because in reality this is  $N_c$  that affects  $N_p$ . We must be careful that the initially proposed model is not symmetric in terms of neurons and that a support in  $\mathbb{R}_+$  does not really allow to answer the question of dependence. The causality is indeed represented by the fact that a child appears after its parent and therefore  $h$  has to be supported in  $\mathbb{R}_+$ . Heuristically, a consequence is the following interpretation: if a parent has a child before its own birth, it may represent that the child is the parent and the parent the real child. Looking at both sides of the support makes the procedure in some sense adaptive to the causality of parent/child roles. Another advantage of this second writing is that it allows applications to other disciplines such as genomics where one studies for

instance the favored or avoided distances between patterns on a strand of DNA and where it is not always possible to know which pattern rules the other. More details about this application to genomics can be found in Chapter 3, where the author proposes an estimation procedure of the function  $h$ , assumed to be well localized, based on wavelet thresholding methods, in a very similar model to the one studied here. The interested reader will find other estimation procedures of the function  $h$  in this DNA context, by using a Hawkes' model in Gusto and Schbath [54] and Reynaud-Bouret and Schbath [93]. In this chapter, we consider the model defined by (4.2). For the simulation study, parents process  $(U_i)_i$  is simulated according to a homogenous Poisson process of intensity  $\mu_p$ .

### 4.1.3 Detection of dependence and tests in the Poisson setting

Since the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " means that conditionally on the total number of points of  $N_c$ , the points of the process  $N_c$  are i.i.d. (independent and identically distributed) with uniform distribution, a first rather naive approach is to perform a Kolmogorov-Smirnov test (for example, see [30]). But this test is not powerful, as illustrated in the section devoted to simulations. The aim of this chapter is then to build a more powerful and nonparametric test  $\Phi_\alpha$  with values in  $\{0, 1\}$  of  $\mathcal{H}_0$ : " $h = 0$ " against the alternative  $\mathcal{H}_1$ : " $h \neq 0$ ", rejecting  $\mathcal{H}_0$  when  $\Phi_\alpha = 1$ , with prescribed probabilities of first and second kind errors. The performance of the test  $\Phi_\alpha$  is measured by its uniform separation rate (for instance, see [4]).

In neurosciences, parametric methods exist to detect such dependence. For instance, the Unitary Event (UE) (see [52]) and the Multiple Tests based on a Gaussian Approximation of the UE (MTGAUE) (see [109]) methods answer partially the problem by considering coincidences (see Section 4.4 for more details). In the one-sample Poisson process model (that is to say  $n = 1$  and  $\mu_c = 0$  in our model), many papers deal with different problems of testing the simple hypothesis that an observed point process is a Poisson process with a known intensity. We can cite for example the papers of Fazli and Kutoyants [44] where the alternative is also a Poisson process with a known intensity, Fazli [43] where the alternatives are Poisson processes with one-sided parametric intensities or Dachian and Kutoyants [28] where the alternatives are self-exciting point processes (namely, Hawkes processes). In the nonparametric framework, Ingster and Kutoyants [63] propose a goodness-of-fit test where the alternatives are Poisson processes with nonparametric intensities in a Sobolev  $\mathcal{S}_2^\delta(R)$  or a Besov ball  $\mathcal{B}_{2,q}^\delta(R)$  with  $1 \leq q < \infty$  and known smoothness parameter  $\delta$ . They establish its uniform separation rate over a Sobolev or a Besov ball and show the adaptivity of their testing procedure in a minimax sense.

In some practical cases like the study of the expression of neuronal interactions or the study of favored or avoided distances between patterns on a strand of DNA, such smooth alternatives (Sobolev or Besov balls) cannot be considered. Indeed, the intensity of the Poisson process  $N_c$  in these cases may burst at a particular position of special interest for the neuroscientist or the biologist. So we have to develop a testing procedure able to distinguish a constant function (or here a null function) from a function that has some small localized spikes. These features are not well captured by using classical Besov spaces. Hence we focus in particular on alternatives based on sparsity rather than on alternatives based on smoothness. For this, we are interested in the computation of uniform separation rates over weak versions of Besov balls. Such alternatives have already been considered. For instance, Fromont *et al.* [45] propose non-asymptotic and nonparametric tests of the homogeneity of a Poisson process that are adaptive over various Besov

bodies simultaneously and in particular over weak Besov bodies. Another example is Fromont *et al.* [46] which construct non-asymptotic and nonparametric multiple tests of the equality of the intensities of two independent Poisson processes, that are adaptive in the minimax sense over a large variety of classes of alternatives based on classical and weak Besov bodies in particular.

#### 4.1.4 The general results

The test  $\Phi_\alpha$  proposed in this chapter consists in a multiple testing procedure obtained by aggregating several single tests based on a wavelet thresholding method as in Fromont *et al.* [45, 46] (they also consider model selection and kernel estimation methods). First, Proposition 4.2 proves that the multiple test is an  $\alpha$ -level test and Theorem 4.2 gives a condition on the alternative to ensure that our multiple test has a prescribed second kind error. This result reveals two regimes as in Chapter 3. Indeed our model presents a double asymptotic through the number  $n$  of parents and the recording time  $T$  (namely, the length of the observations interval), which is not usual. Since  $N_p$  is a homogeneous Poisson process with constant intensity  $\mu_p$ , the number  $n$  of points of  $N_p$  falling into  $[0; T]$  is the realization of a Poisson random variable with parameter  $\mu_p T$ . As a consequence with very high probability,  $T$  is proportional to  $n$  and in this case, the uniform separation rates of the multiple test over weak Besov bodies are established by Theorem 4.3. Thus, our testing procedure is near adaptive in the minimax sense over a class of such alternatives. The proofs of these results are essentially based on concentration inequalities (see [77]) and on exponential inequalities for  $U$ -statistics (see [61]). Secondly, some simulations are carried out to validate our procedure, which is compared with the classical Kolmogorov-Smirnov test and a testing procedure proposed by Tuleau-Malot *et al.* [109], which formalized a well-known procedure in neurosciences, namely the UE method (see Grün *et al.* [52]).

#### 4.1.5 Overview of the chapter and notations

The chapter is organized as follows. Section 4.2 deals with the description of our testing procedure. Section 4.3 is devoted to the general results of the chapter. The control of the probability of second kind error is ensured by Theorem 4.1 for the single testing procedures and by Theorem 4.2 for the multiple test. The uniform separation rates of the multiple test over weak Besov bodies are provided in Theorem 4.3. Section 4.4 presents the simulation study. The proofs of our main theoretical results are finally postponed in Section 4.6.

To end this section we introduce some notations that will be used along the chapter. We denote by  $dN_p$  and  $dN_c$  the point measures associated with  $N_p$  and  $N_c$  respectively. We denote by  $\mathbb{P}_0$  the distribution of the aggregated process  $N_c$  under  $\mathcal{H}_0$ ,  $\mathbb{P}_h$  the distribution of  $N_c$  whose intensity conditionally on  $U_1, \dots, U_n$  is given by the function  $t \mapsto \mu_c + \sum_{i=1}^n h(t - U_i)$  for any alternative  $h$  and by  $\mathbb{E}_h$  the corresponding expectation. The uniform distribution on  $[0; T]$  is named  $\pi$  and  $\mathbb{E}_\pi(f(U))$  denotes the expectation of  $f(U)$  where  $U \sim \pi$  (an independent copy of  $U_1, \dots, U_n$ ) for any measurable function  $f$ . For an orthonormal basis  $\{\varphi_\lambda, \lambda \in L\}$  of a finite dimensional subspace  $S_L$  of  $L_2(\mathbb{R})$ , we denote by  $D_L$  the dimension of  $S_L$  (namely the cardinal of  $L$ ) and by  $h_L$  the orthogonal projection of  $h$  onto  $S_L$ .

## 4.2 Description of our testing procedure

In the sequel, we assume that  $h$  is compactly supported (there is a maximal time of synchronization during a task according to the neuroscientists). Without loss of generality, we suppose now that the support of  $h$  is strictly included in  $[-1; 1]$  and that we observe the  $U_i$ 's (the parents) on  $[0; T]$  and realizations of the process  $N_c$  (the children) on  $[-1; T+1]$ . Consequently,  $h$  belongs to  $\mathbb{L}_1(\mathbb{R})$ ,  $\mathbb{L}_2(\mathbb{R})$  and  $\mathbb{L}_\infty(\mathbb{R})$ . Then, we can consider the decomposition of  $h$  on the Haar basis denoted by  $\{\varphi_\lambda, \lambda \in \Lambda\}$ :

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda \quad \text{with} \quad \beta_\lambda = \int_{\mathbb{R}} h(x) \varphi_\lambda(x) dx,$$

where

$$\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$$

and for any  $\lambda \in \Lambda$  and any  $x \in \mathbb{R}$ ,

$$\varphi_\lambda(x) = \begin{cases} \phi(x - k) & \text{if } \lambda = (-1, k) \\ 2^{j/2} \psi(2^j x - k) & \text{if } \lambda = (j, k) \text{ with } j \geq 0 \end{cases},$$

with

$$\phi = \mathbf{1}_{[0;1]} \quad \text{and} \quad \psi = \mathbf{1}_{] \frac{1}{2}; 1]} - \mathbf{1}_{[0; \frac{1}{2}]}$$

The functions  $\phi$  and  $\psi$  are respectively the father and the mother wavelets. Since the goal is to detect a signal and not to reconstruct it, the Haar basis is suitable in our context. Furthermore from a practical point of view, the use of the Haar basis yields fast algorithms, easy to implement. Nevertheless the theoretical results of the present chapter can be generalized to a biorthogonal wavelet basis (see [23] for a definition of this particular basis) as in Chapter 3 and in [91, 92]. We precise that we can easily extend our results to a function  $h$  compactly supported in  $[-A; A]$  for any  $A > 0$  by scaling the data by  $[A] + 1$ .

By considering this wavelet decomposition of  $h$ , the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " means that all the coefficients  $\beta_\lambda$  are null and the alternative hypothesis  $\mathcal{H}_1$ : " $h \neq 0$ " means that there exists at least one non-zero coefficient. Since  $h$  is strictly supported in  $[-1; 1]$ , if one coefficient  $\beta_{(-1,k)}$  is non-zero, then there exists at least one coefficient  $\beta_{(j,k)}$  with  $j \geq 0$  which is also non-zero. Therefore, we focus only on the coefficients  $\beta_{(j,k)}$  with  $j \geq 0$  and we introduce the following subset  $\Gamma$  of  $\Lambda$

$$\Gamma = \{\lambda = (j, k) \in \Lambda : j \geq 0, k \in \mathcal{K}_j\}$$

with  $\mathcal{K}_j = \{k \in \mathbb{Z} : -2^j \leq k \leq 2^j - 1\}$  ( $\mathcal{K}_j$  is the set of integers  $k$  such that the intersection of the support of  $\varphi_\lambda$  and  $[-1; 1]$  is not empty, with  $\lambda = (j, k)$ ).

For every  $\lambda$  in  $\Gamma$ , the coefficient  $\beta_\lambda$  is estimated by

$$\hat{\beta}_\lambda = \frac{\mathcal{G}(\varphi_\lambda)}{n}, \quad \text{with} \quad \mathcal{G}(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x - U)) \right] dN_c(x).$$

These estimates, inspired by those proposed in Chapter 3 for a simpler model, namely with  $\mu_c = 0$ , are unbiased:

**Proposition 4.1.** *For all  $\lambda = (j, k)$  in  $\Gamma$ ,  $\hat{\beta}_\lambda$  is an unbiased estimator of  $\beta_\lambda$ .*

The proof of Proposition 4.1 uses the fact that for all  $\lambda$  in  $\Gamma$ ,  $\int_{-1}^1 \varphi_\lambda(t) dt = 0$ , avoiding boundary effects (see Section 4.6.1).

In order to test the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " against  $\mathcal{H}_1$ : " $h \neq 0$ ", namely " $\exists \lambda \in \Gamma, \beta_\lambda \neq 0$ ", we first propose to test for all  $\lambda \in \Gamma$ , the null hypothesis  $\mathcal{H}_0$  against the alternative  $\mathcal{H}_1^\lambda$ : " $\beta_\lambda \neq 0$ ". For each  $\lambda \in \Gamma$ , the associated simple test actually consists in testing " $\beta_\lambda = 0$ " against " $\beta_\lambda \neq 0$ " or more precisely, in testing the absence of variation of the function  $h$  on a small interval. Then in a second time, we will aggregate these simple tests to test the nullity of  $h$  on its complete support.

### 4.2.1 The single testing procedures

Let us fix some  $\alpha \in ]0; 1[$  and  $\lambda \in \Gamma$ . We want to construct an  $\alpha$ -level test of the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " against  $\mathcal{H}_1^\lambda$ : " $\beta_\lambda \neq 0$ ", from the observation of the parents  $U_1, \dots, U_n$  and the realization of the Poisson process  $N_c$ . We notice first that the null hypothesis entails in particular that  $\beta_\lambda = 0$ .

We introduce the testing statistic  $\hat{T}_\lambda$  defined by

$$\hat{T}_\lambda = |\hat{\beta}_\lambda|.$$

Our single test consists in rejecting the null hypothesis when  $\hat{T}_\lambda$  is too large and more precisely, when

$$\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha),$$

where  $N_{c, tot}$  is the (random) number of points of the process  $N_c$  falling into  $[-1; T + 1]$  and for any  $m \in \mathbb{N}^*$ ,  $q_\lambda^{[U_1, \dots, U_n; m]}(\alpha)$  is the  $(1 - \alpha)$ -quantile conditionally on  $U_1, \dots, U_n$  of

$$\hat{T}_{\lambda, m}^0 = \frac{1}{n} \left| \sum_{k=1}^m \sum_{i=1}^n \left[ \varphi_\lambda(V_k^0 - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(V_k^0 - U)) \right] \right|, \quad (4.3)$$

with  $(V_1^0, \dots, V_m^0)$  a  $m$ -sample with uniform distribution on  $[-1; T + 1]$  (namely a  $m$ -sample of the process  $N_c$  under  $\mathcal{H}_0$ ). We can easily prove that conditionally on  $U_1, \dots, U_n$  and  $N_{c, tot} = m$ ,  $\hat{T}_\lambda$  and  $\hat{T}_{\lambda, m}^0$  have exactly the same distribution under  $\mathcal{H}_0$ . Thus, the corresponding test function is defined by

$$\Phi_{\lambda, \alpha} = \mathbf{1}_{\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha)}. \quad (4.4)$$

### 4.2.2 The multiple testing procedure

Previously, testing procedures have been built based on each single empirical coefficient  $\hat{\beta}_\lambda$ . We propose in this subsection to consider a collection of empirical coefficients instead of a single one, and to define a multiple testing procedure by aggregating the corresponding single tests.

Let  $\{w_\lambda, \lambda \in \Gamma\}$  be a collection of positive numbers such that  $\sum_{\lambda \in \Gamma} e^{-w_\lambda} \leq 1$ . This set allows us to put weights to empirical coefficients according to their index  $\lambda = (j, k) \in \Gamma$ . Given  $\alpha \in ]0; 1[$ , we consider the test which rejects  $\mathcal{H}_0$  when there exists at least one  $\lambda$  in  $\Gamma$  such that

$$\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} e^{-w_\lambda}),$$



where

$$u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} = \sup \left\{ u > 0 : \mathbb{P} \left( \max_{\lambda \in \Gamma} \left( \hat{T}_{\lambda, N_{c, tot}}^0 - q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(ue^{-w_\lambda}) \right) > 0 \mid U_1, \dots, U_n; N_{c, tot} \right) \leq \alpha \right\}. \quad (4.5)$$

The corresponding test function is defined by

$$\Phi_\alpha = \mathbf{1}_{\max_{\lambda \in \Gamma} \left( \hat{T}_{\lambda, N_{c, tot}} - q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} e^{-w_\lambda}) \right) > 0}. \quad (4.6)$$

We mention that, since the set  $\Gamma$  is infinite countable, the number of tests to be performed is infinite and this is not a problem from a theoretical point of view. But in practice, we have to perform a finite number of single tests and so, we will fix a maximal resolution level  $j_0$  and we will carry out the single tests  $\Phi_{\lambda, \alpha}$  for  $\lambda = (j, k)$  in  $\Gamma$  with  $j \leq j_0$ .

In the next section, we study the properties of the single tests  $\Phi_{\lambda, \alpha}$  defined by (4.4) and the multiple test  $\Phi_\alpha$  defined by (4.6), through their probabilities of first and second kind errors.

## 4.3 Main theoretical results

### 4.3.1 Probability of first kind error

We constructed our single and multiple tests in such a way that the first kind error, which measures the probability that the test wrongly rejects the null hypothesis, is less than  $\alpha$ .

**Proposition 4.2.** *Let  $\alpha$  be a fixed level in  $]0; 1[$ . Then the single test  $\Phi_{\lambda, \alpha}$  defined by (4.4) for any  $\lambda \in \Gamma$  and the multiple test  $\Phi_\alpha$  defined by (4.6) are of level  $\alpha$ . Furthermore,  $u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]}$  defined by (4.5) satisfies  $u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} \geq \alpha$ .*

This result shows that the tests are exactly of level  $\alpha$ , which is required for a test from a non-asymptotic point of view (namely  $n$  and  $T$  are not required to tend to infinity).

### 4.3.2 Probability of second kind error

The second kind error, which measures the probability that the test does not wrongly reject the null hypothesis is not fixed by the testing procedure, unlike the first kind error. We have to control the probability of second kind error in such a way that it is close to 0, in order to obtain powerful tests. The following theorem brings out a condition which guarantees that the single tests have a prescribed second kind error.

Since  $h$  belongs to  $\mathbb{L}_1(\mathbb{R})$  and  $\mathbb{L}_\infty(\mathbb{R})$ , we introduce  $R_1$  and  $R_\infty$  two positive real numbers such that  $\|h\|_1 \leq R_1$  and  $\|h\|_\infty \leq R_\infty$ .

**Theorem 4.1.** *Let  $\alpha, \beta$  be fixed levels in  $]0; 1[$ . Let  $\zeta$  and  $\kappa$  be positive constants depending on  $\beta, \mu_c, R_1$  and  $R_\infty$ . For all  $\lambda \in \Gamma$ , let  $\Phi_{\lambda, \alpha}$  be the test function defined by (4.4). Assume that*

$$|\beta_\lambda| > \sqrt{\frac{2\zeta}{\beta} \left( \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2} \right)} + \kappa \left\{ \sqrt{\ln(2/\alpha)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\}, \quad (4.7)$$

for  $\lambda = (j, k)$ . Then,

$$\mathbb{P}_h(\Phi_{\lambda, \alpha} = 0) \leq \beta.$$

Note that the quantity  $\frac{1}{n} + \frac{1}{T} + \frac{2^{-jn}}{T^2}$  that appears under the square root of the first term of the right hand side of (4.7) is of the same order as the upper bound of the variance of the estimates  $\hat{\beta}_\lambda$  (see Proposition 3.1 of Chapter 3). Consequently, the right hand side of (4.7) can be viewed as a standard deviation term, since the other terms are not asymptotically larger than the first term if we assume that  $2^j \leq n^2/(\ln n)^2$ , where asymptotic means  $\min(n, T) \rightarrow +\infty$ .

Theorem 4.1 means that if the coefficient  $\beta_\lambda$  is far enough from 0, then the probability of second kind error is controlled. This result gives a threshold for  $\beta_\lambda$  from which our associated single testing procedure is able to detect a signal and shows that its power is larger than  $1 - \beta$ . Furthermore, if we consider the regime " $T$  proportional to  $n$ " in order to compare our result with known asymptotic rates of testing, Condition (4.7) can be easily obtained for instance if  $\beta_\lambda^2 > C/n$  by assuming that  $2^j \leq n^2/(\ln n)^2$ , with  $C$  a positive constant.

Now we are interested in the power of the multiple testing procedure and the following theorem gives a condition on the alternative in order to ensure that our multiple test has a prescribed second kind error.

**Theorem 4.2.** *Let  $\alpha, \beta$  be fixed levels in  $]0; 1[$ . Let  $\Phi_\alpha$  be the test function defined by (4.6). Assume that there exists at least one finite subset  $L$  of  $\Gamma$  such that*

$$\begin{aligned} \|h_L\|_2^2 &> \left( C_1 D_L + C_2 \sum_{\lambda \in L} w_\lambda \right) \left[ \frac{1}{n} + \frac{n}{T^2} \right] \\ &+ \left( C_3 D_L + C_4 \sum_{\lambda \in L} w_\lambda + C_5 \sum_{\lambda \in L} w_\lambda^2 \right) \left[ \frac{j_L}{n^2} + \frac{j_L^2 2^{j_L}}{n^3} + \frac{1}{n^2 T^2} \right], \end{aligned} \quad (4.8)$$

where  $j_L = \max\{j \geq 0 : (j, k) \in L \text{ with } k \in \mathcal{K}_j\}$  and  $C_1, C_2, C_3, C_4$  and  $C_5$  are positive constants depending on  $\alpha, \beta, \mu_c, R_1$  and  $R_\infty$ . Then,

$$\mathbb{P}_h(\Phi_\alpha = 0) \leq \beta.$$

This theorem means that if there exists one subspace  $S_L$  of  $L_2(\mathbb{R})$  such that  $h_L$  (the orthogonal projection of  $h$  onto  $S_L$ ) lies outside a small ball around 0, then the probability of second kind error is controlled. This result gives a threshold for the energy of  $h_L$  from which our multiple testing procedure is able to detect a signal and shows that its power is larger than  $1 - \beta$ . Furthermore, if we consider the regime " $T$  proportional to  $n$ " in order to compare our result with known asymptotic rates of testing, Condition (4.8) can be easily obtained for instance if  $\|h_L\|_2^2 > C \times (D_L + \sum_{\lambda \in L} w_\lambda + \sum_{\lambda \in L} w_\lambda^2) / n$  by assuming that  $2^{j_L} \leq n^2/(\ln n)^4$ , with  $C$  a positive constant. Then, the separation rate between the null and the alternative hypotheses is of order  $D_L/n$ , and this is typical for testing procedures based on a thresholding approach (see [45, 46] for instance). Usually, nested tests (namely based on model selection) achieve a faster rate of separation of order  $\sqrt{D_L}/n$  (see [4, 5] for instance). But these latter tests are not adaptive over weak Besov bodies. Consequently, the separation rate established by Theorem 4.2 leads to sharp upper bounds for the uniform separation rates over such particular classes of alternatives and so, our multiple testing procedure will be proved to be adaptive over particular classes of alternatives, based on weak Besov bodies.

### 4.3.3 Uniform separation rates

Given some  $\alpha, \beta \in ]0; 1[$ , we have previously built an  $\alpha$ -level test  $\Phi_\alpha$  defined by (4.6), with a probability of second kind error at most equals to  $\beta$  if Condition (4.8) is satisfied. Then, given a class  $\mathcal{S}_\delta$  of alternatives  $h$ , it is natural to measure the performance of the test via its uniform separation rate  $\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta)$  over  $\mathcal{S}_\delta$  (see [4]) defined by

$$\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta) = \inf \left\{ \rho > 0 : \sup_{h \in \mathcal{S}_\delta, \|h\|_2 > \rho} \mathbb{P}_h(\Phi_\alpha = 0) \leq \beta \right\}.$$

In order to compare our result with known asymptotic rates of testing, we consider the regime " $T$  proportional to  $n$ " in this subsection.

We introduce for  $\delta > 0$ ,  $R > 0$  the Besov body

$$\mathcal{B}_{2,\infty}^\delta(R) = \left\{ f \in \mathbb{L}_2(\mathbb{R}) : f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \quad \forall j \geq 0, \sum_{k \in \mathcal{K}_j} \beta_{(j,k)}^2 \leq R^2 2^{-2j\delta} \right\}.$$

We also consider a weaker version of the above Besov bodies defined for  $p > 0$ ,  $R' > 0$  by

$$\mathcal{W}_p^*(R') = \left\{ f \in \mathbb{L}_2(\mathbb{R}) : f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \quad \sup_{s > 0} s^p \sum_{\lambda \in \Gamma} \mathbf{1}_{|\beta_\lambda| > s} \leq R'^p \right\}.$$

Whereas the spaces  $\mathcal{B}_{2,\infty}^\delta(R)$  constitute an ideal class to measure the regularity of the possible alternatives  $h$ , the spaces  $\mathcal{W}_p^*(R')$  constitute an ideal class to measure the sparsity of a wavelet decomposed signal  $h$ . Indeed, if  $f = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda \in \mathcal{W}_p^*(R')$ , then the associated sequence  $\beta = (\beta_\lambda)_{\lambda \in \Gamma}$  satisfies  $\sup_{\ell \in \mathbb{N}^*} \ell^{1/p} |\beta|_{(\ell)} < \infty$ , where the sequence  $(|\beta|_{(\ell)})_\ell$  is the non-increasing rearrangement of  $\beta$ :  $|\beta|_{(1)} \geq |\beta|_{(2)} \geq \dots \geq |\beta|_{(\ell)} \geq \dots$ . This condition gives a polynomial control of the decreasing rate of the sequence  $(|\beta|_{(\ell)})_\ell$  and smaller  $p$ , sparser the signal. There exists an embedding between Besov and weak Besov balls:

$$\mathcal{B}_{2,\infty}^\delta(R) \subset \mathcal{W}_{\frac{2}{1+2\delta}}^*(r),$$

where the radius  $r$  of the weak Besov ball depends on  $\delta$  and  $R$  (more precisely,  $r = 4^\delta R / \sqrt{2^{2\delta} - 1}$ ). See [66, 94, 95] for more details and for extensions in a more general setting. So, we consider in this paper such alternatives based on the intersection of Besov and weak Besov bodies, namely sparse functions with a small regularity, see below.

To evaluate the uniform separation rates, we choose the following collection of weights  $\{w_\lambda, \lambda \in \Gamma\}$  defined by

$$w_\lambda = 2(\ln(j+1) + \ln(\pi/\sqrt{6})) + \ln|\mathcal{K}_j|, \quad (4.9)$$

for any  $\lambda = (j, k) \in \Gamma$ , where  $|\mathcal{K}_j|$  is the cardinal of  $\mathcal{K}_j$  which is of order  $2^j$ . With this choice, the collection of weights satisfies the condition  $\sum_{\lambda \in \Gamma} e^{-w_\lambda} \leq 1$ . The following theorem gives the uniform separation rates over  $\mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R')$ , where the parameter  $\delta$  measures the regularity and the parameter  $\gamma$  the sparsity.

**Theorem 4.3.** *Let  $\alpha, \beta$  be fixed levels in  $]0; 1[$ . Assume that  $T$  is proportional to  $n$ . Let  $\Phi_\alpha$  be the test function defined by (4.6) with the weights  $w_\lambda$ 's defined by (4.9). Then, for any  $\delta > 0$ ,  $\gamma > 0$ ,  $R > 0$ ,  $R' > 0$ , if  $2\delta > \gamma/(1+2\gamma)$*

$$\rho \left( \Phi_\alpha, \mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R'), \beta \right) \leq C \left( \frac{\ln n}{n} \right)^{\frac{\gamma}{1+2\gamma}},$$

with  $C$  a positive constant depending on  $\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1$  and  $R_\infty$ .

If  $\delta > \gamma$ , then the set  $\mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R')$  is reduced to  $\mathcal{B}_{2,\infty}^\delta(R)$  (given the above embedding between Besov and weak Besov balls) that only measures the regularity. Since we are interested in sparse functions (with a small regularity), this is not the purpose here. Then we restrain our interpretation to the case  $\gamma \geq \delta$ . Note that Theorem 4.3 holds for instance with  $\delta = 1/4$  and for all  $\gamma > 0$ . In this case,  $\delta = 1/4$  corresponds to the small regularity mentioned previously. Consequently, the main index  $\gamma$ , the sparsity index, governs the rates of convergence.

Theorem 4.3 illustrates the optimality of our testing procedure in the minimax setting. Indeed, considering the regime " $T$  proportional to  $n$ ", uniform separation rates of the test  $\Phi_\alpha$  match the minimax separation rates obtained by Fromont *et al.* [45, 46], if  $2\delta > \gamma/(1+2\gamma)$  and also  $\delta < \gamma/2$  and  $\gamma > 1/2$  (see Theorem 1 of [45]). Furthermore, the upper bound of uniform separation rates of our test  $\Phi_\alpha$  over  $\mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R')$  has been already obtained, up to a logarithmic term, for a wavelet thresholding estimation method proposed in Chapter 3 in a very similar context and more precisely, this is equal to the minimax estimation rates of the maxisets of the thresholding estimation procedure (see [66, 91, 95] for more details). This means that it is at least as difficult to test as to estimate over such classes of alternatives. Note that on Sobolev or Besov spaces, testing rates are usually faster than estimation rates.

## 4.4 Simulation study

In this section, we study our testing procedure from a practical point of view and we compare it to the conditional Kolmogorov-Smirnov (KS) test and a Gaussian Approximation of the Unitary Events (GAUE) method developed by Tuleau-Malot *et al.* [109].

### 4.4.1 Description of the data

We create different data sets that are to a certain extent a reflection of a neurobiological reality. We consider the spike trains of two neurons  $N_p$  and  $N_c$  which are modeled by two point processes with respective conditional intensity  $\tilde{\lambda}_p$  and  $\tilde{\lambda}_c$  defined by (4.1).

For real spike trains it is not reasonable to postulate the stationarity of  $N_p$  and  $N_c$ , i.e.  $\mu_p$  and  $\mu_c$  are constant and considering the same function  $h$  on the entire recording period  $[0; T]$  (see Grün *et al.* [52]). But this assumption is quite feasible on smaller time ranges (see Grün [51] and Grammont and Riehle [50]). However, to date, we have no algorithmic and statistical tool to clearly identify the stationarity ranges. Several methods (UE and MTGAUE, see [109] for instance) propose to perform many tests on different small windows of time and to use multiple testing procedure (see Benjamini and Hochberg [8] for instance) to combine them. Hence those methods can solve, at least in practice, this stationarity problem. The aim of this simulation study is not to show how our testing procedure can be incorporated in a Benjamini and Hochberg's approach, which lies outside the scope of the present thesis, but to discuss the advantage of our method on one small window of time. This explains the use of the simulated data described below.

We need therefore to simulate dependence between  $N_p$  and  $N_c$  on  $[0; T]$  and to take into account the major part of the neurobiological reality. So, we simulate processes  $N_p$  and  $N_c$

whose intensities are respectively given by

$$\tilde{\lambda}_p = 50 \quad \text{and} \quad \tilde{\lambda}_c = 20 + \int_{-\infty}^t h(t-u) dN_p(u), \quad \text{with } h = \theta \mathbf{1}_{[\nu; 0.01]}.$$

In order to evaluate the performance of different procedures, several parameters  $\theta$  and  $\nu$  are tested. The parameter  $\theta$  represents the influence strength of  $N_p$  on  $N_c$ : larger the parameter  $\theta$  and more important the influence of  $N_p$  on  $N_c$ . The parameter  $\nu$  introduces a possible minimal delay in the synchronization, i.e. the synchronization of the neuronal activity occurs with a delay  $\delta$  uniform on  $[\nu; 0.01]$ . We consider nine different data sets denoted  $Data_0$ ,  $Data_{10}$ ,  $Data_{30}$ ,  $Data_{50}$ ,  $Data_{80}$ ,  $Data_{10r}$ ,  $Data_{30r}$ ,  $Data_{50r}$  and  $Data_{80r}$ . For  $k \in \{0, 10, 30, 50, 80\}$ ,  $Data_k$  is simulated with  $\theta = k$  and  $\nu = 0$  while  $Data_{kr}$  is simulated with  $\theta = k$  and  $\nu = 0.005$ .

#### 4.4.2 The Kolmogorov-Smirnov test

First, we look at the performance of the Kolmogorov-Smirnov test (see Darling [30]) to convince us that this commonly used test is not reliable in this context. Indeed, even if the KS test is not a test of independence, the KS test may provide an answer to the problem. Since as said before, under  $\mathcal{H}_0$  and conditionally on  $U_1, \dots, U_n$  and  $N_{c,tot} = m$ , the observations of  $N_c$  are i.i.d. with common law the uniform distribution on  $[-1; T + 1]$ , looking for the adequation of  $N_c$  with this law could be an idea to detect the rejection of  $\mathcal{H}_0$ . So, the use of the KS test is relevant.

First, we focus on the empirical rate of the type I error which is an approximation of the level of the test. Thus, we simulate 5000 independent realizations of  $Data_0$ , simulations on which we perform the KS test with level  $\alpha = 0.05$ . The empirical rate of the type I error evaluated on those data is 0.051, which is as desired.

What about the number of wrong rejections of  $\mathcal{H}_1$ ? We consider the power of the tests which is the proportion of correct rejections of  $\mathcal{H}_0$ . We simulate 1000 independent realizations of  $Data_{10}$ ,  $Data_{30}$ ,  $Data_{50}$ ,  $Data_{80}$ ,  $Data_{10r}$ ,  $Data_{30r}$ ,  $Data_{50r}$  and  $Data_{80r}$ , data on which we perform the KS test with level  $\alpha = 0.05$  and we evaluate the empirical power of the test. Table 4.1 summarizes the obtained results. The KS test power is extremely close to the expected level. Hence the KS test is not able to clearly detect the dependence.

Data set	Power for KS
$Data_{10}$	0.040
$Data_{30}$	0.051
$Data_{50}$	0.087
$Data_{80}$	0.113
$Data_{10r}$	0.054
$Data_{30r}$	0.059
$Data_{50r}$	0.053
$Data_{80r}$	0.073

Table 4.1: *Power of the KS test with level  $\alpha = 0.05$ , evaluated for various interactions.*

### 4.4.3 The GAUE method adapted to our context

Before comparing both methods, we briefly return to the principle of the GAUE method. The aim of the GAUE method is to detect the dependence on a single window  $[0; T]$ . This method is based on the coincidences with delay. More precisely for the couple of processes  $(N_p, N_c)$ , we compute the number of coincidences with delay  $\delta$  on  $[0; T]$ , i.e. the variable  $X_T = \int_{[0; T]^2} \mathbf{1}_{|x-y| \leq \delta} dN_p(x) dN_c(y)$ , that represents the number of pairs  $(x, y)$  in  $N_p \times N_c$  such that  $|x - y| \leq \delta$ . Let us define  $\hat{\lambda}_p = N_p([0; T])/T$  and  $\hat{\lambda}_c = N_c([0; T])/T$  where  $N_p([0; T])$  and  $N_c([0; T])$  denote respectively the number of spikes of  $N_p$  and  $N_c$  among  $[0; T]$ . The quantities  $\hat{\lambda}_p$  and  $\hat{\lambda}_c$  are estimators of  $\tilde{\lambda}_p$  and  $\tilde{\lambda}_c$ .

We reject the null hypothesis  $\mathcal{H}_0$ : " $h = 0$ " when  $X_T \geq \hat{m}_0 + \hat{\sigma} u_{1-\alpha/2}$ , where  $\hat{m}_0 = \hat{\lambda}_p \hat{\lambda}_c (2T\delta - \delta^2)$ ,  $\hat{\sigma}^2 = \hat{\lambda}_p \hat{\lambda}_c (2T\delta - \delta^2) + \hat{\lambda}_p \hat{\lambda}_c \left( \hat{\lambda}_p + \hat{\lambda}_c \right) \left( \frac{2}{3} \delta^3 - \frac{1}{T} \delta^4 \right)$  and  $u_{1-\alpha/2}$  is the  $(1 - \alpha/2)$ -quantile of a standard normal. This threshold comes from the theory developed in [109] and is adapted to our context. The quantity  $\hat{m}_0$  is a plug-in estimator of the expectation of  $X_T$  under  $\mathcal{H}_0$  and  $\hat{\sigma}^2$  is an estimator of the variance. It can be shown that under the assumptions " $N_p$  and  $N_c$  are Poisson processes" and " $N_p$  and  $N_c$  are stationary", this test is asymptotically of level  $\alpha$ . Further details about the meaning of those different estimators are given in [109].

The GAUE method was developed jointly with a neurophysiologist and it fits in line the UE method developed by Grün and coauthors (for instance, see [51] and [52]), which is a commonly used method in neurosciences. One of its main disadvantage is that  $\delta$  has to be chosen beforehand. Part of the aim of this work is to propose a more adaptive method.

### 4.4.4 Our procedure in practice

From a theoretical point of view, considering all the resolution levels is not a problem. However, in practice, we have a maximal resolution level, denoted  $j_0$  in the sequel. We choose  $j_0$  quite small for time computational reasons ( $j_0 = 3$ ). Thus, it is better to have the support  $[-A; A]$  of  $h$  with  $A$  close to 1. Nevertheless, if in addition to a global detection, we are interested in a more local detection, i.e. the coefficients  $\lambda \in \Gamma$  for which  $\Phi_{\alpha, \lambda}$  is rejected,  $j_0$  should not be too small. For instance, if  $h = \mathbf{1}_{[0; A]}$  and if the order of magnitude of  $A$  is  $2^{-J}$  or  $1 - 2^{-J}$ , with  $J > j_0 + 1$ , our procedure does not allow to detect locally the jump of  $h$  at  $A$ . Consequently, taking  $A$  close to  $1/2$  may appear reasonable. Hence, the data are multiplied by 50 before being treated.

Let us recall that our test rejects  $\mathcal{H}_0$  when there exists at least one  $\lambda = (j, k)$  in  $\Gamma$  with  $j \leq j_0$  such that

$$\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]} (u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} e^{-w_\lambda}),$$

where  $j_0 \geq 1$  denotes the maximal resolution level,  $u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]}$  is defined by (4.5) and the  $w_\lambda$ 's are given by (4.9). Hence, for each observation of the process  $N_c$  whose number of points is denoted by  $N_{c, tot} = m$ , given the points of  $N_p$  denoted  $U_1, \dots, U_n$ , we estimate  $u_\alpha^{[U_1, \dots, U_n; m]}$  and the quantiles  $q_\lambda^{[U_1, \dots, U_n; m]}$  by classical Monte Carlo methods based on the simulations of  $B$  independent sequences  $\{V^b, 1 \leq b \leq B\}$ , where  $V^b = (V_1^b, \dots, V_m^b)$  is a  $m$ -sample of uniform variables on  $[-1; T+1]$  (i.e. the law of  $N_c$  under  $\mathcal{H}_0$ , conditionally on  $U_1, \dots, U_n$  and  $N_{c, tot} = m$ ). We fix  $B = 20000$  in the sequel since for larger values of  $B$ , the gain in precision for the estimates of  $u_\alpha^{[U_1, \dots, U_n; m]}$  and  $q_\lambda^{[U_1, \dots, U_n; m]}$  becomes negligible. We define for any  $\lambda = (j, k)$  in  $\Gamma$  with  $j \leq j_0$ ,

for  $1 \leq b \leq B$ :

$$\hat{T}_{\lambda,m}^{0,b} = \frac{1}{n} \left| \sum_{k=1}^m \sum_{i=1}^n \left[ \varphi_{\lambda}(V_k^b - U_i) - \frac{n-1}{n} \mathbb{E}_{\pi}(\varphi_{\lambda}(V_k^b - U)) \right] \right|.$$

We compute these  $\hat{T}_{\lambda,m}^{0,b}$ 's with a cascade algorithm (see Mallat [74]).

Half of the  $m$ -samples is used to estimate the quantiles by putting in ascending order the  $\hat{T}_{\lambda,m}^{0,b}$ 's for any  $\lambda$ . The other half is used to approximate the conditional probabilities occurring in (4.5). Then,  $u_{\alpha}^{[U_1, \dots, U_n; m]}$  is obtained by dichotomy, such that the estimated conditional probability occurring in (4.5) is less than  $\alpha$ , but as close as possible to  $\alpha$ . Choosing  $j_0 = 3$ , our procedure considers 15 single tests  $\Phi_{\lambda,\alpha}$  involving wavelets whose support length is respectively 0.125, 0.25, 0.5 and 1. This allows us to make detections at the positions  $m \times 2^{-3}$  ( $m$  in  $\{0, \dots, 7\}$ ) with a range of  $2^{-3}$ . Due to the scaling of the data in our procedure, we need to divide the positions and the range of the possible detections by 50. Consequently, in the real time, the positions and the range become  $m \times 0.0025$  ( $m$  in  $\{0, \dots, 7\}$ ) and 0.0025.

The associated programs are available in Section A.2 of Appendix A.

#### 4.4.5 Results

We compare our testing procedure and the GAUE method on the different data sets. As for the KS test, we look first at the level of both tests. We simulate 5000 independent realizations of  $Data_0$ , simulations on which we perform the present method and the GAUE ones with  $\alpha = 0.05$ . For the GAUE, the tuning parameter  $\delta$  varies on a regular grid of  $[0.001; 0.04]$  with a step 0.001. The order of magnitude of  $\delta$  is similar to the range of the possible detections done with our method. On those data, we evaluate the empirical rate of type I error. Those results, for both methods, are summarized in Table 4.2: both testing methods seem to have a correct level in practice. This means that the number of wrong rejections of  $\mathcal{H}_0$  is well controlled.

Data set	our procedure	GAUE
$Data_0$	0.0486	0.0446/0.0510/0.0548

Table 4.2: *Empirical rate of type I error associated to our procedure and the GAUE method. The theoretical level is  $\alpha = 0.05$ . Since the GAUE method depends on the tuning parameter  $\delta$ , the given value is the minimum/median/maximum of the empirical rate over all the  $\delta$ .*

Now we want to see if the number of wrong rejections of  $\mathcal{H}_1$  is also controlled. To evaluate the power of both tests, we simulate 1000 independent realizations of  $Data_{10}$ ,  $Data_{30}$ ,  $Data_{50}$ ,  $Data_{80}$ ,  $Data_{10r}$ ,  $Data_{30r}$ ,  $Data_{50r}$  and  $Data_{80r}$  as for the KS test (see Section 4.4.2). The results of the empirical power are given by Table 4.3: both methods are comparable in terms of power when  $\nu = 0$  (i.e. for the  $Data_k$ ). However when  $\nu \neq 0$ , our method seems to have better performance since the power is higher.

Moreover, if both methods are comparable in terms of performance, it remains that the testing procedure proposed in this chapter has an advantage over the GAUE method. In fact, our method is statistically adaptive. Indeed, the parameter  $\delta$  which appears in the GAUE method is not calibrated in practice. In our method, we aggregate the single tests over  $(j, k)$ . So on one hand, we do not need to specify this parameter but just an upper bound  $j_0$ , the maximal

Data set	our procedure	GAUE
$Data_{10}$	0.095	0.068/0.1085/0.168
$Data_{30}$	0.478	0.154/0.3795/0.707
$Data_{50}$	0.864	0.278/0.6645/0.953
$Data_{80}$	0.993	0.451/0.9160/0.998
$Data_{10r}$	0.073	0.047/0.0575/0.077
$Data_{30r}$	0.282	0.050/0.1415/0.277
$Data_{50r}$	0.664	0.053/0.2825/0.589
$Data_{80r}$	0.968	0.048/0.4900/0.879

Table 4.3: Powers associated with our procedure and the GAUE method, evaluated for various interactions. The theoretical level is  $\alpha = 0.05$ . Since the GAUE method depends on the tuning parameter  $\delta$ , the given value is the minimum/median/maximum of the empirical rate over all the  $\delta$ .

resolution level: the method through weights (4.9), adapts to this unspecified parameter  $(j, k)$ . But on the other hand, by looking at the single tests  $\Phi_{\lambda, \alpha}$  that have supported the rejection, we are able to partially recover an important information for the practitioner: the position  $(k2^{-j})$  and the range  $(2^{-j})$  of the influence. In fact, by looking only at this single testing procedure, we get an upper value for 0.01 and a lower value for  $\nu$  on the range of delay  $\delta$  of synchronization. To obtain more precise estimations of the support of  $h$ , we can consider an estimate of  $h$ , for instance the one proposed in Chapter 3. The capacity to our method to get an information on  $\nu$  is due to the fact that for a resolution level  $j$  we consider different positions  $k$ . This is not possible with the GAUE method. This explains why the results on  $Data_{10r}$ ,  $Data_{30r}$ ,  $Data_{50r}$  and  $Data_{80r}$  are better with our method.

## 4.5 Conclusion

In our chapter, we have investigated the influence of a point process on another one. We have built a multiple testing procedure based on wavelet thresholding. The main results of the chapter have revealed the optimality of the procedure. Furthermore, our test is adaptive in the minimax sense over classes of alternatives essentially based on weak Besov bodies. Then, from a practical point of view, our method answers several practical questions. However, a number of challenges remain before applying our method on real data. To overcome the problem of stationarity, we could use a Benjamini and Hochberg's approach as for the GAUE method. Finally, we could consider a more sophisticated model that takes into account the phenomenon of self-excitation (as for the complete Hawkes model). But this model raises serious difficulties from the theoretical point of view. This is an exciting challenge.

## 4.6 Proofs

All along the proofs, we introduce some positive constants denoted by  $C(\xi, \dots)$  meaning that they may depend on  $\xi, \dots$ . They do not depend on  $j, n$  and  $T$  (which drive the asymptotic). Furthermore, the values of these constants may vary from line to line.



We recall that  $\{\varphi_\lambda, \lambda \in \Lambda\}$  is the Haar basis and consequently, we have:

$$\|\varphi_\lambda\|_1 = 2^{-j/2}, \quad \|\varphi_\lambda\|_2 = 1 \quad \text{and} \quad \|\varphi_\lambda\|_\infty = 2^{j/2}.$$

In the case of a biorthogonal wavelet basis,  $\|\varphi_\lambda\|_1$ ,  $\|\varphi_\lambda\|_2$  and  $\|\varphi_\lambda\|_\infty$  are of the same order as above, up to a positive constant respectively depending on  $\|\psi\|_1$ ,  $\|\psi\|_2$  and  $\|\psi\|_\infty$ , where  $\psi$  is the mother wavelet associated to the considered biorthogonal wavelet basis. Consequently, the same proofs potentially lead to the results on a biorthogonal wavelet basis as well as in Chapter 3 for the wavelet thresholding estimation.

#### 4.6.1 Proof of Proposition 4.1

We first notice that for any  $\lambda$  in  $\Gamma$ , for any  $u \in [0; T]$ ,

$$\int_{-1}^{T+1} \varphi_\lambda(t-u) dt = 0. \quad (4.10)$$

Let  $\lambda \in \Gamma$  be fixed. By considering the aggregated process (4.2), we can write

$$\mathcal{G}(\varphi_\lambda) = G^0(\varphi_\lambda) + G(\varphi_\lambda), \quad (4.11)$$

with

$$G^0(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x - U)) \right] dN_c^0(x)$$

and

$$G(\varphi_\lambda) = \int_{\mathbb{R}} \sum_{i=1}^n \left[ \varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x - U)) \right] \sum_{j=1}^n dN_c^j(x).$$

On the one hand, we notice that  $G(\varphi_\lambda)$  is the same quantity as the one defined by equation (3.5) of Chapter 3. Thus, by applying the first part of Proposition 3.1 of Chapter 3, we obtain

$$\mathbb{E}(G(\varphi_\lambda)) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx.$$

On the other hand, we have

$$G^0(\varphi_\lambda) = \int_{\mathbb{R}} \varphi_\lambda(x - U_1) dN_c^0(x) + \sum_{i=2}^n \int_{\mathbb{R}} [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dN_c^0(x).$$

Thus,

$$\mathbb{E}(G^0(\varphi_\lambda) | U_1, \dots, U_n) = \int_{-1}^{T+1} \varphi_\lambda(x - U_1) \mu_c dx + \sum_{i=2}^n \int_{-1}^{T+1} [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] \mu_c dx$$

and by using (4.10), we obtain

$$\mathbb{E}(G^0(\varphi_\lambda)) = \sum_{i=2}^n \int_{-1}^{T+1} \mathbb{E}[\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] \mu_c dx = 0.$$

Finally,

$$\mathbb{E}(\hat{\beta}_\lambda) = \mathbb{E}\left(\frac{\mathcal{G}(\varphi_\lambda)}{n}\right) = \int_{\mathbb{R}} \varphi_\lambda(x) h(x) dx = \beta_\lambda,$$

which proves Proposition 4.1.

### 4.6.2 Proof of Proposition 4.2

Let  $\alpha$  be a fixed level in  $]0; 1[$ . Let  $\lambda \in \Gamma$  be fixed. First, the probability that the single test defined by (4.4) wrongly detects a signal is

$$\mathbb{P}_0(\Phi_{\lambda,\alpha} = 1) = \mathbb{P}_0\left(\hat{T}_\lambda > q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(\alpha)\right).$$

Since conditionally on  $U_1, \dots, U_n$  and  $N_{c,tot}$ ,  $\hat{T}_\lambda$  and  $\hat{T}_{\lambda, N_{c,tot}}^0$  have exactly the same distribution under  $\mathcal{H}_0$ ,  $q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(\alpha)$  is also the  $(1 - \alpha)$ -quantile of  $\hat{T}_\lambda | U_1, \dots, U_n; N_{c,tot}$  under  $\mathcal{H}_0$ . Thus,

$$\mathbb{P}_0(\Phi_{\lambda,\alpha} = 1) \leq \alpha$$

and the level of the single test is  $\alpha$ .

Then, the probability that the multiple test defined by (4.6) wrongly detects a signal is

$$\mathbb{P}_0(\Phi_\alpha = 1) = \mathbb{P}_0\left(\max_{\lambda \in \Gamma} \left(\hat{T}_\lambda - q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(u_\alpha^{[U_1, \dots, U_n; N_{c,tot}]} e^{-w_\lambda})\right) > 0\right).$$

By definition (4.5) of  $u_\alpha^{[U_1, \dots, U_n; N_{c,tot}]}$ ,

$$\mathbb{P}_0\left(\max_{\lambda \in \Gamma} \left(\hat{T}_\lambda - q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(u_\alpha^{[U_1, \dots, U_n; N_{c,tot}]} e^{-w_\lambda})\right) > 0 \mid U_1, \dots, U_n; N_{c,tot}\right) \leq \alpha,$$

because conditionally on  $U_1, \dots, U_n$  and  $N_{c,tot}$ ,  $\hat{T}_\lambda$  and  $\hat{T}_{\lambda, N_{c,tot}}^0$  have exactly the same distribution under  $\mathcal{H}_0$ . By taking the expectation over  $U_1, \dots, U_n$  and  $N_{c,tot}$ , we obtain that

$$\mathbb{P}_0(\Phi_\alpha = 1) \leq \alpha$$

and the level of the multiple test is  $\alpha$ .

Furthermore, by Bonferroni's inequality we have

$$\begin{aligned} & \mathbb{P}\left(\max_{\lambda \in \Gamma} \left(\hat{T}_{\lambda, N_{c,tot}}^0 - q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(\alpha e^{-w_\lambda})\right) > 0 \mid U_1, \dots, U_n; N_{c,tot}\right) \\ & \leq \sum_{\lambda \in \Gamma} \mathbb{P}\left(\hat{T}_{\lambda, N_{c,tot}}^0 - q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(\alpha e^{-w_\lambda}) > 0 \mid U_1, \dots, U_n; N_{c,tot}\right) \\ & \leq \sum_{\lambda \in \Gamma} \alpha e^{-w_\lambda} \\ & \leq \alpha \end{aligned}$$

and consequently  $u_\alpha^{[U_1, \dots, U_n; N_{c,tot}]} \geq \alpha$  by definition (4.5) of  $u_\alpha^{[U_1, \dots, U_n; N_{c,tot}]}$ , which concludes the proof of Proposition 4.2.

### 4.6.3 Proof of Theorem 4.1

Let  $\lambda \in \Gamma$  be fixed. Here we want to find a condition which will guarantee that

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leq \beta,$$

given  $\beta \in ]0; 1[$ .

Let us introduce  $q_{1-\beta/2}^\alpha$  the  $(1 - \beta/2)$ -quantile of the conditional quantile  $q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha)$ . Then for any  $h$ ,

$$\begin{aligned} \mathbb{P}_h(\Phi_{\lambda, \alpha} = 0) &= \mathbb{P}_h\left(\hat{T}_\lambda \leq q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha), q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha) \leq q_{1-\beta/2}^\alpha\right) \\ &\quad + \mathbb{P}_h\left(\hat{T}_\lambda \leq q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha), q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha) > q_{1-\beta/2}^\alpha\right) \\ &\leq \mathbb{P}_h(\hat{T}_\lambda \leq q_{1-\beta/2}^\alpha) + \beta/2 \end{aligned}$$

and a condition which guarantees  $\mathbb{P}_h(\hat{T}_\lambda \leq q_{1-\beta/2}^\alpha) \leq \beta/2$  will be enough to ensure that

$$\mathbb{P}_h(\Phi_{\lambda, \alpha} = 0) \leq \beta.$$

The following lemma gives such a condition.

**Lemma 4.1.** *Let  $\alpha, \beta$  be fixed levels in  $]0; 1[$ . For any  $\lambda = (j, k) \in \Gamma$ , if*

$$\mathbb{E}_h(\hat{T}_\lambda) > \sqrt{\frac{2\zeta Q_{j, n, T}}{\beta}} + q_{1-\beta/2}^\alpha \quad (4.12)$$

for a particular  $\zeta$  which is a positive constant depending on  $\mu_c, R_1$  and  $R_\infty$ , where

$$Q_{j, n, T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2},$$

then

$$\mathbb{P}_h(\hat{T}_\lambda \leq q_{1-\beta/2}^\alpha) \leq \beta/2,$$

so that

$$\mathbb{P}_h(\Phi_{\lambda, \alpha} = 0) \leq \beta.$$

The proof of this lemma is postponed in Section 4.6.6.1.

In order to have an idea of the order of the right hand side of (4.12), we are now interested in the control of  $q_{1-\beta/2}^\alpha$ , the  $(1 - \beta/2)$ -quantile of  $q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha)$ . A sharp upper bound for  $q_{1-\beta/2}^\alpha$  is given by the following lemma.

**Lemma 4.2.** *Let  $\alpha, \beta$  be fixed levels in  $]0; 1[$ . For any  $\lambda = (j, k) \in \Gamma$ , there exists some positive constant  $\kappa$  depending on  $\beta, \mu_c$  and  $R_1$  such that*

$$q_{1-\beta/2}^\alpha \leq \kappa \left\{ \sqrt{\ln(2/\alpha)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\}.$$

The proof of this lemma is postponed in Section 4.6.6.2.

Now, observe that if Condition (4.7) of Theorem 4.1 is satisfied, namely

$$|\beta_\lambda| > \sqrt{\frac{2\zeta Q_{j, n, T}}{\beta}} + \kappa \left\{ \sqrt{\ln(2/\alpha)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\},$$

then by Lemma 4.2,

$$|\beta_\lambda| > \sqrt{\frac{2\zeta Q_{j, n, T}}{\beta}} + q_{1-\beta/2}^\alpha.$$

We notice by Jensen's inequality that  $|\beta_\lambda| = |\mathbb{E}_h(\hat{\beta}_\lambda)| \leq \mathbb{E}_h(|\hat{\beta}_\lambda|) = \mathbb{E}_h(\hat{T}_\lambda)$ . Thus, Condition (4.12) of Lemma 4.1 is satisfied and by Lemma 4.1,

$$\mathbb{P}_h(\Phi_{\lambda, \alpha} = 0) \leq \beta,$$

which concludes the proof of Theorem 4.1.

#### 4.6.4 Proof of Theorem 4.2

Since  $u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} \geq \alpha$  (see Proposition 4.2) and by setting  $\alpha_\lambda = \alpha e^{-w_\lambda}$ , we have

$$\begin{aligned} \mathbb{P}_h(\Phi_\alpha = 0) &= \mathbb{P}_h\left(\forall \lambda \in \Gamma, \hat{T}_\lambda \leq q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(u_\alpha^{[U_1, \dots, U_n; N_{c, tot}]} e^{-w_\lambda})\right) \\ &\leq \mathbb{P}_h\left(\forall \lambda \in \Gamma, \hat{T}_\lambda \leq q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha_\lambda)\right) \\ &\leq \min_{\lambda \in \Gamma} \mathbb{P}_h\left(\hat{T}_\lambda \leq q_\lambda^{[U_1, \dots, U_n; N_{c, tot}]}(\alpha_\lambda)\right) \\ &\leq \min_{\lambda \in \Gamma} \mathbb{P}_h(\Phi_{\lambda, \alpha_\lambda} = 0) \\ &\leq \beta, \end{aligned}$$

as soon as there exists  $\lambda$  in  $\Gamma$  such that  $\mathbb{P}_h(\Phi_{\lambda, \alpha_\lambda} = 0) \leq \beta$ .

First, let us give the precise values of the constants that appear in Condition (4.8) of Theorem 4.2:

$$C_1 = 8\left(\frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha)\right), \quad C_2 = 24\kappa^2, \quad C_3 = 12\kappa^2 \ln^2(2/\alpha), \quad C_4 = 24\kappa^2 \ln(2/\alpha) \text{ and } C_5 = 12\kappa^2,$$

where  $\zeta$  and  $\kappa$  are the constants defined respectively by Lemma 4.1 and Lemma 4.2. We recall that  $Q_{j,n,T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2}$  and we denote  $R_{j,n,T} = \frac{j}{n^2} + \frac{j^2 2^j}{n^3} + \frac{2^{-j}}{n^2 T^2}$ .

Let us assume that there exists one finite subset  $L$  of  $\Gamma$  such that Condition (4.8) of Theorem 4.2 is satisfied. Thus,

$$\begin{aligned} \|h_L\|_2^2 &> 8 \left( \left( \frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha) \right) D_L + 3\kappa^2 \sum_{\lambda \in L} w_\lambda \right) \left[ \frac{1}{n} + \frac{n}{T^2} \right] \\ &\quad + \left( 12\kappa^2 \ln^2(2/\alpha) D_L + 24\kappa^2 \ln(2/\alpha) \sum_{\lambda \in L} w_\lambda + 12\kappa^2 \sum_{\lambda \in L} w_\lambda^2 \right) \left[ \frac{j_L}{n^2} + \frac{j_L^2 2^{j_L}}{n^3} + \frac{1}{n^2 T^2} \right]. \end{aligned}$$

Since  $\ln(2/\alpha) + w_\lambda = \ln(2/\alpha_\lambda)$ ,

$$\sum_{\lambda \in L} \beta_\lambda^2 > \sum_{\lambda \in L} \left\{ 8 \left( \frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha_\lambda) \right) \left[ \frac{1}{n} + \frac{n}{T^2} \right] + 12\kappa^2 \ln^2(2/\alpha_\lambda) \left[ \frac{j_L}{n^2} + \frac{j_L^2 2^{j_L}}{n^3} + \frac{1}{n^2 T^2} \right] \right\}$$

and it implies that there exists one coefficient  $\lambda = (j, k)$  in  $\Gamma$  such that

$$\beta_\lambda^2 > 8 \left( \frac{\zeta}{\beta} + 3\kappa^2 \ln(2/\alpha_\lambda) \right) \left[ \frac{1}{n} + \frac{n}{T^2} \right] + 12\kappa^2 \ln^2(2/\alpha_\lambda) \left[ \frac{j}{n^2} + \frac{j^2 2^j}{n^3} + \frac{1}{n^2 T^2} \right].$$

Seeing that  $Q_{j,n,T} \leq 2 \left[ \frac{1}{n} + \frac{n}{T^2} \right]$  and  $R_{j,n,T} \leq \left[ \frac{j}{n^2} + \frac{j^2 2^j}{n^3} + \frac{1}{n^2 T^2} \right]$ , we have:

$$\beta_\lambda^2 > 4 \frac{\zeta}{\beta} Q_{j,n,T} + 12\kappa^2 \ln(2/\alpha_\lambda) Q_{j,n,T} + 12\kappa^2 \ln^2(2/\alpha_\lambda) R_{j,n,T}.$$

Since  $(\sqrt{a} + \sqrt{b} + \sqrt{c})^2 \leq 3(a + b + c)$  for all  $a, b, c$  nonnegative real numbers,

$$\beta_\lambda^2 > 4 \frac{\zeta}{\beta} Q_{j,n,T} + 4\kappa^2 \ln(2/\alpha_\lambda) \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2} \sqrt{n}}{T} \right)^2 + 4\kappa^2 \ln^2(2/\alpha_\lambda) \left( \frac{\sqrt{j}}{n} + \frac{j 2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right)^2$$

and then,

$$\beta_\lambda^2 > \left( \sqrt{\frac{2\zeta}{\beta} Q_{j,n,T} + \kappa} \left\{ \sqrt{\ln(2/\alpha_\lambda)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha_\lambda) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\} \right)^2.$$

Finally, it is equivalent to

$$|\beta_\lambda| > \sqrt{\frac{2\zeta}{\beta} Q_{j,n,T} + \kappa} \left\{ \sqrt{\ln(2/\alpha_\lambda)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T} \right) + \ln(2/\alpha_\lambda) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\},$$

which is exactly Condition (4.7) of Theorem 4.1 and we conclude the proof of Theorem 4.2 by applying Theorem 4.1.

#### 4.6.5 Proof of Theorem 4.3

With  $T$  proportional to  $n$ , Condition (4.8) of Theorem 4.2 is satisfied if there exists one finite subset  $L$  of  $\Gamma$  such that

$$\|h\|_2^2 \geq \|h - h_L\|_2^2 + C(\alpha, \beta, \mu_c, R_1, R_\infty) \left\{ \left( D_L + \sum_{\lambda \in L} w_\lambda \right) \frac{1}{n} + \left( D_L + \sum_{\lambda \in L} w_\lambda + \sum_{\lambda \in L} w_\lambda^2 \right) \left[ \frac{j_L}{n^2} + \frac{j_L^2 2^{j_L}}{n^3} \right] \right\},$$

with  $j_L = \max\{j \geq 0 : (j, k) \in L\}$ ,  $\sum_{\lambda \in L} w_\lambda \leq C \times (j_L + 1)D_L$  and  $\sum_{\lambda \in L} w_\lambda^2 \leq C \times (j_L + 1)^2 D_L$ . Consequently, Condition (4.8) is satisfied if there exists one finite subset  $L$  of  $\Gamma$  such that

$$\|h\|_2^2 \geq \|h - h_L\|_2^2 + C(\alpha, \beta, \mu_c, R_1, R_\infty) \frac{(j_L + 1)}{n} D_L, \quad (4.13)$$

with the maximal resolution level  $j_L$  such that  $2^{j_L} \leq n^2 / (\ln n)^4$ .

Let  $J \geq 1$  that will be chosen later. We consider the following finite subset  $\Gamma_J$  of  $\Gamma$

$$\Gamma_J = \{\lambda = (j, k) \in \Gamma : 0 \leq j \leq J, k \in \mathcal{K}_j\}.$$

We introduce for all integer  $D \leq |\Gamma_J|$  the subset  $L$  of  $\Gamma_J$  such that  $\{\beta_\lambda, \lambda \in L\}$  is the set of the  $D$  largest coefficients among  $\{\beta_\lambda, \lambda \in \Gamma_J\}$ . We can notice that

$$\|h - h_L\|_2^2 = \|h - h_{\Gamma_J}\|_2^2 + \|h_{\Gamma_J} - h_L\|_2^2.$$

On the one hand, since  $h$  belongs to  $\mathcal{B}_{2,\infty}^\delta(R)$ ,

$$\|h - h_{\Gamma_J}\|_2^2 = \sum_{j>J} \sum_{k \in \mathcal{K}_j} \beta_{(j,k)}^2 \leq C(\delta) R^2 2^{-2J\delta}.$$

On the other hand, using equivalent definitions of weak Besov balls given by Lemma 2.2 of [66] and using for instance page 211 of [45], we obtain:

$$\|h_{\Gamma_J} - h_L\|_2^2 \leq C(\gamma) R''^{2+4\gamma} D^{-2\gamma},$$

since  $h$  belongs to  $\mathcal{W}_{\frac{2}{1+2\gamma}}^*(R')$ , with  $R''$  an absolute positive constant depending eventually on  $\gamma$  and  $R'$ .

Taking

$$J = \lfloor \log_2(n^\varepsilon) \rfloor + 1$$

for some  $0 < \varepsilon < 2$ , we obtain that the right hand side of (4.13) is upper bounded by

$$C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left( n^{-2\varepsilon\delta} + D^{-2\gamma} + \frac{\varepsilon D \ln n}{n} \right).$$

Taking  $D = \lfloor (n/\ln n)^{1/(1+2\gamma)} \rfloor$  and  $\varepsilon > \gamma/(\delta(1+2\gamma))$ , we obtain that the right hand side of (4.13) is upper bounded by

$$C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left( \frac{n}{\ln n} \right)^{\frac{-2\gamma}{1+2\gamma}}$$

when  $2\delta > \gamma/(1+2\gamma)$  and so,

$$\rho(\Phi_\alpha, \mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_{\frac{2}{1+2\gamma}}^*(R'), \beta) \leq C(\delta, \gamma, R, R', \alpha, \beta, \mu_c, R_1, R_\infty) \left( \frac{n}{\ln n} \right)^{\frac{-\gamma}{1+2\gamma}},$$

which concludes the proof of Theorem 4.3.

## 4.6.6 Proof of lemmas

### 4.6.6.1 Proof of Lemma 4.1

Let  $\lambda \in \Gamma$  be fixed. From Markov's inequality, we have that for any  $x > 0$ ,

$$\mathbb{P}_h \left( \left| \hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \right| \geq x \right) \leq \frac{\text{Var}(\hat{T}_\lambda)}{x^2}. \quad (4.14)$$

Let us control  $\text{Var}(\hat{T}_\lambda) = \mathbb{E}_h(\hat{T}_\lambda^2) - \mathbb{E}_h^2(\hat{T}_\lambda)$ . We easily obtain by Jensen's inequality and by considering the decomposition (4.11) of  $\mathcal{G}(\varphi_\lambda)$ :

$$\begin{aligned} \text{Var}(\hat{T}_\lambda) &\leq \text{Var}(\hat{\beta}_\lambda) \\ &\leq \frac{1}{n^2} \text{Var}(G^0(\varphi_\lambda) + G(\varphi_\lambda)) \\ &\leq \frac{2}{n^2} [\text{Var}(G^0(\varphi_\lambda)) + \text{Var}(G(\varphi_\lambda))], \end{aligned}$$

with

$$\text{Var}(G(\varphi_\lambda)) \leq C(R_1, R_\infty) \left\{ n + \frac{n^2}{T} + \frac{2^{-j} n^3}{T^2} \right\},$$

by applying the second part of Proposition 3.1 of Chapter 3. It remains to compute  $\text{Var}(G^0(\varphi_\lambda))$ . For this purpose, we apply the same methodology developed in Section 3.6.1.2 of Chapter 3. We have the following decomposition of  $\text{Var}(G^0(\varphi_\lambda))$  into two terms:

$$\text{Var}(G^0(\varphi_\lambda)) = \mathbb{E}(\text{Var}(G^0(\varphi_\lambda)|U_1, \dots, U_n)) + \text{Var}(\mathbb{E}(G^0(\varphi_\lambda)|U_1, \dots, U_n)). \quad (4.15)$$

We start by dealing with the first term of (4.15). We have

$$\begin{aligned} &\text{Var}(G^0(\varphi_\lambda)|U_1, \dots, U_n) \\ &= \int_{-1}^{T+1} \left( \sum_{i=1}^n \left[ \varphi_\lambda(x - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(x - U)) \right] \right)^2 \mu_c dx \\ &= \mu_c \int_{-1}^{T+1} \left( \varphi_\lambda(x - U_1) + \sum_{i=2}^n [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] \right)^2 dx \end{aligned}$$

$$\begin{aligned}
&= \mu_c \int_{-1}^{T+1} \varphi_\lambda^2(x - U_1) dx + 2\mu_c \int_{-1}^{T+1} \varphi_\lambda(x - U_1) \sum_{i=2}^n [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx \\
&\quad + \mu_c \int_{-1}^{T+1} \sum_{i=2}^n \sum_{k=2}^n [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] [\varphi_\lambda(x - U_k) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx.
\end{aligned}$$

In the first integral, write  $y = x - U_1$ . So,

$$\begin{aligned}
&\text{Var}(G^0(\varphi_\lambda)|U_1, \dots, U_n) \\
&= \mu_c \|\varphi_\lambda\|_2^2 + 2\mu_c \int_{-1}^{T+1} \varphi_\lambda(x - U_1) \sum_{i=2}^n [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx \\
&\quad + \mu_c \int_{-1}^{T+1} \sum_{i=2}^n \sum_{k=2}^n [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] [\varphi_\lambda(x - U_k) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}(\text{Var}(G^0(\varphi_\lambda)|U_1, \dots, U_n)) &= \mu_c \|\varphi_\lambda\|_2^2 + \mu_c \int_{-1}^{T+1} \sum_{i=2}^n \mathbb{E} \left( [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))]^2 \right) dx \\
&= \mu_c \|\varphi_\lambda\|_2^2 + (n-1)\mu_c \int_{-1}^{T+1} \text{Var}_\pi(\varphi_\lambda(x - U)) dx \\
&\leq \mu_c \|\varphi_\lambda\|_2^2 + (n-1)\mu_c(T+2) \frac{\|\varphi_\lambda\|_2^2}{T} \\
&\leq C(\mu_c)n,
\end{aligned} \tag{4.16}$$

by using (4.10) and Lemma 3.1 of Chapter 3.

Now, we deal with the second term of (4.15). We have

$$\begin{aligned}
\mathbb{E}(G^0(\varphi_\lambda)|U_1, \dots, U_n) &= \int_{-1}^{T+1} \varphi_\lambda(x - U_1) \mu_c dx + \sum_{i=2}^n \int_{-1}^{T+1} [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] \mu_c dx \\
&= \mu_c \sum_{i=2}^n \int_{-1}^{T+1} [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx,
\end{aligned}$$

by using (4.10). Therefore,

$$\begin{aligned}
\text{Var}(\mathbb{E}(G^0(\varphi_\lambda)|U_1, \dots, U_n)) &= \mu_c^2 \text{Var} \left( \sum_{i=2}^n \int_{-1}^{T+1} [\varphi_\lambda(x - U_i) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx \right) \\
&= \mu_c^2 (n-1) \text{Var} \left( \int_{-1}^{T+1} [\varphi_\lambda(x - U_1) - \mathbb{E}_\pi(\varphi_\lambda(x - U))] dx \right) \\
&\leq \mu_c^2 (n-1) \mathbb{E} \left[ \left( \int_{-1}^{T+1} |\varphi_\lambda(x - U_1)| dx \right)^2 \right] \\
&\leq \mu_c^2 (n-1) \|\varphi_\lambda\|_1^2 \\
&\leq C(\mu_c) 2^{-j} n.
\end{aligned} \tag{4.17}$$

Finally, by combining inequalities (4.15), (4.16) and (4.17), we obtain:

$$\text{Var}(G^0(\varphi_\lambda)) \leq C(\mu_c)n.$$

Thus,

$$\text{Var}(\hat{T}_\lambda) \leq \frac{C(\mu_c, R_1, R_\infty)}{n^2} \left\{ n + \frac{n^2}{T} + \frac{2^{-j}n^3}{T^2} \right\} \leq \zeta Q_{j,n,T},$$

with

$$Q_{j,n,T} = \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2}$$

and  $\zeta$  a positive constant depending on  $\mu_c$ ,  $R_1$  and  $R_\infty$ .

Taking  $x = \sqrt{2\zeta Q_{j,n,T}/\beta}$  in (4.14) and using the previous inequality leads to

$$\mathbb{P}_h \left( \left| \hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \right| \geq \sqrt{2\zeta Q_{j,n,T}/\beta} \right) \leq \frac{\beta}{2}.$$

Therefore, if  $\mathbb{E}_h(\hat{T}_\lambda) > \sqrt{2\zeta Q_{j,n,T}/\beta} + q_{1-\beta/2}^\alpha$ , then

$$\begin{aligned} \mathbb{P}_h(\hat{T}_\lambda \leq q_{1-\beta/2}^\alpha) &= \mathbb{P}_h(\hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \leq q_{1-\beta/2}^\alpha - \mathbb{E}_h(\hat{T}_\lambda)) \\ &\leq \mathbb{P}_h \left( \left| \hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \right| \geq \mathbb{E}_h(\hat{T}_\lambda) - q_{1-\beta/2}^\alpha \right) \\ &\leq \mathbb{P}_h \left( \left| \hat{T}_\lambda - \mathbb{E}_h(\hat{T}_\lambda) \right| \geq \sqrt{2\zeta Q_{j,n,T}/\beta} \right) \\ &\leq \beta/2 \end{aligned}$$

and so

$$\mathbb{P}_h(\Phi_{\lambda,\alpha} = 0) \leq \beta,$$

which concludes the proof of Lemma 4.1.

#### 4.6.6.2 Proof of Lemma 4.2

We focus first on the control of the conditional quantile  $q_\lambda^{[U_1, \dots, U_n; N_{c,tot}]}(\alpha)$ . For all  $m \in \mathbb{N}^*$ , the  $(1 - \alpha)$ -quantile  $q_\lambda^{[U_1, \dots, U_n; m]}(\alpha)$  is the smallest real number such that

$$\mathbb{P} \left( \hat{T}_{\lambda,m}^0 > q_\lambda^{[U_1, \dots, U_n; m]}(\alpha) \mid U_1, \dots, U_n; N_{c,tot} = m \right) \leq \alpha,$$

where  $\hat{T}_{\lambda,m}^0$  is defined by (4.3). Let  $m \in \mathbb{N}^*$  be fixed. We write

$$\hat{T}_{\lambda,m}^0 = \frac{1}{n} \left| \sum_{k=1}^m S(\varphi_\lambda)(V_k^0) \right|,$$

where  $(V_1^0, \dots, V_m^0)$  is a  $m$ -sample with uniform distribution on  $[-1; T + 1]$  and for any  $v \in [-1; T + 1]$ ,

$$S(\varphi_\lambda)(v) = \sum_{i=1}^n \left[ \varphi_\lambda(v - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(v - U)) \right].$$

Since  $\mathbb{E}(\varphi_\lambda(V - U) | U) = 0$  for independent random variables  $U$  and  $V$  uniformly distributed on  $[0; T]$  and  $[-1; T + 1]$  respectively, the  $S(\varphi_\lambda)(V_k^0)$ 's are centered and independent conditionally on  $U_1, \dots, U_n$ . Then we apply Bernstein's inequality (for instance, see Proposition 2.9 of [77]) to get that for all  $\omega > 0$ , with probability larger than  $1 - 2e^{-\omega}$ ,

$$\left| \sum_{k=1}^m S(\varphi_\lambda)(V_k^0) \right| \leq \sqrt{2m \text{Var}(S(\varphi_\lambda)(V_1^0) | U_1, \dots, U_n) \omega} + \frac{\omega}{3} \sup_{v \in [-1; T+1]} |S(\varphi_\lambda)(v)|.$$



Thus, with probability larger than  $1 - \alpha$ ,

$$\hat{T}_{\lambda,m}^0 \leq f(U_1, \dots, U_n; m),$$

with

$$f(U_1, \dots, U_n; m) = \frac{1}{n} \left\{ \sqrt{2m \ln(2/\alpha) V_S} + \frac{\ln(2/\alpha)}{3} B_S \right\}, \quad (4.18)$$

where

$$V_S = \text{Var}(S(\varphi_\lambda)(V_1^0) | U_1, \dots, U_n) \quad \text{and} \quad B_S = \sup_{v \in [-1; T+1]} |S(\varphi_\lambda)(v)|.$$

Therefore we have  $q_\lambda^{[U_1, \dots, U_n; m]}(\alpha) \leq f(U_1, \dots, U_n; m)$  by definition of the quantile  $q_\lambda^{[U_1, \dots, U_n; m]}(\alpha)$ .

Let us now provide a control in probability of  $f(U_1, \dots, U_n; m)$ . We control first  $V_S$ .

$$\begin{aligned} V_S &= \text{Var} \left( \sum_{i=1}^n \varphi_\lambda(V_1^0 - U_i) - (n-1) \mathbb{E}_\pi(\varphi_\lambda(V_1^0 - U)) \middle| U_1, \dots, U_n \right) \\ &\leq \mathbb{E} \left[ \left( \sum_{i=1}^n \varphi_\lambda(V_1^0 - U_i) - (n-1) \mathbb{E}_\pi(\varphi_\lambda(V_1^0 - U)) \right)^2 \middle| U_1, \dots, U_n \right] \\ &\leq \frac{1}{T+2} \int_{v=-1}^{T+1} \left( \sum_{i=1}^n \varphi_\lambda(v - U_i) - (n-1) \mathbb{E}_\pi(\varphi_\lambda(v - U)) \right)^2 dv \\ &\leq \frac{2}{T+2} \int_{v=-1}^{T+1} \left( \sum_{1 \leq i, k \leq n} \varphi_\lambda(v - U_i) \varphi_\lambda(v - U_k) + (n-1)^2 \mathbb{E}_\pi^2(\varphi_\lambda(v - U)) \right) dv \\ &\leq \frac{2}{T+2} \left\{ \int_{v=-1}^{T+1} \sum_{i=1}^n \varphi_\lambda^2(v - U_i) dv + \int_{v=-1}^{T+1} \sum_{1 \leq i \neq k \leq n} \varphi_\lambda(v - U_i) \varphi_\lambda(v - U_k) dv \right. \\ &\quad \left. + \frac{(n-1)^2}{T^2} \int_{v=-1}^{T+1} \left( \int_0^T |\varphi_\lambda|(v-u) du \right)^2 dv \right\} \\ &\leq \frac{2}{T+2} \left\{ n \|\varphi_\lambda\|_2^2 + \int_{v=-1}^{T+1} \sum_{1 \leq i \neq k \leq n} \varphi_\lambda(v - U_i) \varphi_\lambda(v - U_k) dv + \frac{(n-1)^2}{T^2} (T+2) \|\varphi_\lambda\|_1^2 \right\} \\ &\leq \frac{C}{T} \left\{ n + \sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} \varphi_\lambda(v - U_i) \varphi_\lambda(v - U_k) dv + \frac{2^{-j} n^2}{T} \right\}, \quad (4.19) \end{aligned}$$

with  $C$  an absolute positive constant. We have a decomposition of the second term in a sum of degenerate  $U$ -statistics of order 0, 1 and 2. Indeed

$$\sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} \varphi_\lambda(v - U_i) \varphi_\lambda(v - U_k) dv = W_0 + W_{1,1} + W_{1,2} + W_2,$$

with

$$W_2 = \sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} [\varphi_\lambda(v - U_i) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] [\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] dv,$$

$$W_{1,1} = \sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} \varphi_\lambda(v - U_i) \mathbb{E}_\pi(\varphi_\lambda(v - U)) dv,$$

$$W_{1,2} = \sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} \mathbb{E}_\pi(\varphi_\lambda(v-U)) \varphi_\lambda(v-U_k) dv$$

and

$$W_0 = - \sum_{1 \leq i \neq k \leq n} \int_{v=-1}^{T+1} \mathbb{E}_\pi^2(\varphi_\lambda(v-U)) dv.$$

First we control  $W_0$ :

$$\begin{aligned} |W_0| &\leq \frac{n(n-1)(T+2)}{T^2} \|\varphi_\lambda\|_1^2 \\ &\leq C \frac{2^{-j} n^2}{T}, \end{aligned} \quad (4.20)$$

with  $C$  an absolute positive constant. Next we deal with the control of  $W_{1,1}$  and  $W_{1,2}$ . We notice that

$$W_{1,1} = W_{1,2} = (n-1) \sum_{i=1}^n \int_{v=-1}^{T+1} \varphi_\lambda(v-U_i) \mathbb{E}_\pi(\varphi_\lambda(v-U)) dv$$

and consequently we have by using Lemma 3.3 of Chapter 3

$$\begin{aligned} |W_{1,1}| = |W_{1,2}| &\leq (n-1) \sum_{i=1}^n \int_{v=-1}^{T+1} |\varphi_\lambda|(v-U_i) dv \frac{\|\varphi_\lambda\|_1}{T} \\ &\leq C \frac{2^{-j} n^2}{T}, \end{aligned} \quad (4.21)$$

with  $C$  an absolute positive constant.

Now it remains to control  $W_2$ , with

$$W_2 = \sum_{1 \leq i < k \leq n} g(U_i, U_k),$$

where

$$g(U_i, U_k) = 2 \int_{v=-1}^{T+1} [\varphi_\lambda(v-U_i) - \mathbb{E}_\pi(\varphi_\lambda(v-U))] [\varphi_\lambda(v-U_k) - \mathbb{E}_\pi(\varphi_\lambda(v-U))] dv.$$

One can apply Theorem 3.4 of [61] to  $W_2$  and  $-W_2$ . It implies that there exist absolute positive constants  $c_1, c_2, c_3$  and  $c_4$  such that with probability larger than  $1 - 2 \times 2.77e^{-\omega}$ ,

$$|W_2| \leq c_1 C \sqrt{\omega} + c_2 D \omega + c_3 B \omega^{3/2} + c_4 A \omega^2$$

for all  $\omega > 0$ , where

- $A = \|g\|_\infty \leq 8 \|\varphi_\lambda\|_1 \|\varphi_\lambda\|_\infty \leq 8$ ;
- $C^2 = \mathbb{E}(W_2^2)$  and we have

$$\begin{aligned} C^2 &= \sum_{1 \leq i < k \leq n} \mathbb{E}(g^2(U_i, U_k)) \\ &\leq 4n(n-1) \mathbb{E} \left[ \left( \int_{v=-1}^{T+1} [\varphi_\lambda(v-U_1) - \mathbb{E}_\pi(\varphi_\lambda(v-U))] [\varphi_\lambda(v-U_2) - \mathbb{E}_\pi(\varphi_\lambda(v-U))] dv \right)^2 \right]. \end{aligned}$$

We denote  $\mathbb{E}_{(U,U') \sim \pi \otimes \pi}(f(U, U'))$  the expectation of  $f(U, U')$  where  $U \sim \pi$  and  $U' \sim \pi$  are independent and  $f^U(v) = f(v - U)$ . Hence,

$$\begin{aligned} \mathcal{C}^2 &\leq 4n(n-1) \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left[ \left( \int_{v=-1}^{T+1} [\varphi_\lambda^U(v) - \mathbb{E}_\pi(\varphi_\lambda^U(v))] [\varphi_\lambda^{U'}(v) - \mathbb{E}_\pi(\varphi_\lambda^{U'}(v))] dv \right)^2 \right] \\ &\leq 4n^2 \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left[ \left( \int_{v=-1}^{T+1} \varphi_\lambda^U(v) \varphi_\lambda^{U'}(v) dv - \mathbb{E}_{U' \sim \pi} \left( \int_{v=-1}^{T+1} \varphi_\lambda^U(v) \varphi_\lambda^{U'}(v) dv \right) \right. \right. \\ &\quad \left. \left. - \mathbb{E}_{U \sim \pi} \left( \int_{v=-1}^{T+1} \varphi_\lambda^U(v) \varphi_\lambda^{U'}(v) dv \right) + \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left( \int_{v=-1}^{T+1} \varphi_\lambda^U(v) \varphi_\lambda^{U'}(v) dv \right) \right)^2 \right] \\ &\leq Cn^2 \left\{ \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left[ \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U(v)| |\varphi_\lambda^{U'}(v)| dv \right)^2 \right] \right. \\ &\quad \left. + \left[ \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U(v)| |\varphi_\lambda^{U'}(v)| dv \right) \right]^2 \right\}, \end{aligned}$$

with  $C$  an absolute positive constant. But,

$$\begin{aligned} &\mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left[ \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U(v)| |\varphi_\lambda^{U'}(v)| dv \right)^2 \right] \\ &\leq \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U|^2(v) |\varphi_\lambda^{U'}(v)| dv \int_{v=-1}^{T+1} |\varphi_\lambda^{U'}(v)| dv \right) \\ &= \mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U|^2(v) |\varphi_\lambda^{U'}(v)| dv \right) \|\varphi_\lambda\|_1 \\ &\leq \|\varphi_\lambda\|_2^2 \frac{\|\varphi_\lambda\|_1^2}{T} \end{aligned}$$

and

$$\mathbb{E}_{(U,U') \sim \pi \otimes \pi} \left( \int_{v=-1}^{T+1} |\varphi_\lambda^U(v)| |\varphi_\lambda^{U'}(v)| dv \right) = \int \mathbb{E}_\pi(|\varphi_\lambda^U(v)|) \mathbb{E}_\pi(|\varphi_\lambda^{U'}(v)|) dv \leq \frac{(T+2)\|\varphi_\lambda\|_1^2}{T^2},$$

by using Lemma 3.3 of Chapter 3. So,

$$\mathcal{C}^2 \leq Cn^2 \left\{ \frac{2^{-j}}{T} + \frac{2^{-2j}}{T^2} \right\},$$

with  $C$  an absolute positive constant;

- $D = \sup \left\{ \mathbb{E} \left( \sum_{1 \leq k < i \leq n} g(U_i, U_k) a_i(U_i) b_k(U_k) \right) : \mathbb{E} \left( \sum_{i=2}^n a_i(U_i)^2 \right) \leq 1, \mathbb{E} \left( \sum_{k=1}^{n-1} b_k(U_k)^2 \right) \leq 1 \right\}$ .

But, with the conditions on the  $a_i$ 's and the  $b_k$ 's, we have:

$$\begin{aligned} &\mathbb{E} \left( \sum_{1 \leq k < i \leq n} g(U_i, U_k) a_i(U_i) b_k(U_k) \right) \\ &= 2 \mathbb{E} \left( \sum_{i=2}^n \sum_{k=1}^{i-1} \int [\varphi_\lambda(v - U_i) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] [\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] dv a_i(U_i) b_k(U_k) \right) \end{aligned}$$

$$\begin{aligned}
&\leq 2 \int \mathbb{E} \left( \sum_{i=2}^n |\varphi_\lambda(v - U_i) - \mathbb{E}_\pi(\varphi_\lambda^U(v))| |a_i(U_i)| \right) \mathbb{E} \left( \sum_{k=1}^{n-1} |\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda^U(v))| |b_k(U_k)| \right) dv \\
&\leq 2 \int_{v=-1}^{T+1} \sqrt{(n-1) \text{Var}_\pi(\varphi_\lambda(v - U))} \mathbb{E} \left( \sum_{k=1}^{n-1} |\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda^U(v))| |b_k(U_k)| \right) dv \\
&\leq 2 \sqrt{(n-1) \frac{\|\varphi_\lambda\|_2^2}{T}} \mathbb{E} \left( \sum_{k=1}^{n-1} \int_{v=-1}^{T+1} |\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda(v - U))| |b_k(U_k)| dv \right) \\
&\leq 2 \sqrt{\frac{n-1}{T}} \|\varphi_\lambda\|_2 \mathbb{E} \left( 2 \|\varphi_\lambda\|_1 \sum_{k=1}^{n-1} |b_k(U_k)| \right) \\
&\leq 4 \sqrt{\frac{n-1}{T}} \|\varphi_\lambda\|_2 \|\varphi_\lambda\|_1 \sqrt{n-1} \\
&\leq 4 \frac{n-1}{\sqrt{T}} \|\varphi_\lambda\|_1 \|\varphi_\lambda\|_2,
\end{aligned}$$

by using Lemma 3.1 of Chapter 3. Then,

$$D \leq C \frac{2^{-j/2} n}{\sqrt{T}},$$

with  $C$  an absolute positive constant;

- $B^2 = \sup_u \left( \sum_{k=1}^{n-1} \mathbb{E}(g^2(u, U_k)) \right)$ , with

$$\begin{aligned}
&\mathbb{E}(g^2(u, U_k)) \\
&= 4 \mathbb{E} \left[ \left( \int_{v=-1}^{T+1} [\varphi_\lambda(v - u) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] [\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] dv \right)^2 \right] \\
&\leq 4 \mathbb{E} \left[ \int_{v=-1}^{T+1} [\varphi_\lambda^u(v) - \mathbb{E}_\pi(\varphi_\lambda^U(v))]^2 |\varphi_\lambda^{U_k}(v) - \mathbb{E}_\pi(\varphi_\lambda^U(v))| dv \int_{v=-1}^{T+1} |\varphi_\lambda^{U_k}(v) - \mathbb{E}_\pi(\varphi_\lambda^U(v))| dv \right] \\
&\leq 8 \mathbb{E} \left[ \int_{v=-1}^{T+1} [\varphi_\lambda(v - u) - \mathbb{E}_\pi(\varphi_\lambda(v - U))]^2 |\varphi_\lambda(v - U_k) - \mathbb{E}_\pi(\varphi_\lambda(v - U))| dv \right] \|\varphi_\lambda\|_1 \\
&\leq \frac{16}{T} \int_{v=-1}^{T+1} [\varphi_\lambda(v - u) - \mathbb{E}_\pi(\varphi_\lambda(v - U))]^2 dv \|\varphi_\lambda\|_1^2 \\
&\leq \frac{64}{T} \|\varphi_\lambda\|_1^2 \|\varphi_\lambda\|_2^2,
\end{aligned}$$

by using Lemma 3.3 of Chapter 3. Hence,

$$B^2 \leq C \frac{2^{-j} n}{T},$$

with  $C$  an absolute positive constant.

Finally, we obtain for all  $\omega > 0$ , with probability larger than  $1 - 2 \times 2.77e^{-\omega}$ ,

$$|W_2| \leq C \left\{ \frac{2^{-j/2} n}{\sqrt{T}} \sqrt{\omega} + \frac{2^{-j} n}{T} \sqrt{\omega} + \frac{2^{-j/2} n}{\sqrt{T}} \omega + \frac{2^{-j/2} \sqrt{n}}{\sqrt{T}} \omega^{3/2} + \omega^2 \right\}, \quad (4.22)$$

with  $C$  an absolute positive constant.

Thus, by inequalities (4.19), (4.20), (4.21) and (4.22), for all  $\omega > 0$ , with probability larger than  $1 - 2 \times 2.77e^{-\omega}$ ,

$$V_S \leq \frac{C(\omega)}{T} \left\{ n + \frac{2^{-j}n^2}{T} + \frac{2^{-j/2}n}{\sqrt{T}} \right\}. \quad (4.23)$$

Then it remains to compute  $B_S$ . We recall that

$$\begin{aligned} B_S &= \sup_{v \in [-1; T+1]} \left| \sum_{i=1}^n \left[ \varphi_\lambda(v - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(v - U)) \right] \right| \\ &\leq \tilde{B}_S + \frac{\|\varphi_\lambda\|_1}{T}, \end{aligned}$$

with  $\tilde{B}_S = \sup_{v \in [-1; T+1]} \left| \sum_{i=1}^n [\varphi_\lambda(v - U_i) - \mathbb{E}_\pi(\varphi_\lambda(v - U))] \right|$ . Since the Haar basis is considered here, we can write for any  $x \in \mathbb{R}$ :

$$\varphi_\lambda(x) = 2^{j/2} \left( \mathbf{1}_{(2k+1)2^{-(j+1)} < x \leq (k+1)2^{-j}} - \mathbf{1}_{k2^{-j} \leq x \leq (2k+1)2^{-(j+1)}} \right),$$

with  $\lambda = (j, k)$ . Thus,

$$\tilde{B}_S \leq 2^{j/2} \left( \tilde{B}_S^1 + \tilde{B}_S^2 \right),$$

where

$$\tilde{B}_S^1 = \sup_{v \in [-1; T+1]} \left| \sum_{i=1}^n \left[ \mathbf{1}_{k2^{-j} \leq v - U_i \leq (2k+1)2^{-(j+1)}} - \mathbb{E}_\pi(\mathbf{1}_{k2^{-j} \leq v - U \leq (2k+1)2^{-(j+1)}}) \right] \right|$$

and

$$\tilde{B}_S^2 = \sup_{v \in [-1; T+1]} \left| \sum_{i=1}^n \left[ \mathbf{1}_{(2k+1)2^{-(j+1)} < v - U_i \leq (k+1)2^{-j}} - \mathbb{E}_\pi(\mathbf{1}_{(2k+1)2^{-(j+1)} < v - U \leq (k+1)2^{-j}}) \right] \right|.$$

We observe that

$$\tilde{B}_S^1 \leq \sup_{B_v, v \in \mathbb{R}} \left| \sum_{i=1}^n [\mathbf{1}_{B_v}(U_i) - \mathbb{E}_\pi(\mathbf{1}_{B_v}(U))] \right|,$$

where for any  $v \in \mathbb{R}$ ,  $B_v = [v - (2k+1)2^{-(j+1)}; v - k2^{-j}]$ . We set  $\mathcal{B} = \{B_v, v \in \mathbb{R}\}$  and for every integer  $n$ ,  $m_n(\mathcal{B}) = \sup_{A \subset \mathbb{R}, |A|=n} |\{A \cap B_v, v \in \mathbb{R}\}|$ . It is easy to see that

$$m_n(\mathcal{B}) \leq 1 + \frac{n(n+1)}{2}$$

and so, the VC-dimension  $\mathcal{V}$  of  $\mathcal{B}$  defined by  $\sup\{n \geq 0, m_n(\mathcal{B}) = 2^n\}$  is bounded by 2 (see Definition 6.2 of [77]). Let us define  $\sigma^2 = \max \left\{ 2^{-(j+1)}, K^2 \mathcal{V} \left( 1 + \frac{j+1}{2} \ln 2 \right) / n \right\}$  with  $K$  the absolute constant given by Lemma 6.4 of [77]. The quantity  $\sigma^2$  satisfies in particular the two following assertions:

$$\forall B \in \mathcal{B}, \mathbb{P}_\pi[U \in B] \leq \sigma^2 \quad \text{and} \quad \sigma \geq K \sqrt{\mathcal{V}(1 + \ln(\sigma^{-1} \vee 1))} / n.$$

Indeed, if  $\sigma^2 = 2^{-(j+1)}$ , we have  $K^2 \mathcal{V}(1 + \ln(\sigma^{-1} \vee 1)) / n \leq K^2 \mathcal{V}(1 + \ln(2^{(j+1)/2})) / n \leq \sigma^2$ , or else if  $\sigma^2 = K^2 \mathcal{V}(1 + \frac{j+1}{2} \ln 2) / n$ , we have  $\sigma^{-1} \leq 2^{(j+1)/2}$  and so,

$$K^2 \mathcal{V}(1 + \ln(\sigma^{-1} \vee 1)) / n \leq K^2 \mathcal{V}(1 + \ln(2^{(j+1)/2})) / n = \sigma^2.$$

By applying Lemma 6.4 of [77], we obtain:

$$\begin{aligned}\mathbb{E}(\tilde{B}_S^1) &\leq \frac{K}{2}\sigma\sqrt{\mathcal{V}(1+|\ln\sigma|)} \\ &\leq \frac{K}{2}2^{-(j+1)/2}\sqrt{\mathcal{V}\left(1+\frac{j+1}{2}\ln 2\right)} + \frac{K}{2}K\mathcal{V}\left(1+\frac{j+1}{2}\ln 2\right)/\sqrt{n} \\ &\leq C\left\{2^{-j/2}\sqrt{j} + \frac{j}{\sqrt{n}}\right\},\end{aligned}$$

with  $C$  a positive absolute constant. So, with a similar argument for  $\tilde{B}_S^2$ , we obtain for any  $\lambda$  in  $\Gamma$

$$\mathbb{E}(\tilde{B}_S) \leq C\left\{\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}}\right\}.$$

Consequently,

$$\mathbb{E}(B_S) \leq C\left\{\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right\},$$

with  $C$  an absolute positive constant and from Markov's inequality, we have that for all  $\omega > 0$

$$\mathbb{P}\left(B_S > C(\omega)\left\{\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right\}\right) \leq e^{-\omega}. \quad (4.24)$$

Thus, by combining inequalities (4.18), (4.23) and (4.24), we obtain for all  $\omega > 0$ , with probability larger than  $1 - (1 + 2 \times 2.77)e^{-\omega}$ ,

$$\begin{aligned}f(U_1, \dots, U_n; m) \\ \leq \frac{C(\omega)}{n}\left\{\sqrt{m \ln(2/\alpha)\left(\frac{n}{T} + \frac{2^{-j}n^2}{T^2} + \frac{2^{-j/2}n}{T^{3/2}}\right)} + \ln(2/\alpha)\left(\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\}.\end{aligned}$$

Furthermore,  $N_{[-1;T+1]} \sim \mathcal{P}((T+2)\mu_c + n\|h\|_1)$ . Hence,

$$\mathbb{E}(N_{[-1;T+1]}) \leq C(\mu_c, R_1)(n+T).$$

From Markov's inequality, we have that for all  $\omega > 0$

$$\mathbb{P}(N_{[-1;T+1]} > C(\omega, \mu_c, R_1)(n+T)) \leq e^{-\omega}.$$

Then, we choose  $\omega$  such that this quantity  $(2 \times 2.77 + 2)e^{-\omega}$  is equal to  $\beta/2$ . So, with probability larger than  $1 - \beta/2$ ,

$$\begin{aligned}f(U_1, \dots, U_n; m) \\ \leq \frac{C(\beta, \mu_c, R_1)}{n}\left\{\sqrt{\ln(2/\alpha)}\sqrt{(n+T)\left(\frac{n}{T} + \frac{2^{-j}n^2}{T^2} + \frac{2^{-j/2}n}{T^{3/2}}\right)} + \ln(2/\alpha)\left(\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\} \\ \leq \frac{C(\beta, \mu_c, R_1)}{n}\left\{\sqrt{\ln(2/\alpha)}\sqrt{n + \frac{n^2}{T} + \frac{2^{-j}n^3}{T^2}} + \ln(2/\alpha)\left(\sqrt{j} + \frac{j2^{j/2}}{\sqrt{n}} + \frac{2^{-j/2}}{T}\right)\right\} \\ \leq C(\beta, \mu_c, R_1)\left\{\sqrt{\ln(2/\alpha)}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2}\sqrt{n}}{T}\right) + \ln(2/\alpha)\left(\frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT}\right)\right\}.\end{aligned}$$

Therefore by definition of  $q_{1-\beta/2}^\alpha$ ,

$$q_{1-\beta/2}^\alpha \leq C(\beta, \mu_c, R_1) \left\{ \sqrt{\ln(2/\alpha)} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} + \frac{2^{-j/2} \sqrt{n}}{T} \right) + \ln(2/\alpha) \left( \frac{\sqrt{j}}{n} + \frac{j2^{j/2}}{n^{3/2}} + \frac{2^{-j/2}}{nT} \right) \right\},$$

which concludes the proof of Lemma 4.2.

# Chapitre 5

## Lasso estimation for Poissonian interactions on the circle

This chapter proposes an extension of the Poissonian interactions model, introduced in Chapter 3, in a high-dimensional discrete setting. In this context, an adaptive Lasso-type procedure is proposed and we aim at obtaining oracle inequalities with large probability. For this purpose, Restricted Isometry Properties are provided and the Lasso procedure is calibrated by namely obtaining the convenient shape of the  $\ell_1$ -penalty weights. Finally, some immediate perspectives are given hoping to complete this work in progress.

*This chapter is a work in progress and is the fruit of a collaboration with Patricia Reynaud-Bouret (Université Nice Sophia-Antipolis), Vincent Rivoirard (Université Paris Dauphine) and Rebecca Willett (Duke University).*

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>128</b>
<b>5.2</b>	<b>Formulation of the problem</b>	<b>129</b>
5.2.1	Detailed description of the model	129
5.2.2	Related work	130
<b>5.3</b>	<b>Isometry properties</b>	<b>130</b>
<b>5.4</b>	<b>Lasso-type procedure</b>	<b>132</b>
5.4.1	Choice of the $d_k$ 's	132
5.4.2	The performance of the Lasso estimate	133
<b>5.5</b>	<b>Perspectives</b>	<b>134</b>
<b>5.6</b>	<b>Proofs</b>	<b>134</b>
5.6.1	Proof of Theorem 5.1	134
5.6.2	Proof of Theorem 5.2	136
5.6.3	Proof of Theorem 5.3	144
5.6.4	Proof of Theorem 5.4	147
5.6.5	Proof of Theorem 5.5	149

---



## 5.1 Introduction

In Chapter 3, we proposed a wavelet thresholding method to estimate interactions in a Poissonian interactions model, that is particularly suitable for the treatment of genomic data (see Section 3.4 of Chapter 3) with a fast and low computational algorithm (see Section 3.3 of Chapter 3). However, this method needs the knowledge of an a priori bound  $A$  on the support of interaction functions (see also [56, 93] for instance).

In this chapter, we aim at overcoming knowledge of  $A$  in a high-dimensional discrete setting, that can be viewed as a discrete version of the Poissonian interactions model that is in the heart of Chapter 3. The model is described as follows. Let  $U_1, \dots, U_n$  be a collection of  $n$  i.i.d. realizations of a uniform random variable on the set  $\{0, \dots, M-1\}$ ; these have to be thought of as points equally distributed on a circle. Considering a circular model allows us to avoid boundary effects. Furthermore, from a biological point of view, some genomes turn out to be circular (more precisely, contained in a circular DNA molecule), as prokaryote (typically bacteria and single-celled microorganisms). See [55, 106] for instance. Then, each  $U_i$  will give birth to some Poisson variables. If  $h^* = (h_0^*, \dots, h_{M-1}^*)^T$  is a vector in  $\mathbb{R}_+^M$ , then let us define  $N_{U_i+j}^i$  to be a Poisson variable with distribution  $\mathcal{P}(h_j^*)$  independent of anything else. The variable  $N_{U_i+j}^i$  represents the number of children that a certain individual  $U_i$  has at distance  $j$ . Here we understand  $U_i + j$  in a cyclic way, i.e. this is actually  $U_i + j$  modulus  $M$ .

This model can be reformulated with the introduction of a random matrix  $G$  depending on the  $U_i$ 's as we will see at Section 5.2. The aim is to recover  $h^*$ , with the observations of the  $U_i$ 's and the number of children at each position on the circle  $\{0, 1, \dots, M-1\}$ . More precisely, we are focussing on the problem of recovering an input vector  $h^* \in \mathbb{R}_+^M$  from corrupted measurements, which can be viewed as an inverse problem, potentially ill-posed (see Section 5.2 for more details). Finding sparse solutions to this problem is our aim. For this, we propose a nonparametric estimation method based on the Lasso, introduced by Tibshirani [105] (or a Dantzig selector [20]) that achieves sparsity of an estimated parameter vector via  $\ell_1$ -penalization. More precisely, our Lasso-type estimator requires the establishment of a type condition RIP (Restricted Isometry Property, introduced by Candès and Tao [20]) on the random matrix  $G$  with high probability (see Section 5.3 for more details), with non-constant data-driven weights (for instance, Zou [117] and Huang *et al.* [62] used these weights to obtain consistent and efficient Lasso estimators and Bertin *et al.* [9] and Hansen *et al.* [56] to control coefficient fluctuations).

The chapter is organized as follows. Section 5.2 describes the problem with the introduction of the sensing matrix  $G$  and Section 5.3 discusses the isometry properties and the RIP. Then, Section 5.4 presents the Lasso-type estimators with a convenient calibration of the weights, needed to obtain an oracle inequality with large probability. Finally, Section 5.5 gives immediate perspectives for this not achieved work and Section 5.6 is devoted to the proofs.

## 5.2 Formulation of the problem

### 5.2.1 Detailed description of the model

We observe at each position  $k$  between 0 and  $M - 1$  the total number of children regardless of who their parent might be. So, we set

$$Y_k = \sum_{i=1}^n N_k^i$$

and the problem can be translated as follows: we observe the  $U_i$ 's and the  $Y_k$ 's whose distribution conditioned on the  $U_i$ 's is given by

$$Y_k \sim \mathcal{P}\left(\sum_{i=1}^n h_{k-U_i}^*\right).$$

So, we may write  $Y \triangleq (Y_0, \dots, Y_{M-1})^T$  and we consider  $G \equiv G(U)$  the following  $M \times M$  circulant matrix:

$$G = \sum_{i=1}^n G_i, \quad \text{where } G_i \triangleq \begin{bmatrix} e_{U_i} & e_{U_i+1} & e_{U_i+2} & \cdots & e_{U_i+(M-1)} \end{bmatrix},$$

where the subscripts are understood to be modulus  $M$  and  $e_i$  denotes the  $i^{\text{th}}$  column of an  $M \times M$  identity matrix. Using this notation, we may write

$$Y \sim \mathcal{P}(Gh^*).$$

Estimating  $h^*$  is equivalent to an inverse problem where the operator is random and depends on the  $U_i$ 's. In particular, the sensing matrix  $G$  corresponds to a random convolution.

We consider the discrete Fourier transform and we introduce the matrix

$$F = \left( \frac{e^{-2\pi\mathbf{i}\ell m/M}}{\sqrt{M}} \right)_{\ell, m=0, \dots, M-1}.$$

Its Hermitian adjoint is

$$F^H = \left( \frac{e^{2\pi\mathbf{i}\ell m/M}}{\sqrt{M}} \right)_{\ell, m=0, \dots, M-1}$$

and so we have:  $F^H F = F F^H = \mathbf{I}_M$ . Since the  $G_i$ 's and  $G$  are circulant matrices, we can diagonalize these matrices as follows:

$$G_i = F^H \Sigma(U_i) F,$$

where

$$\Sigma(U_i) = \text{diag}(1, \dots, e^{-2\pi\mathbf{i}\ell U_i/M}, \dots, e^{-2\pi\mathbf{i}(M-1)U_i/M})$$

and

$$G = F^H \Sigma F,$$

where

$$\Sigma \triangleq \sum_{i=1}^n \Sigma_i = \text{diag}(\sigma_0, \dots, \sigma_{M-1}),$$

with

$$\sigma_k = \sum_{i=1}^n e^{-2\pi\mathbf{i}kU_i/M}.$$

Note that  $\sigma_0 = n$  and for  $k \neq 0[M]$ ,

$$\mathbb{E}(e^{-2\pi\mathbf{i}kU_i/M}) = \frac{1}{M} \sum_{u=0}^{M-1} e^{-2\pi\mathbf{i}ku/M} = \frac{1}{M} \frac{1 - e^{-2\pi\mathbf{i}kM/M}}{1 - e^{-2\pi\mathbf{i}k/M}} = 0.$$

Therefore, we obtain:

$$\mathbb{E}(\sigma_k) = \begin{cases} n & \text{if } k = 0[M] \\ 0 & \text{if } k \neq 0[M] \end{cases}.$$

Hence, since most of the eigenvalues are zero-mean, our problem is potentially ill-posed, which justifies the use of nonparametric procedures, such as Lasso, even if  $M$  is not large with respect to  $n$ .

### 5.2.2 Related work

Other authors have studied similar sensing matrices (albeit without considering Poisson noise), but those analyses do not extend to the problem considered here. For example, Candès and Plan [18] consider random convolutions in which they observe a random subset of elements of the product  $Ah^*$ , for a certain matrix  $A$  that needs to be an isometry. As we will see, here the matrix  $G$  is not an isometry and actually when  $n$  grows,  $G$  is closer and closer to a non invertible matrix. On the other hand, we do not observe just a random subset but the whole set of elements, in a noisy Poissonian way. In particular, the ratio of  $M$  (the total number of measurements) to  $n$  (the number of uniformly distributed trigger events) will play a crucial role in our analysis but is not explored in the existing literature. Other analyses of random convolution include [58, 76, 98], but these works consider specific generative models for the convolution kernel which are not compatible with the current setting.

## 5.3 Isometry properties

Ideally,  $G$  would satisfy an "isometry property" of the following form: with high probability,

$$r\|x\|_2^2 \leq \|Gx\|_2^2 \leq R\|x\|_2^2, \quad \forall x \in \mathbb{C}^M, \quad (5.1)$$

for some positive constants  $r$  and  $R$  satisfying  $r \approx R$ . Unfortunately, we cannot achieve (5.1), as proved by the following heuristic argument. First, note that

$$\|Gx\|_2^2 = x^H F^H \Sigma^H \Sigma F x = \sum_{k=0}^{M-1} |\sigma_k|^2 |(Fx)_k|^2.$$

Furthermore,  $|\sigma_0|^2 = n^2$  and for  $k \neq 0[M]$ ,

$$|\sigma_k|^2 = \sum_{1 \leq i, j \leq n} e^{2\pi\mathbf{i}k(U_i - U_j)/M} = n + \sum_{1 \leq i \neq j \leq n} e^{2\pi\mathbf{i}k(U_i - U_j)/M} = n + 2 \sum_{1 \leq i < j \leq n} \cos(2\pi k(U_i - U_j)/M)$$

and we obtain:

$$\mathbb{E}[|\sigma_k|^2] = \begin{cases} n^2 & \text{if } k = 0[M] \\ n & \text{if } k \neq 0[M] \end{cases}.$$

Note also that

$$\sum_{k=0}^{M-1} |(Fx)_k|^2 = \|x\|_2^2.$$

Hence we introduce

$$\tilde{\sigma}_k \triangleq \begin{cases} \sqrt{n} & \text{if } k = 0[M] \\ \sigma_k & \text{otherwise} \end{cases}$$

and

$$\|Gx\|_r^2 = \sum_{k=0}^{M-1} |\tilde{\sigma}_k|^2 |(Fx)_k|^2.$$

Consequently,

$$\|Gx\|_2^2 = (n^2 - n) \frac{(\mathbf{1}^T x)^2}{M} + \|Gx\|_r^2.$$

Hence, we are now interested in proving the analog of (5.1) with  $\|Gx\|_r^2$  instead of  $\|Gx\|_2^2$  and with  $r \approx R \approx n$ . The upper bound can be easily derived.

**Theorem 5.1.** *For all  $\delta \in ]0; 1[$ , with probability at least  $1 - \delta$ , for all  $x \in \mathbb{C}^M$ ,*

$$\|Gx\|_r^2 \leq n \left( 1 + \alpha + \frac{\beta}{n} \right) (\log(3(M-1)) - \log(\delta))^2 \|x\|_2^2,$$

where  $\alpha$  and  $\beta$  are absolute positive constants.

Unfortunately, a similar lower bound does not hold with high probability for all  $x \in \mathbb{C}^M$ . This is the reason why we will also derive a *Restricted Isometry Property* (RIP in the sequel) stating that there exist positive constants  $r$  and  $R$  such that with high probability, for any  $K$ -sparse vector  $x$  (i.e. with at the most  $K$  non-zero coordinates)

$$r\|x\|_2^2 \leq \|Gx\|_r^2 \leq R\|x\|_2^2,$$

with  $R$  as close to  $r$  as possible and of the order of  $n$ . More precisely, we have the following result.

**Theorem 5.2.** *Let  $\delta \in ]0; 1[$  and let  $K$  be an integer strictly smaller than  $M$ . Assume that*

$$c \triangleq \kappa \delta^{-2} \max \left\{ \frac{1}{n}, \frac{K^{3/4}}{M^{1/2}} \right\} K^2 (\log(M) + \log(n))^2 < 1, \quad (5.2)$$

for  $\kappa$  some absolute positive constant. Then with probability larger than  $1 - \delta$ , for any  $K$ -sparse vector  $x$ ,

$$(1 - c)n\|x\|_2^2 \leq \|Gx\|_r^2 \leq (1 + c)n\|x\|_2^2. \quad (5.3)$$

In particular, if  $K = O(M^\alpha)$  and  $K = O(n^\beta)$  when  $\min(M, n) \rightarrow +\infty$  with  $\alpha < 2/11$  and  $\beta < 1/2$ , then for any absolute constant  $c \in ]0; 1[$ , there exists  $\delta(n, M)$  that tends to 0 at a polynomial rate such that (5.3) holds with probability larger than  $1 - \delta(n, M)$ .

This property will enable us to derive oracle inequalities for our Lasso-type estimate. Before presenting our procedure and its theoretical results, let us say some words about the proof of the lower bound of Theorem 5.2. Our approach is similar to Rudelson and Vershynin's one [99] but controls on Rademacher processes are replaced with controls on suprema of  $U$ -statistics. See Section 5.6.2 for more details.

## 5.4 Lasso-type procedure

We first introduce the following Lasso estimate

$$\tilde{h}_L := \operatorname{argmin}_{h \in \mathbb{R}^M} \left\{ \|Y - Gh\|_{\ell_2}^2 + 2 \sum_{k=0}^{M-1} d_k |h_k| \right\},$$

which is based on random weights  $d_k$ . The weighted Lasso has already been used for instance by van de Geer *et al.* [111] with deterministic non-negative weights. Our approach where weights are random is similar to [9, 56, 62, 117]. In some sense, the weights play the same role as the thresholds in the estimation procedure proposed in Chapter 3 and also in [37, 65, 91, 92]. In the next subsection, we tune the  $d_k$ 's and give their order of magnitude.

For the moment, we are not able to prove oracle inequalities for  $\tilde{h}_L$ , so, for  $K_0$  a given positive integer, we introduce the following modified version of our Lasso estimate:

$$\hat{h}_L \equiv \hat{h}_L(K_0) := \operatorname{argmin}_{h \in \mathbb{R}^M \text{ } K_0\text{-sparse}} \left\{ \|Y - Gh\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k |h_k| \right\},$$

for some  $\eta > 0$ , which forces sparsity of the estimate. The integer  $K_0$ , that may depend on  $n$  and  $M$ , can be viewed as the a priori maximal length of the vector  $h^*$  known by the practitioner. Unfortunately  $\hat{h}_L$  cannot be implemented for large values of  $M$  in a low computational time. However it is likely that in practice  $\hat{h}_L = \tilde{h}_L$  with high probability but for the moment we are not able to prove such result.

### 5.4.1 Choice of the $d_k$ 's

The  $d_k$ 's are chosen so that they satisfy the following properties.

1. To be used in practice, they only depend on observable quantities.
2. They are as small as possible, which can be justified by practical considerations or by theoretical ones by using the shape of the oracle inequalities in Theorem 5.5.
3. The  $d_k$ 's need to control the deviations of centered quantities. More precisely, we need to control, with high probability, the event

$$|G^H(Y - Gh^*)|_k \leq d_k, \quad \forall k \in \{0, \dots, M-1\}.$$

See the proofs for more justifications. Consequently, under these three constraints, the  $d_k$ 's are non-constant data-driven weights in the Lasso-type procedure and the following result specifies a convenient choice for the  $d_k$ 's needed to obtain oracle inequalities with large probability.

**Theorem 5.3.** *Let  $\delta \in ]0; 1[$  and define for all  $k \in \{0, \dots, M-1\}$ ,*

$$d_k = \sqrt{2 \log(3/\delta) \tilde{V}_k} + \frac{1}{3} \log(3/\delta) B_k, \quad (5.4)$$

with

$$\tilde{V}_k = \hat{V}_k + \sqrt{2 \log(3/\delta) \hat{V}_k B_k^2} + 3 \log(3/\delta) B_k^2, \quad (5.5)$$

$$\hat{V}_k = \sum_{1 \leq i, j \leq n} \mathbf{1}_{U_i=U_j} Y_{k+U_i} \quad \text{and} \quad B_k = \sup_{t \in \mathcal{C}_{\{0, \dots, M-1\}}} \sum_{i=1}^n \mathbf{1}_{U_i+k=t},$$

where  $\mathcal{C}_{\{0, \dots, M-1\}}$  is the discrete circle on  $\{0, \dots, M-1\}$ . Then, with probability larger than  $1 - \delta$ ,

$$\forall k \in \{0, \dots, M-1\}, \quad |G^H(Y - Gh^*)|_k \leq d_k. \quad (5.6)$$

The order of magnitude of the  $d_k$ 's is given by the following result. More precisely, Theorem 5.4 provides a control of the moment of order 2 of the  $d_k$ 's.

**Theorem 5.4.** *For any  $\delta \in ]0; 1[$ ,*

$$\mathbb{E}(d_k^2) \leq C \left\{ \log(3/\delta) \left( n + \frac{n^3}{M^2} \right) + (\log(3/\delta))^2 \left( n + \frac{n^2}{M^2} \right) \right\},$$

where  $C$  is a positive constant depending on  $h_k^*$  and  $\|h^*\|_{\ell_1}$ .

The next subsection, where we establish oracle inequalities, shows, that if  $M$  is at least equal to  $n$  this order of magnitude leads to suitable theoretical results. Note that, to obtain this result, assumptions on  $K$  (and then on  $K_0$ ) are very mild.

### 5.4.2 The performance of the Lasso estimate

We establish the theoretical properties of  $\hat{h}_L := \hat{h}_L(K_0)$  by using the classical oracle approach. More precisely, we establish a bound on the risk of  $\hat{h}_L$  under some conditions.

**Theorem 5.5.** *Let  $\Omega$  be the event such that the lower bound of (5.3) is true with  $K \geq 2K_0$  and  $c < 1/2$  and such that (5.6) holds. It means that on  $\Omega$ , for all  $x$  at least  $2K_0$  sparse in  $\mathbb{C}^M$*

$$\|Gx\|_2^2 \geq \|Gx\|_r^2 \geq (1-c)n\|x\|_{\ell_2}^2 = r\|x\|_{\ell_2}^2$$

and

$$|G^H(Y - Gh^*)|_k \leq d_k \quad \forall k.$$

We have the following inequalities on the event  $\Omega$  (i.e. with probability at least  $1 - 2\delta$ ).

1. For the prediction-loss, for  $\eta \geq 0$  and for any  $0 < \alpha < 1$ , the estimate satisfies:

$$\|G\hat{h}_L - Gh^*\|_{\ell_2}^2 \leq \inf_{h: |S(h)| \leq K_0} \left\{ \frac{(1+\alpha)}{(1-\alpha)} \|Gh^* - Gh\|_{\ell_2}^2 + \frac{2(2+\eta)^2}{\alpha(1-\alpha)r} \sum_{k \in S(h)} d_k^2 \right\}, \quad (5.7)$$

where  $S(h)$  denotes the support of  $h$ . In particular, if  $h^*$  is  $K_0$ -sparse,

$$\|G\hat{h}_L - Gh^*\|_{\ell_2}^2 \leq \frac{8(2+\eta)^2}{r} \sum_{k \in S(h^*)} d_k^2, \quad (5.8)$$

2. For the  $\ell_1$ -loss, if  $\eta > 0$  and if  $h^*$  is  $K_0$ -sparse,

$$\|\hat{h}_L - h^*\|_{\ell_1} \leq \frac{(4+2\eta)^2}{\eta} \frac{|S(h^*)|}{r} \frac{\sup_{k \in S(h^*)} d_k^2}{\inf_k d_k}, \quad (5.9)$$

3. For the  $\ell_2$ -loss, if  $\eta > 0$  and if  $h^*$  is  $K_0$ -sparse,

$$\|\hat{h}_L - h^*\|_{\ell_2}^2 \leq \frac{8(2+\eta)^2}{r^2} \sum_{k \in S(h^*)} d_k^2. \quad (5.10)$$

By using Theorem 5.4, if  $r$  is of the order  $n$ , then the expectation of the right hand side of (5.10) is of order  $\frac{1}{n} + \frac{n}{M^2}$ , up to logarithmic terms. We observe that this rate is the same as the rate  $\frac{1}{n} + \frac{n}{T^2}$  exhibited in Chapter 3 with a wavelet thresholding estimation procedure since the asymptotic parameters  $M$  and  $T$  play the same role.

Here, we draw attention in the last inequality (5.10). Assume that  $M$  is at least equal to  $n$ . Theorem 5.2 shows that if  $K = O(M^\alpha)$  and  $K = O(n^\beta)$  when  $\min(M, n) \rightarrow +\infty$  with  $\alpha < 2/11$  and  $\beta < 1/2$ , then we can take  $\delta = M^{-\gamma}$  with  $\gamma > 0$  such that (5.3) holds with probability larger than  $1 - M^{-\gamma}$  and so that  $r$  is of the order  $n$ . Theorem 5.4 shows that the  $d_k$ 's are of the order  $\log M\sqrt{n}$  if  $n = O(M)$ . So the inequality (5.10) presents the classical right hand side of oracle inequalities, up to a logarithmic term, for Lasso estimates (see for instance [9, 16, 17, 110]). Indeed, the right hand side of (5.10) is of the order  $(\log M)^2/n$ . We mention that, in the case of  $n = O(M)$  and also  $n \geq (\log M)^2$ , we can obtain a small improvement of the order of magnitude of the  $d_k$ 's, namely they are upper bounded by  $\sqrt{n \log M} + (\log M)^2$ .

## 5.5 Perspectives

Of course, the paper associated to this chapter cannot be submitted in this shape. Before submission, we aim to deal with the following problems and extensions.

- We wish to lead a simulation study. Of course, for large values of  $M$ ,  $\hat{h}_L$  cannot be performed in a low computational time. So, we shall only carry out simulations for  $\tilde{h}_L$  and shall check that  $\tilde{h}_L$  is sparse.
- We wish to establish that with high probability, the true Lasso estimate  $\tilde{h}_L$  satisfies the  $K_0$  sparsity property if  $h^*$  is  $K_0$  sparse. For this purpose, we wish to mimick Bunea's proof (see [15]). If this result is valid, then  $\tilde{h}_L$  will also satisfy oracular properties.
- In general context, Restricted Isometry Properties provide oracle properties with high probability for Dantzig estimate (see [20]) defined by

$$\hat{h}_D := \operatorname{argmin}_{h \in \mathbb{R}^M} \{ \|h\|_{\ell_1} : \forall k \in \{0, \dots, M-1\}, |G^H(Y - Gh^*)|_k \leq d_k \}.$$

But for the moment, Theorem 5.2 gives such a result with  $\|\cdot\|_r$  instead of  $\|\cdot\|_2$ . We have to overcome this problem.

- From a genomic point of view, even if some organisms have circular genomes, the real 3D structure of the genome is not circular and consequently, the 3D distances are more influential. However, it could be that this kind of analysis can be generalized to more complex graphs than the circle, that would represent the real 3D structure of DNA. Furthermore, recent technological progresses could allow us to obtain such data for the study of genomes in three dimensions (see [101]).

## 5.6 Proofs

### 5.6.1 Proof of Theorem 5.1

To prove Theorem 5.1, it is sufficient to have an inequality for the deviation of  $|\sigma_k|^2$  around its mean for  $k \neq 0[M]$ .

**Lemma 5.1.** *There exists some absolute positive constants  $\alpha$  and  $\beta$  such that if  $k \neq 0[M]$*

$$\mathbb{P}(|\sigma_k|^2 \geq n(1 + \alpha z + \beta z/n)) \leq 3e^{-z},$$

for any  $z > 1$ .

*Proof.* It is sufficient to prove that there exist constants  $\alpha$  and  $\beta$  such that

$$W_n := \sum_{i < j} \cos(2\pi k(U_i - U_j)/M)$$

satisfies

$$\mathbb{P}(W_n \geq n/2(\alpha z + \beta z/n)) \leq 3e^{-z}.$$

But

$$\begin{aligned} \mathbb{E}(\cos(2\pi k(U_i - U_j)/M)|U_j) &= \mathbb{E}\left(\frac{1}{2} \left[ e^{2\pi \mathbf{i}k(U_i - U_j)/M} + e^{2\pi \mathbf{i}k(U_j - U_i)/M} \right] |U_j\right) \\ &= \frac{1}{2} \left[ \mathbb{E}\left(e^{2\pi \mathbf{i}kU_i/M}\right) e^{-2\pi \mathbf{i}kU_j/M} + e^{2\pi \mathbf{i}kU_j/M} \mathbb{E}\left(e^{-2\pi \mathbf{i}kU_i/M}\right) \right] \\ &= 0. \end{aligned}$$

Hence  $W_n$  is a completely degenerate U-statistics. Hence one can apply [61] (Theorem 3.4). It implies that there exists absolute constants  $c_1, c_2, c_3, c_4$  such that

$$\mathbb{P}(W_n \geq c_1 C\sqrt{z} + c_2 Dz + c_3 Bz^{3/2} + c_4 Az^2) \leq 3e^{-z},$$

with  $A = 1$  an upper bound on the infinite norm of  $\cos$ ,  $C^2 = \sum_{i < j} \mathbb{E}(\cos^2(2\pi k(U_i - U_j)/M)) \leq n(n-1)/2$  and  $D$  and  $B$  defined as follows.

$$D = \sup \left\{ \mathbb{E}\left(\sum_{i < j} \cos(2\pi k(U_i - U_j)/M) a_i(U_i) b_j(U_j)\right) / \mathbb{E}\left(\sum_i a_i(U_i)^2\right) \leq 1, \mathbb{E}\left(\sum_j b_j(U_j)^2\right) \leq 1 \right\}.$$

But

$$\begin{aligned} \mathbb{E}\left(\sum_{i < j} \cos(2\pi k(U_i - U_j)/M) a_i(U_i) b_j(U_j)\right) &\leq \mathbb{E}\left(\sum_{i < j} |a_i(U_i)| |b_j(U_j)|\right) \\ &\leq \mathbb{E}\left(\left(\sum_i |a_i(U_i)|\right) \left(\sum_j |b_j(U_j)|\right)\right) \\ &\leq \sqrt{\mathbb{E}\left(\left(\sum_i |a_i(U_i)|\right)^2\right) \mathbb{E}\left(\left(\sum_j |b_j(U_j)|\right)^2\right)} \\ &\leq n \sqrt{\mathbb{E}\left(\left(\sum_i |a_i(U_i)|^2\right)\right) \mathbb{E}\left(\left(\sum_j |b_j(U_j)|^2\right)\right)} \end{aligned}$$

Hence  $D \leq n$ . Finally

$$B^2 = \sup_u \sum_{j=2}^n \mathbb{E}(\cos^2(2\pi k(u - U_j)/M)) \leq n.$$



Hence we obtain that with probability larger than  $1 - 3e^{-z}$

$$W_n \leq n \left[ c_1 \sqrt{z} + c_2 z + c_3 \frac{1}{\sqrt{n}} z^{3/2} + c_4 \frac{1}{n} z^2 \right].$$

But for  $z > 1$ ,  $\sqrt{z} < z$  and we also have

$$2 \frac{1}{\sqrt{n}} z^{3/2} \leq z + z^2/n,$$

which gives the result. □

Now, we prove Theorem 5.1. From Lemma 5.1 and the union bound, we have that

$$\begin{aligned} & \mathbb{P}(\exists k \in \{1, \dots, M-1\} : |\sigma_k|^2 > n(1 + \alpha z + \beta z^2/n)) \\ & \leq \sum_{k=1}^{M-1} \mathbb{P}(|\sigma_k|^2 > n(1 + \alpha z + \beta z^2/n)) \\ & \leq 3(M-1)e^{-z}. \end{aligned}$$

In other words, with probability at least  $1 - \delta$ ,

$$|\tilde{\sigma}_k|^2 < R \quad \forall k$$

where

$$R = n(1 + \alpha + \beta/n)(\log(3(M-1)) - \log(\delta))^2.$$

Note that  $\alpha, \beta > 0$  and  $M-1 > \delta$ , which implies  $R \geq n$ .

### 5.6.2 Proof of Theorem 5.2

The proof is quite long and needs to be divided in 4 steps.

#### Step 1. Reduction of the problem to the control of a supremum of $U$ -statistics.

The idea is to mimic Rudelson and Vershynin's proof [99] directly on the matrix we are interested in. We focus on  $\|Gx\|_r^2$  when  $x$  is sparse with support  $S$  ( $S$ -sparse for short).

Generally, to any vector  $v$  of  $\mathbb{C}^M$ , we associate  $v^S$  the vector of  $\mathbb{C}^{|S|}$  whose coordinates are given by the coordinates of  $v$  with indices in  $S$ . In particular if  $x$  is  $S$  sparse,  $\|x\|_{\mathbb{C}^M} = \|x^S\|_{\mathbb{C}^{|S|}$ . The spaces  $\mathbb{C}^d$  or  $\mathbb{R}^k$  are put in indices in the norm and in the hermitian/scalar products when necessary to avoid confusion

We take in the sequel  $x$   $S$  sparse. Then, we have that

$$\|Gx\|_r^2 = \sum_{k=0}^{M-1} |\tilde{\sigma}_k|^2 |(Fx)_k|^2.$$

But denoting by  $A_j$  the  $j$ th column of a matrix  $A$

$$\begin{aligned} (Fx)_k &= \sum_{j=0}^{M-1} x_j F_j^{\{k\}} \\ &= (F_k^H)^H x \\ &= \langle F_k^H, x \rangle_{\mathbb{C}^M} \end{aligned}$$

Let us denote by  $z_k = F_k^H$  and by  $z_k^S$  its reduced form to  $S$ . Then

$$(Fx)_k = \langle z_k^S, x^S \rangle_{\mathbb{C}^{|S|}}.$$

Let us finally denote by  $z_k^S \otimes z_k^S$  the tensor product ie the matrix of  $\mathbb{C}^{|S|}$  equal to  $z_k^S (z_k^S)^H$ . Then

$$\|Gx\|_r^2 = (x^S)^H B_S x^S$$

with

$$B_S = \sum_{k=0}^{M-1} |\tilde{\sigma}_k|^2 z_k^S \otimes z_k^S$$

an hermitian matrix of  $\mathbb{C}^{|S|}$  with positive eigenvalues.

Note that for all  $S \subset \{0, \dots, M-1\}$

$$\sum_{k=0}^{M-1} z_k^S \otimes z_k^S = Id_{\mathbb{C}^{|S|}},$$

since the  $(i, j)$  term of this matrix is  $(1/M) \sum_k e^{\frac{2\pi\mathfrak{I}(i-j)k}{M}}$ .

Therefore for all  $S \subset \{0, \dots, M-1\}$ ,  $\mathbb{E}(B_S) = n Id_{\mathbb{C}^{|S|}}$ .

Now let us introduce the following supremum of operator norms, which controls the fluctuation of  $B_S$  around its mean

$$X = \sup_{S/|S| \leq K} \|Id_{\mathbb{C}^{|S|}} - \frac{1}{n} B_S\|_{op} = \sup_{S/|S| \leq K} \|Id_{\mathbb{C}^{|S|}} - \frac{1}{n} \sum_{k=0}^{M-1} |\tilde{\sigma}_k|^2 z_k^S \otimes z_k^S\|_{op} \quad (5.11)$$

Since the matrices  $Id_{\mathbb{C}^{|S|}} - \frac{1}{n} B_S$  are hermitian, their operator norm corresponds to their largest eigenvalue in absolute value (all their eigenvalues are real).

Next if we are able to control  $X$ , we will have on the event  $\{X < c\}$  with  $c < 1$  a fixed positive constant, that for all  $S$ , for all  $v \in \mathbb{C}^{|S|}$ ,

$$|v^H \left( Id_{\mathbb{C}^{|S|}} - \frac{1}{n} B_S \right) v| \leq c v^H v.$$

Consequently

$$(1+c)v^H v \geq \frac{1}{n} v^H B_S v \geq (1-c)v^H v.$$

It remains to apply this to  $v = x^S$  to get

$$\|Gx\|_r^2 = (x^S)^H B_S x^S \geq n(1-c) \|x^S\|_{\mathbb{C}^{|S|}}^2 = n(1-c) \|x\|_{\mathbb{C}^M}^2,$$

and

$$n(1+c) \|x\|_{\mathbb{C}^M}^2 \geq \|Gx\|_r^2.$$

So we will have the desired RIP with  $r = n(1-c)$  and  $R = n(1+c)$  on the event  $\{X < c\}$ .

So to prove that this holds with probability larger than  $1 - \delta$ , it remains to prove that  $\mathbb{P}(X \geq c)$  is smaller than  $\delta$ .

But since  $|\tilde{\sigma}_0|^2 = n$  and  $|\tilde{\sigma}_k|^2 = n + \sum_{i \neq j} \cos(2\pi k(U_i - U_j)/M)$

$$X = \sup_{S/|S| \leq K} \sup_{v \in \mathbb{C}^{|S|} / \|v\|_{\mathbb{C}^{|S|}} = 1} \left| \frac{1}{n} \sum_{k=1}^{M-1} \sum_{i \neq j} \cos(2\pi k(U_i - U_j)/M) v^H z_k^S \otimes z_k^S v \right|.$$

From now on we will consider vector with real coordinates and not complex anymore.

So let  $h(u, u')$  be vectors of  $\mathbb{R}^M$  such that  $h(u, u')_0 = 0$  and  $h(u, u')_k = \cos(2\pi k(u - u')/M)$  for  $k = 1, \dots, M - 1$ , for any  $u, u'$  in  $\{0, \dots, M - 1\}$ . For  $v \in \mathbb{C}^{|S|}$ , let  $\Psi_{v,S}$  be also a vector of  $\mathbb{R}^M$  defined by

$$(\Psi_{v,S})_k = v^H z_k^S \otimes z_k^S v = |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^2 \text{ for } k = 0, \dots, M - 1. \quad (5.12)$$

Then

$$X = \sup_{S/|S| \leq K} \sup_{v \in \mathbb{C}^{|S|} / \|v\|_{\mathbb{C}^{|S|}} = 1} \left| \frac{1}{n} \sum_{i \neq j} \langle h(U_i, U_j) | \Psi_{v,S} \rangle_{\mathbb{R}^M} \right|.$$

More generally we will study for any  $\Psi \in \mathbb{R}^M$

$$Z_\Psi = \frac{1}{n} \sum_{i \neq j} \langle h(U_i, U_j) | \Psi \rangle_{\mathbb{R}^M},$$

which is a completely degenerate U-statistics in the  $U_i$ 's and a linear form in  $\Psi$ , and

$$\mathcal{Z}_{\mathcal{F}} = \sup_{\Psi \in \mathcal{F}} |Z_\Psi|.$$

### Step 2. Study of one U-statistics.

We will need upper bounds on the moment of  $Z_\Psi$ . By Theorem 3.1.1 of Giné and de la Pena [32] (which decouples the U-statistics) and Theorem 3.3 of Giné, Latala and Zinn [48] (which gives the upper bound for decoupled U-statistics), we have that for any real number  $p \geq 2$ ,

$$\mathbb{E}(|Z_\Psi|^p) \leq \square^p [p^{p/2} C_\Psi^p + p^p D_\Psi^p + p^{3p/2} B_\Psi^p + p^{2p} A_\Psi^p], \quad (5.13)$$

where  $\square$  is an absolute positive constant and  $A_\Psi, B_\Psi, C_\Psi$  and  $D_\Psi$  are defined by

- $A_\Psi = \sup_{u, u'} \left| \frac{1}{n} \langle h(u, u') | \Psi \rangle \right|$
- $C_\Psi^2 = \sum_{i \neq j} \mathbb{E} \left[ \left( \frac{1}{n} \langle h(U_i, U_j) | \Psi \rangle \right)^2 \right]$ .
- $D_\Psi$  is the supremum of

$$\sum_{i \neq j} \mathbb{E} \left[ \frac{1}{n} \langle h(U_i, U_j) | \Psi \rangle a_i(U_i) b_j(U_j) \right]$$

for any choice of  $a_i, b_j$  such that  $\sum_i \mathbb{E}(a_i(U_i)^2) \leq 1$  and  $\sum_j \mathbb{E}(b_j(U_j)^2) \leq 1$ .

- $B_\Psi^2$  is the supremum of

$$\sum_{i=1}^{n-1} \mathbb{E} \left[ \left( \frac{1}{n} \langle h(U_i, u) | \Psi \rangle \right)^2 \right]$$

for any choice of  $u$  in  $\{0, \dots, M - 1\}$ .

Since  $\langle h(u, u') | \Psi \rangle = \sum_{k=1}^{M-1} \cos(2\pi k(u - u')/M) \Psi_k$ ,  $A_\Psi \leq \frac{1}{n} \|\Psi\|_1$ , where  $\|\Psi\|_1$  is the  $\ell^1$  norm of  $\Psi$  in  $\mathbb{R}^M$ . Next

$$\begin{aligned} C_\Psi^2 &\leq \mathbb{E} [\langle h(U_1, U_2) | \Psi \rangle^2] \\ &\leq \sum_{k,l=1}^{M-1} \mathbb{E} [\cos(2\pi k(U_1 - U_2)/M) \cos(2\pi l(U_1 - U_2)/M)] \Psi_k \Psi_l \end{aligned}$$

By linearizing the product of cosinus, one sees that its expectation is 0 except when  $k = l$  or  $l = M - k$ . Therefore

$$C_{\Psi}^2 \leq \sum_{k=1}^{M-1} [\Psi_k^2 + |\Psi_k| |\Psi_{M-k}|] \leq 2\|\Psi\|_2^2,$$

where  $\|\Psi\|_2$  is the  $\ell^2$  norm of  $\Psi$  in  $\mathbb{R}^M$ . For  $D_{\Psi}$ , note that

$$\begin{aligned} D_{\psi} &\leq \sqrt{\mathbb{E} \left[ \sum_i a_i (U_i)^2 \right]} \sqrt{\mathbb{E} \left[ \sum_i \left( \sum_{j=1, j \neq i}^n b_j (U_j) \frac{1}{n} < h(U_i, U_j) | \Psi > \right)^2 \right]} \\ &\leq \sqrt{\sum_i \sum_{j=1, j \neq i}^n \sum_{\ell=1, \ell \neq i}^n \mathbb{E} \left[ b_j (U_j)^2 \frac{1}{n^2} < h(U_i, U_{\ell}) | \Psi >^2 \right]} \end{aligned}$$

But

$$\begin{aligned} \mathbb{E} [ < h(U_i, U_{\ell}) | \Psi >^2 | U_j ] &= \sum_{k,l=1}^{M-1} \mathbb{E} [\cos(2\pi k(U_i - U_{\ell})/M) \cos(2\pi l(U_i - U_{\ell})/M)] \Psi_k \Psi_l \\ &= \frac{1}{4} \sum_{k,l=1}^{M-1} \mathbb{E} \left[ e^{\frac{2\pi i(k+l)(U_i - U_{\ell})}{M}} + e^{\frac{2\pi i(l-k)(U_i - U_{\ell})}{M}} + e^{\frac{2\pi i(k-l)(U_i - U_{\ell})}{M}} + e^{\frac{-2\pi i(k+l)(U_i - U_{\ell})}{M}} \right] \Psi_k \Psi_l, \end{aligned}$$

Since the expectation is null except if  $k = l$  or  $l = M - k$ , even if  $j = \ell$  (since  $j \neq i$ ), one has that

$$\mathbb{E} [ < h(U_i, U_{\ell}) | \Psi >^2 | U_j ] \leq 2\|\Psi\|_2^2 \quad (5.14)$$

and therefore

$$D_{\psi} \leq \sqrt{\sum_i \sum_{\ell=1, \ell \neq i}^n \frac{1}{n^2} \|\Psi\|_2^2 \mathbb{E} \left[ \sum_{j=1, j \neq i}^n b_j (U_j)^2 \right]} \leq \sqrt{2} \|\Psi\|_2.$$

Finally, (5.14) also shows that

$$B_{\Psi}^2 \leq \frac{1}{n} \mathbb{E} ( < h(U_i, u) | \Psi >^2 ) \leq 2 \frac{\|\Psi\|_2^2}{n}.$$

Hence (5.13) implies: for all  $p \geq 2$ ,

$$\mathbb{E} ( |Z_{\Psi}|^p ) \leq \square^p p^{2p} \max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)^p \quad (5.15)$$

where  $\square$  is a positive absolute constant which can change from line to line. We can therefore upper bound the Laplace transform of  $\sqrt{|Z_{\Psi}|}$ . Indeed for all  $\lambda > 0$ ,

$$\mathbb{E} \left[ \exp(\lambda \sqrt{|Z_{\Psi}|}) \right] = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbb{E} ( |Z_{\Psi}|^{k/2} ).$$

But for  $k \leq 4$ ,

$$\mathbb{E} ( |Z_{\Psi}|^{k/2} ) \leq \mathbb{E} ( |Z_{\Psi}|^2 )^{k/4} \leq \left( \square \max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right) \right)^{k/2}$$

and for larger  $k$  one can apply directly (5.15) with  $p = k/2$ . Consequently (5.15) holds for all  $p = k/2$  and all integer  $k \geq 0$ . Hence

$$\begin{aligned} \mathbb{E} \left[ \exp(\lambda \sqrt{|Z_\Psi|}) \right] &\leq \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \square^{k/2} (k/2)^k \max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)^{k/2} \\ &\leq \sum_{k=0}^{\infty} \frac{k^k}{k!} \left[ \square \lambda \sqrt{\max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)} \right]^k \end{aligned}$$

But by Stirling's formula  $k! \geq e^{-k} k^k$  and therefore for any  $\lambda$  such that

$$\square \lambda \sqrt{\max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)} \leq 1,$$

one has that

$$\mathbb{E} \left[ \exp(\lambda \sqrt{|Z_\Psi|}) \right] \leq \frac{1}{1 - \square \lambda \sqrt{\max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)}}.$$

We will denote by  $\|\Psi\|_{mix} = \max \left( \|\Psi\|_2, \frac{\|\Psi\|_1}{n} \right)$  which is a norm on  $\mathbb{R}^M$ .

**Step 3. Maximal inequalities for the  $Z_\Psi$ 's.**

This section mimics the results of Chapter 6 of [77] and [115] on Rademacher processes. Let us start first with  $\mathcal{A}$  a finite set of  $\mathbb{R}^M$  of cardinality  $N$ . We want to upper bound

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right].$$

By Jensen, for all  $\lambda$  such that

$$\square \lambda \sqrt{\max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}} \leq 1,$$

one has that

$$\begin{aligned} \exp \left( \lambda \mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right] \right) &\leq \mathbb{E} \left[ \exp \left( \lambda \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right) \right] \\ &\leq \sum_{\Psi \in \mathcal{A}} \mathbb{E} \left[ \exp(\lambda \sqrt{|Z_\Psi|}) \right] \\ &\leq N \frac{1}{1 - \square \lambda \sqrt{\max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}}} \\ &\leq N \exp \left( \frac{1}{1 - \square \lambda \sqrt{\max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}}} - 1 \right), \end{aligned}$$

since  $1 + x \leq \exp(x)$ . Let  $\omega = \square \sqrt{\max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}}$ . Hence

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right] \leq \frac{\log(N)}{\lambda} + \frac{\omega}{1 - \lambda \omega}.$$

It remains to optimize this in  $\lambda$ . If  $N = 1$ , one can choose  $\lambda \rightarrow 0$  and

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right] \leq \omega.$$

If  $N \geq 2$ , the best choice for  $\lambda$  is

$$\lambda = \frac{\log(N) - \sqrt{\log(N)}}{\omega[\log(N) - 1]}$$

which leads to

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right] \leq \omega[\log(N) - 1] \left( \frac{\log(N)}{\log(N) - \sqrt{\log(N)}} + \frac{1}{\sqrt{\log(N) - 1}} \right).$$

But the last term is smaller than an absolute constant and therefore, for all  $N$  we have

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{A}} \sqrt{|Z_\Psi|} \right] \leq \square \sqrt{\max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}} [\log(N) + 1]. \quad (5.16)$$

Now for any set  $\mathcal{F}$  and any  $\eta > 0$ , let us define a net of size  $\eta$  of  $\mathcal{F}$ , denoted  $\mathcal{F}_\eta$ , as a set of points of  $\mathcal{F}$  such that for any  $\Psi$  in  $\mathcal{F}$  there exists a  $\Psi_\eta$  of  $\mathcal{F}_\eta$  such that  $\|\Psi - \Psi_\eta\|_{mix} \leq \eta$ . Let us denote by  $H_{mix}(\eta, \mathcal{F}) = \inf_{\mathcal{F}_\eta} \log(|\mathcal{F}_\eta|)$ , which is decreasing in  $\eta$ , since any net of size  $\eta$  is also a net of size  $\eta' > \eta$ .

Assume that there exists

$$\xi \geq \max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix} \quad (5.17)$$

and assume temporarily that  $\mathcal{F}$  is finite.

Let us set for all integer  $j \geq 0$ ,  $\xi_j = \xi 2^{-j}$ . For  $j \geq 1$ , let  $\mathcal{F}_{\xi_j}$  be a net of size  $\xi_j$  of  $\mathcal{F}$  with minimal cardinality (this exists since  $\mathcal{F}$  is finite). Let  $\Pi_j$  be the application which associate to all  $\Psi \in \mathcal{F}$  its closest point in  $\mathcal{F}_{\xi_j}$ . Let also  $\Pi_0$  be the application which associate 0 to any  $\Psi$ . We have for all  $\Psi \in \mathcal{F}$  that  $\|\Psi - \Pi_j \Psi\|_{mix} \leq \xi_j$ , for all  $j \geq 0$ . We also have that for all  $j \geq 1$ ,  $\log(|\Pi_j(\mathcal{F})|) = H_{mix}(\xi_j, \mathcal{F})$  and  $\log(|\Pi_0(\mathcal{F})|) = 0$ .

Next since  $\mathcal{F}$  is finite, and  $Z_\Psi$  is linear in  $\Psi$ , there exists a maximal level  $J$  such that for all  $\Psi \in \mathcal{F}$ ,

$$Z_\Psi = \sum_{j=0}^J Z_{\Pi_{j+1}(\Psi)} - Z_{\Pi_j(\Psi)} = \sum_{j=0}^J Z_{\Pi_{j+1}(\Psi) - \Pi_j(\Psi)}.$$

Therefore

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_\Psi|} \right] \leq \sum_{j=0}^J \mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_{\Pi_{j+1}(\Psi) - \Pi_j(\Psi)}|} \right].$$

But we have at the most  $\exp(2H_{mix}(\xi_{j+1}, \mathcal{F}))$  points that are  $\Pi_{j+1}(\Psi) - \Pi_j(\Psi)$  and (5.16) tells us that

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_{\Pi_{j+1}(\Psi) - \Pi_j(\Psi)}|} \right] \leq \square [2H_{mix}(\xi_{j+1}, \mathcal{F}) + 1] \sqrt{\max_{\Psi \in \mathcal{F}} \|\Pi_{j+1}(\Psi) - \Pi_j(\Psi)\|_{mix}}.$$

But  $\|\cdot\|_{mix}$  is a norm. Hence

$$\begin{aligned} \mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_{\Pi_{j+1}(\Psi) - \Pi_j(\Psi)}|} \right] &\leq \square H_{mix}(\xi_{j+1}, \mathcal{F}) \sqrt{\max_{\Psi \in \mathcal{F}} \|\Pi_{j+1}(\Psi) - \Psi\|_{mix} + \|\Pi_j(\Psi) - \Psi\|_{mix}} \\ &\leq \square \sqrt{\xi} 2^{-j/2} H_{mix}(\xi_{j+1}, \mathcal{F}) \end{aligned}$$

Hence

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_\Psi|} \right] \leq \square \sqrt{\xi} \sum_{j=0}^J 2^{-j/2} H_{mix}(\xi_{j+1}, \mathcal{F}) \leq \square \sqrt{\xi} \sum_{j=0}^{\infty} 2^{-j/2} H_{mix}(\xi_{j+1}, \mathcal{F}).$$

But since the  $H_{mix}(\xi_{j+1}, \mathcal{F})$  increase when  $\mathcal{F}$  increase, this also holds for any countable set  $\mathcal{F}$ . Finally, by density, the following inequality also holds for any subset  $\mathcal{F}$  of  $\mathbb{R}^M$ , for all  $\xi$  satisfying (5.17),

$$\mathbb{E} \left[ \sup_{\Psi \in \mathcal{F}} \sqrt{|Z_{\Psi}|} \right] \leq \square \sqrt{\xi} \sum_{j=0}^{\infty} 2^{-j/2} H_{mix}(2^{-(j+1)}\xi, \mathcal{F}) \quad (5.18)$$

**Step 4. Application to the variable  $X$ .**

To compute a bound on  $\mathbb{E}(\sqrt{X})$  it remains to compute an upper bound of the entropy  $H_{mix}(\eta, \mathcal{F})$  where

$$\mathcal{F} = \left\{ \Psi_{v,S} \ / \ v \in \mathbb{C}^{|S|}, \|v\|_{\mathbb{C}^{|S|}} = 1, |S| \leq K \right\} = \cup_{S \ / \ |S| \leq K} \mathcal{F}^S,$$

with  $\mathcal{F}^S = \left\{ \Psi_{v,S} \ / \ v \in \mathbb{C}^{|S|}, \|v\|_{\mathbb{C}^{|S|}} = 1 \right\}$ . To have an upper bound it is sufficient to find a net of size  $\eta$  for each  $\mathcal{F}^S$  and their union will be a net of size  $\eta$  for  $\mathcal{F}$ . But for  $v, w$  in  $\mathbb{C}^{|S|}$ ,

$$\|\Psi_{v,S} - \Psi_{w,S}\|_{mix} = \max \left( \|\Psi_{v,S} - \Psi_{w,S}\|_2, \frac{1}{n} \|\Psi_{v,S} - \Psi_{w,S}\|_1 \right).$$

But we have

$$\begin{aligned} \|\Psi_{v,S} - \Psi_{w,S}\|_1 &= \sum_{k=0}^{M-1} \left| |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^2 - |\langle z_k^S, w \rangle_{\mathbb{C}^{|S|}}|^2 \right| \\ &\leq \sum_{k=0}^{M-1} |\langle z_k^S, v - w \rangle_{\mathbb{C}^{|S|}}| \left( |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}| + |\langle z_k^S, w \rangle_{\mathbb{C}^{|S|}}| \right) \end{aligned}$$

But since

$$\sum_{k=0}^{M-1} |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^2 = \|v\|_{\mathbb{C}^{|S|}}^2 = 1,$$

one has for all  $k$ ,  $|\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}| \leq 1$ . Hence

$$\begin{aligned} \|\Psi_{v,S} - \Psi_{w,S}\|_1 &\leq \frac{2}{\sqrt{M}} \sum_{k=0}^{M-1} \left| \sum_{j \in S} e^{2\pi i j k / M} (v - w)_j \right| \leq 2\sqrt{M} \sum_{j \in S} |(v - w)_j| \\ &\leq 2\sqrt{M|S|} \|v - w\|_{\mathbb{C}^{|S|}} \\ &\leq 2\sqrt{MK} \|v - w\|_{\mathbb{C}^{|S|}} \end{aligned}$$

We also have with similar computations

$$\begin{aligned} \|\Psi_{v,S} - \Psi_{w,S}\|_2^2 &= \sum_{k=0}^{M-1} \left[ |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^2 - |\langle z_k^S, w \rangle_{\mathbb{C}^{|S|}}|^2 \right]^2 \\ &\leq \sum_{k=0}^{M-1} |\langle z_k^S, v - w \rangle_{\mathbb{C}^{|S|}}|^2 \left( |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}| + |\langle z_k^S, w \rangle_{\mathbb{C}^{|S|}}| \right)^2 \\ &\leq 4 \sum_{k=0}^{M-1} |\langle z_k^S, v - w \rangle_{\mathbb{C}^{|S|}}|^2 \\ &= 4 \|v - w\|_{\mathbb{C}^{|S|}}^2. \end{aligned}$$

Therefore

$$\|\Psi_{v,S} - \Psi_{w,S}\|_{mix} \leq \max\left(2, \frac{2\sqrt{MK}}{n}\right) \|v - w\|_{\mathbb{C}^{|S|}}$$

If we want a net of size  $\eta$  for  $(\mathcal{F}_S, \|\cdot\|_{mix})$ , it is sufficient to have a net of size  $\epsilon = \eta \max\left(2, \frac{\sqrt{MK}}{n}\right)^{-1}$  of a sphere in  $(\mathbb{C}^{|S|}, \|\cdot\|_{\ell_2})$  which is of size less than  $\epsilon^{-|S|}$ .

So roughly speaking the entropy  $H_{mix}(\eta, \mathcal{F})$  is upper bounded by

$$\begin{aligned} \log\left(\sum_{s=0}^K \binom{M}{s} \max\left(2, \frac{\sqrt{MK}}{n}\right)^s \eta^{-s}\right) \\ \leq \log\left(K \binom{M}{K} \max\left(2, \frac{\sqrt{MK}}{n}\right)^K \eta^{-K}\right) \\ \leq \square K[\log(M) - \log(\eta)] \end{aligned}$$

Next one needs to find  $\xi \geq \max_{\Psi \in \mathcal{A}} \|\Psi\|_{mix}$ .

But for any unitary  $v \in \mathbb{C}^{|S|}$ ,

$$\|\Psi_{v,S}\|_1 = \sum_{k=0}^{M-1} |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^2 = 1.$$

and

$$\begin{aligned} \|\Psi_{v,S}\|_2^2 &= \sum_{k=0}^{M-1} |\langle z_k^S, v \rangle_{\mathbb{C}^{|S|}}|^4 \\ &= \sum_{k=0}^{M-1} \left| \sum_{j \in S} (z_k^{\{j\}})^H v_j \right|^4 \\ &= \frac{1}{M^2} \sum_{a,b,c,d \in S} \left( \sum_{k=0}^{M-1} e^{-\frac{2\pi k \mathfrak{I}(-a+b-c+d)}{M}} \right) v_a v_b^H v_c v_d^H. \end{aligned}$$

But

$$\left( \sum_{k=0}^{M-1} e^{-\frac{2\pi k \mathfrak{I}(-a+b-c+d)}{M}} \right) = 0$$

as soon as  $-a + b - c + d \neq 0[M]$ , and if  $-a + b - c + d = 0[M]$

$$\left( \sum_{k=0}^{M-1} e^{-\frac{2\pi k \mathfrak{I}(-a+b-c+d)}{M}} \right) = M.$$

But for  $a, b, c$  fixed less than  $M$  there is at the most 3 possible  $d$  such that  $-a + b - c + d = 0[M]$ , and  $|v_d| \leq \|v\|_{\mathbb{C}^{|S|}}^2 = 1$ . Therefore we obtain

$$\begin{aligned} \|\Psi_{v,S}\|_2^2 &\leq \frac{\square}{M} \sum_{a,b,c \in S} |v_a| |v_b| |v_c| \\ &\leq \frac{\square}{M} \left( \sum_{a \in S} |v_a| \right)^3 \end{aligned}$$



$$\begin{aligned} &\leq \frac{\square}{M} \sqrt{|S|}^3 \|v\|_{\mathcal{C}^{|S|}}^3 \\ &\leq \square \frac{K^{3/2}}{M} \end{aligned}$$

Therefore one can take

$$\xi = \square \max \left( \frac{K^{3/4}}{\sqrt{M}}, \frac{1}{n} \right),$$

and (5.18) tells us that

$$\begin{aligned} \mathbb{E}(\sqrt{X}) &\leq \square \max \left( \frac{K^{3/8}}{M^{1/4}}, \frac{1}{n^{1/2}} \right) \sum_{j=0}^{\infty} 2^{-j/2} K [\log(M) - \log(\xi) + (j+1) \log(2)] \\ &\leq \square \max \left( \frac{K^{3/8}}{M^{1/4}}, \frac{1}{n^{1/2}} \right) K [\log(M) + \log(n)]. \end{aligned} \quad (5.19)$$

Let  $\sqrt{\kappa}$  be the absolute constant  $\square$  in the previous equation. This defines the absolute constant  $\kappa$  of Theorem 5.2.

Therefore, with the choice of  $c$  in Theorem 5.2

$$\mathbb{P}(X > c) \leq \frac{\mathbb{E}(\sqrt{X})}{\sqrt{c}} \leq \delta,$$

and since  $c < 1$  by (5.2), we have on the event  $X \geq c$  that

$$(1-c)n\|x\|_2^2 \leq \|Gx\|_r^2 \leq (1+c)n\|x\|_2^2,$$

which concludes the proof of Theorem 5.2.

### 5.6.3 Proof of Theorem 5.3

Our problem, consisting in counting the total number of children at each position  $k$  between 0 and  $M-1$ , can be translated in another way by introducing the continuous circle  $\mathcal{C}_{[0;M]}$  and by considering the associated Poisson process. Indeed,  $U_1, \dots, U_n$  are  $n$  i.i.d. realizations of a uniform random variable on the set  $\{0, \dots, M-1\}$ . Then, since we work modulus  $M$ , each  $U_i$  gives birth independently to a Poisson process whose intensity is given by the function  $t \mapsto h(t - U_i)$ , where for any  $t \in \mathcal{C}_{[0;M]}$ ,

$$h(t) = \sum_{k=0}^{M-1} h_k^* \mathbf{1}_{[k; k+1[}(t).$$

We precise that the function  $h$  is defined modulus  $M$ . We denote  $\mathbf{1}^k = \mathbf{1}_{[k; k+1[}$  modulo  $M$  in the sequel. We observe the aggregated process  $N$  with intensity the function  $t \mapsto \sum_{i=1}^n h(t - U_i)$ .

Consequently, for any  $k \in \{0, \dots, M-1\}$ ,

$$Y_k = N_{[k; k+1[} \sim \mathcal{P} \left( \sum_{i=1}^n h_{k-U_i}^* \right)$$

and we can write

$$Y_k = \int_k^{k+1} dN_t = \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t) dN_t$$

and in particular  $\hat{V}_k$  can be rewritten as follows:

$$\begin{aligned}\hat{V}_k &= \sum_{1 \leq i, j \leq n} \mathbf{1}_{U_i=U_j} Y_{k+U_i} = \sum_{1 \leq i, j \leq n} \mathbf{1}_{U_i=U_j} \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_i) dN_t \\ &= \sum_{1 \leq i, j \leq n} \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_i) \mathbf{1}^k(t - U_j) dN_t \\ &= \int_{\mathcal{C}_{[0;M]}} \left( \sum_{i=1}^n \mathbf{1}^k(t - U_i) \right)^2 dN_t\end{aligned}$$

and  $B_k$  as follows:

$$B_k = \sup_{t \in \mathcal{C}_{\{0, \dots, M-1\}}} \sum_{i=1}^n \mathbf{1}_{U_i+k=t} = \sup_{t \in \mathcal{C}_{[0;M]}} \sum_{i=1}^n \mathbf{1}^k(t - U_i).$$

Since  $G = F^H \Sigma F$ , the following matrix products are involved in the computation of  $G^H Y$  and  $G^H G h^*$ :

- $F^H \Sigma^H$  is a  $M \times M$  matrix with  $(F^H \Sigma^H)_{kl} = \frac{e^{2\pi \mathfrak{i} k \ell / M}}{\sqrt{M}} \bar{\sigma}_\ell$  for any  $k, \ell \in \{0, \dots, M-1\}$ ,
- $F^H \Sigma^H \Sigma$  is a  $M \times M$  matrix with  $(F^H \Sigma^H \Sigma)_{kl} = \frac{e^{2\pi \mathfrak{i} k \ell / M}}{\sqrt{M}} |\sigma_\ell|^2$  for any  $k, \ell \in \{0, \dots, M-1\}$ ,
- $FY$  is a  $M \times 1$  matrix with  $(FY)_l = \sum_{m=0}^{M-1} \frac{e^{-2\pi \mathfrak{i} \ell m / M}}{\sqrt{M}} Y_m$  for any  $l \in \{0, \dots, M-1\}$ , and
- $Fh^*$  is a  $M \times 1$  matrix with  $(Fh^*)_l = \sum_{m=0}^{M-1} \frac{e^{-2\pi \mathfrak{i} \ell m / M}}{\sqrt{M}} h_m^*$  for any  $l \in \{0, \dots, M-1\}$ .

Thus, for any  $k \in \{0, \dots, M-1\}$ ,

$$(G^H Y)_k = (F^H \Sigma^H FY)_k = \frac{1}{M} \sum_{\ell, m=0}^{M-1} e^{2\pi \mathfrak{i} (k-\ell)m/M} \bar{\sigma}_m Y_\ell$$

and

$$(G^H G h^*)_k = (F^H \Sigma^H F F^H \Sigma F h^*)_k = (F^H \Sigma^H \Sigma F h^*)_k = \frac{1}{M} \sum_{\ell, m=0}^{M-1} e^{2\pi \mathfrak{i} (k-\ell)m/M} |\sigma_m|^2 h_\ell^*,$$

since  $F$  is a unitary matrix.

Therefore, for any  $k \in \{0, \dots, M-1\}$ , we have:

$$\begin{aligned}(G^H Y)_k &= \frac{1}{M} \sum_{\ell, m=0}^{M-1} e^{2\pi \mathfrak{i} (k-\ell)m/M} \sum_{i=1}^n e^{2\pi \mathfrak{i} m U_i / M} Y_\ell \\ &= \frac{1}{M} \sum_{i=1}^n \sum_{\ell=0}^{M-1} Y_\ell \sum_{m=0}^{M-1} e^{2\pi \mathfrak{i} (k-\ell+U_i)m/M} \\ &= \frac{1}{M} \sum_{i=1}^n \sum_{\ell=0}^{M-1} Y_\ell \times \begin{cases} M & \text{if } k - \ell + U_i = 0 \\ 0 & \text{if } k - \ell + U_i \neq 0 \end{cases} \\ &= \sum_{i=1}^n \sum_{\ell=0}^{M-1} Y_\ell \mathbf{1}_{\ell=k+U_i}\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n Y_{k+U_i} \\
&= \sum_{i=1}^n \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^{k+U_i}(t) dN_t \\
&= \int_{\mathcal{C}_{[0;M]}} \sum_{i=1}^n \mathbf{1}^k(t - U_i) dN_t
\end{aligned}$$

and

$$\begin{aligned}
(G^H G h^*)_k &= \frac{1}{M} \sum_{\ell, m=0}^{M-1} e^{2\pi i(k-\ell)m/M} \sum_{i, j=1}^n e^{2\pi i m(U_i - U_j)/M} h_\ell^* \\
&= \frac{1}{M} \sum_{i, j=1}^n \sum_{\ell=0}^{M-1} h_\ell^* \sum_{m=0}^{M-1} e^{2\pi i(k-\ell+U_i-U_j)m/M} \\
&= \frac{1}{M} \sum_{i, j=1}^n \sum_{\ell=0}^{M-1} h_\ell^* \times \begin{cases} M & \text{if } k - \ell + U_i - U_j = 0 \\ 0 & \text{if } k - \ell + U_i - U_j \neq 0 \end{cases} \\
&= \sum_{i, j=1}^n \sum_{\ell=0}^{M-1} h_\ell^* \mathbf{1}_{\ell=k+U_i-U_j} \\
&= \sum_{i, j=1}^n h_{k+U_i-U_j}^* \\
&= \sum_{i, j=1}^n \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^{k+U_i-U_j}(t) h(t) dt \\
&= \sum_{i, j=1}^n \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^{k+U_i}(t) h(t - U_j) dt \\
&= \int_{\mathcal{C}_{[0;M]}} \sum_{i=1}^n \mathbf{1}^k(t - U_i) \sum_{j=1}^n h(t - U_j) dt.
\end{aligned}$$

We notice that

$$\mathbb{E}((G^H Y)_k | U_1, \dots, U_n) = \int_{\mathcal{C}_{[0;M]}} \sum_{i=1}^n \mathbf{1}^k(t - U_i) \sum_{j=1}^n h(t - U_j) dt = (G^H G h^*)_k.$$

Let  $k \in \{0, \dots, M-1\}$  be fixed. We apply Lemma 6.1 of [91]: for any  $x > 0$ , conditionally on the  $U_i$ 's, we have

$$\mathbb{P} \left( |G^H(Y - G h^*)|_k \geq \sqrt{2xV_k} + \frac{xB_k}{3} \right) \leq 2e^{-x},$$

with

$$V_k = \text{Var}((G^H Y)_k | U_1, \dots, U_n).$$

Moreover, by fixing  $x > 0$  and defining

$$\tilde{V}_k = \hat{V}_k + \sqrt{2x\hat{V}_k B_k^2} + 3xB_k^2,$$

we have also

$$\mathbb{P}(V_k \geq \tilde{V}_k) \leq e^{-x}.$$

Finally, with  $\tilde{V}_k$  defined by (5.5) and with probability larger than  $1 - \delta$ ,

$$|G^H(Y - Gh^*)|_k \leq \sqrt{2 \log(3/\delta) \tilde{V}_k} + \frac{1}{3} \log(3/\delta) B_k = d_k,$$

that concludes the proof.

#### 5.6.4 Proof of Theorem 5.4

In this section, we keep the same notations introduced in Section 5.6.3.

Let  $k \in \{0, \dots, M-1\}$ . By definition (5.4) of  $d_k$ , we have:

$$\mathbb{E}(d_k^2) \leq 2 \left( 2 \log(3/\delta) \mathbb{E}(\tilde{V}_k) + \frac{1}{9} (\log(3/\delta))^2 \mathbb{E}(B_k^2) \right),$$

with

$$\mathbb{E}(\tilde{V}_k) \leq \mathbb{E}(\hat{V}_k) + \sqrt{2 \log(3/\delta) \mathbb{E}(\hat{V}_k) \mathbb{E}(B_k^2) + 3 \log(3/\delta) \mathbb{E}(B_k^2)}.$$

It remains to control  $\mathbb{E}(\hat{V}_k)$  and  $\mathbb{E}(B_k^2)$ .

On the one hand,

$$\mathbb{E}(\hat{V}_k) = \mathbb{E}(V_k),$$

where  $V_k = \mathbb{E}(\hat{V}_k | U_1, \dots, U_n) (= \text{Var}((G^H Y)_k | U_1, \dots, U_n))$ . We have

$$\begin{aligned} V_k &= \int_{\mathcal{C}_{[0;M]}} \left( \sum_{i=1}^n \mathbf{1}^k(t - U_i) \right)^2 \sum_{j=1}^n h(t - U_j) dt \\ &= \int_{\mathcal{C}_{[0;M]}} \sum_{j=1}^n \left( \mathbf{1}^k(t - U_j) + \sum_{i \neq j} \mathbf{1}^k(t - U_i) \right)^2 h(t - U_j) dt \\ &= \sum_{j=1}^n \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_j) h(t - U_j) dt + 2 \sum_{j=1}^n \int_{\mathcal{C}_{[0;M]}} \sum_{i \neq j} \mathbf{1}^k(t - U_i) \mathbf{1}^k(t - U_j) h(t - U_j) dt \\ &\quad + \sum_{j=1}^n \int_{\mathcal{C}_{[0;M]}} \sum_{i \neq j} \sum_{i' \neq j} \mathbf{1}^k(t - U_i) \mathbf{1}^k(t - U_{i'}) h(t - U_j) dt \\ &= n h_k^* + 2 \sum_{i \neq j} \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_i) \mathbf{1}^k(t - U_j) h(t - U_j) dt + \sum_{i \neq j} \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_i) h(t - U_j) dt \\ &\quad + \sum_{i \neq i' \neq j} \int_{\mathcal{C}_{[0;M]}} \mathbf{1}^k(t - U_i) \mathbf{1}^k(t - U_{i'}) h(t - U_j) dt. \end{aligned}$$

We observe, with  $U$  a uniform random variable on the set  $\{0, \dots, M-1\}$ , that

- $\mathbb{E}[\mathbf{1}^k(t - U)] = \frac{1}{M} \sum_{u=0}^{M-1} \mathbf{1}_{k \leq t-u < k+1} = \frac{1}{M},$
- $\mathbb{E}[h(t - U)] = \frac{1}{M} \sum_{u=0}^{M-1} h(t - u) = \frac{1}{M} \sum_{u=0}^{M-1} \sum_{\ell=0}^{M-1} h_\ell^* \mathbf{1}_{\ell \leq t-u < \ell+1} = \frac{1}{M} \sum_{\ell=0}^{M-1} h_\ell^*,$  and
- $\mathbb{E}[\mathbf{1}^k(t - U) h(t - U)] = \frac{1}{M} \sum_{u=0}^{M-1} \mathbf{1}_{k \leq t-u < k+1} h(t - u) = \frac{h_k^*}{M} \sum_{u=0}^{M-1} \mathbf{1}_{k \leq t-u < k+1} = \frac{h_k^*}{M}.$

Consequently, by denoting  $\|h^*\|_{\ell_1} = \sum_{\ell=0}^{M-1} h_\ell^*$ ,

$$\begin{aligned}
\mathbb{E}(\hat{V}_k) &= \mathbb{E}(V_k) = nh_k^* + 2n(n-1) \int_{\mathcal{C}_{[0;M]}} \mathbb{E}(\mathbf{1}^k(t-U_1)) \mathbb{E}(\mathbf{1}^k(t-U_1)h(t-U_1)) dt \\
&\quad + n(n-1) \int_{\mathcal{C}_{[0;M]}} \mathbb{E}(\mathbf{1}^k(t-U_1)) \mathbb{E}(h(t-U_1)) dt \\
&\quad + n(n-1)(n-2) \int_{\mathcal{C}_{[0;M]}} \mathbb{E}(\mathbf{1}^k(t-U_1)) \mathbb{E}(\mathbf{1}^k(t-U_1)) \mathbb{E}(h(t-U_1)) dt \\
&= nh_k^* + 2n(n-1) \frac{h_k^*}{M} + n(n-1) \frac{\|h^*\|_{\ell_1}}{M} + n(n-1)(n-2) \frac{\|h^*\|_{\ell_1}}{M^2} \\
&\leq \square \left\{ nh_k^* + \frac{n^2 h_k^*}{M} + \frac{n^2 \|h^*\|_{\ell_1}}{M} + \frac{n^3 \|h^*\|_{\ell_1}}{M^2} \right\} \\
&\leq C(h_k^*, \|h^*\|_{\ell_1}) \left\{ n + \frac{n^3}{M^2} \right\}, \tag{5.20}
\end{aligned}$$

with  $C(h_k^*, \|h^*\|_{\ell_1})$  a positive constant depending on  $h_k^*$  and  $\|h^*\|_{\ell_1}$ .

On the other hand,

$$\begin{aligned}
B_k &= \sup_{t \in \mathcal{C}_{[0;M]}} \sum_{i=1}^n \mathbf{1}^k(t-U_i) \\
&\leq \sup_{t \in \mathcal{C}_{[0;M]}} \left| \sum_{i=1}^n \left( \mathbf{1}^k(t-U_i) - \mathbb{E}[\mathbf{1}^k(t-U_1)] \right) \right| + n \sup_{t \in \mathcal{C}_{[0;M]}} \left| \mathbb{E}[\mathbf{1}^k(t-U_1)] \right| \\
&\leq \tilde{B}_k + \frac{n}{M},
\end{aligned}$$

with

$$\begin{aligned}
\tilde{B}_k &= \sup_{t \in \mathcal{C}_{[0;M]}} \left| \sum_{i=1}^n \left( \mathbf{1}^k(t-U_i) - \mathbb{E}[\mathbf{1}^k(t-U_1)] \right) \right| \\
&= \sup_{I_t, t \in \mathcal{C}_{[0;M]}} \left| \sum_{i=1}^n \left( \mathbf{1}_{I_t}(U_i) - \mathbb{E}[\mathbf{1}_{I_t}(U_1)] \right) \right|,
\end{aligned}$$

where for any  $t \in \mathcal{C}_{[0;M]}$ ,  $I_t = ]t-k-1; t-k]$ . We set  $\mathcal{B} = \{I_t, t \in \mathcal{C}_{[0;M]}\}$  and for every integer  $n$ ,  $m_n(\mathcal{B}) = \sup_{A \subset \mathcal{C}_{[0;M]}, |A|=n} |\{A \cap I_t, t \in \mathcal{C}_{[0;M]}\}|$ . It is easy to see that

$$m_n(\mathcal{B}) \leq 1 + \frac{n(n+1)}{2}$$

and so, the VC-dimension  $\mathcal{V}$  of  $\mathcal{B}$  defined by  $\sup\{n \geq 0, m_n(\mathcal{B}) = 2^n\}$  is bounded by 2 (see Definition 6.2 of [77]). By applying Lemma 6.4 of [77] (we fix  $\sigma^2 = 1$  in the use of this property), we obtain:

$$\mathbb{E}(\tilde{B}_k) \leq \frac{K\sqrt{2}}{2} \sqrt{n},$$

where  $K$  is an absolute positive constant. But, we want an upper bound of  $\mathbb{E}(\tilde{B}_k^2)$ . For this, we use Theorem 11 of [12]:

$$\left[ \mathbb{E}(\tilde{B}_k^2) \right]^{1/2} \leq \square \left\{ \mathbb{E}(\tilde{B}_k) + (\mathbb{E}[\mathcal{M}^2])^{1/2} \right\},$$

where

$$\mathcal{M} = \max_{1 \leq i \leq n} \sup_{t \in \mathcal{C}_{[0;M]}} |\mathbf{1}^k(t - U_i) - \mathbb{E}(\mathbf{1}^k(t - U_1))| \quad \text{and so} \quad \mathbb{E}[\mathcal{M}^2] \leq 4.$$

Consequently,

$$\mathbb{E}(B_k^2) \leq \square \left\{ n + \frac{n^2}{M^2} \right\}. \quad (5.21)$$

Finally, by combining (5.20) and (5.21), we obtain:

$$\begin{aligned} \mathbb{E}(d_k^2) &\leq \square \left( \log(3/\delta) \mathbb{E}(\hat{V}_k) + \log(3/\delta) \sqrt{\log(3/\delta) \mathbb{E}(\hat{V}_k) \mathbb{E}(B_k^2)} + (\log(3/\delta))^2 \mathbb{E}(B_k^2) \right) \\ &\leq C(h_k^*, \|h^*\|_{\ell_1}) \left( \log(3/\delta) \left\{ n + \frac{n^3}{M^2} \right\} + (\log(3/\delta))^2 \left\{ n + \frac{n^2}{M^2} \right\} \right), \end{aligned}$$

that concludes the proof of Theorem 5.4.

### 5.6.5 Proof of Theorem 5.5

For short, we denote  $\hat{h} = \hat{h}_L$ . We first establish the following lemma.

**Lemma 5.2.** *For any  $h$ , if  $S(h) = \{k : h_k \neq 0\}$  satisfies  $|S(h)| \leq K_0$ , we have:*

$$\eta \sum_{k \notin S(h)} d_k |\hat{h}_k - h_k| \leq \|Gh - Gh^*\|_{\ell_2}^2 + (4 + \eta) \sum_{k \in S(h)} d_k |\hat{h}_k - h_k|, \quad (5.22)$$

$$\|G\hat{h} - Gh^*\|_{\ell_2}^2 \leq \|Gh - Gh^*\|_{\ell_2}^2 + (4 + 2\eta) \sum_{k \in S(h)} d_k |\hat{h}_k - h_k|. \quad (5.23)$$

*Proof.* We have for any  $h$   $K_0$ -sparse,

$$\|Y - G\hat{h}\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k |\hat{h}_k| \leq \|Y - Gh\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k |h_k|.$$

And using standard computations, we obtain:

$$\begin{aligned} \|G\hat{h} - Gh^*\|_{\ell_2}^2 + \eta \sum_{k=0}^{M-1} d_k |\hat{h}_k - h_k| &\leq \|Gh - Gh^*\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k (|h_k| - |\hat{h}_k|) \\ &\quad + \eta \sum_{k=0}^{M-1} d_k |\hat{h}_k - h_k| + 2 \langle G^H(Y - Gh^*), \hat{h} - h \rangle \\ &\leq \|Gh - Gh^*\|_{\ell_2}^2 + (2 + \eta) \sum_{k=0}^{M-1} d_k (|h_k| - |\hat{h}_k| + |h_k - \hat{h}_k|) \\ &\leq \|Gh - Gh^*\|_{\ell_2}^2 + (4 + 2\eta) \sum_{k \in S(h)} d_k |h_k - \hat{h}_k|, \end{aligned}$$

which leads to (5.22) and (5.23).  $\square$

Still by using the constant  $r$  corresponding to a sparsity degree  $K = 2K_0$ , we have for any  $h$  such that  $|S(h)| \leq K_0$ :

$$\left( \sum_{k \in S(h)} |\hat{h}_k - h_k|^2 \right)^{1/2} \leq \|\hat{h} - h\|_{\ell_2}$$

$$\begin{aligned} &\leq r^{-1/2} \|G\hat{h} - Gh\|_{\ell_2} \\ &\leq r^{-1/2} (\|G\hat{h} - Gh^*\|_{\ell_2} + \|Gh - Gh^*\|_{\ell_2}). \end{aligned}$$

Therefore, using (5.23),

$$\begin{aligned} &\|G\hat{h} - Gh^*\|_{\ell_2}^2 \\ &\leq \|Gh - Gh^*\|_{\ell_2}^2 + (4 + 2\eta) \left( \sum_{k \in S(h)} d_k^2 \right)^{1/2} \left( \sum_{k \in S(h)} |\hat{h}_k - h_k|^2 \right)^{1/2} \\ &\leq \|Gh - Gh^*\|_{\ell_2}^2 + (4 + 2\eta) r^{-1/2} \left( \sum_{k \in S(h)} d_k^2 \right)^{1/2} (\|G\hat{h} - Gh^*\|_{\ell_2} + \|Gh - Gh^*\|_{\ell_2}) \\ &\leq \|Gh - Gh^*\|_{\ell_2}^2 + \alpha \|G\hat{h} - Gh^*\|_{\ell_2}^2 + \alpha \|Gh - Gh^*\|_{\ell_2}^2 \\ &\quad + 2\alpha^{-1} (2 + \eta)^2 r^{-1} \sum_{k \in S(h)} d_k^2 \\ &\leq \frac{(1 + \alpha)}{(1 - \alpha)} \|Gh^* - Gh\|_{\ell_2}^2 + \frac{2(2 + \eta)^2}{\alpha(1 - \alpha)r} \sum_{k \in S(h)} d_k^2, \end{aligned}$$

This leads to (5.7). Now, assume that  $h^*$  is  $K_0$ -sparse. Then, by using (5.22) with  $h = h^*$ ,

$$\begin{aligned} \inf_k d_k \sum_{k=0}^{M-1} |\hat{h}_k - h_k^*| &\leq \sum_{k \notin S(h^*)} d_k |\hat{h}_k - h_k^*| + \sum_{k \in S(h^*)} d_k |\hat{h}_k - h_k^*| \\ &\leq \frac{(4 + 2\eta)}{\eta} \sum_{k \in S(h^*)} d_k |\hat{h}_k - h_k^*| \\ &\leq \frac{(4 + 2\eta)}{\eta} \sup_{k \in S(h^*)} d_k \sum_{k \in S(h^*)} |\hat{h}_k - h_k^*| \\ &\leq \frac{(4 + 2\eta)}{\eta} \left( \sup_{k \in S(h^*)} d_k \right) \sqrt{|S(h^*)|} \left( \sum_{k \in S(h^*)} |\hat{h}_k - h_k^*|^2 \right)^{1/2}. \end{aligned}$$

But, by using (5.23) with  $h = h^*$ ,

$$\begin{aligned} r \sum_{k \in S(h^*)} |\hat{h}_k - h_k^*|^2 &\leq r \|\hat{h} - h^*\|_{\ell_2}^2 \\ &\leq \|G\hat{h} - Gh^*\|_{\ell_2}^2 \\ &\leq (4 + 2\eta) \sum_{k \in S(h^*)} d_k |\hat{h}_k - h_k^*| \\ &\leq (4 + 2\eta) \left( \sum_{k \in S(h^*)} d_k^2 \right)^{1/2} \left( \sum_{k \in S(h^*)} |\hat{h}_k - h_k^*|^2 \right)^{1/2} \end{aligned}$$

and

$$\left( \sum_{k \in S(h^*)} |\hat{h}_k - h_k^*|^2 \right)^{1/2} \leq \frac{4 + 2\eta}{r} \left( \sum_{k \in S(h^*)} d_k^2 \right)^{1/2}.$$

We finally obtain:

$$\inf_k d_k \sum_{k=0}^{M-1} |\hat{h}_k - h_k^*| \leq \frac{(4 + 2\eta)^2}{\eta r} \sqrt{|S(h^*)|} \left( \sum_{k \in S(h^*)} d_k^2 \right)^{1/2} \left( \sup_{k \in S(h^*)} d_k \right),$$

This leads to (5.9). Finally, (5.10) is obtained by using (5.8) with  $h = h^*$ .





# Chapitre A

## Programs coded in Scilab

This appendix presents the programs coded in Scilab 5.2 that allowed us to provide the simulation study of Chapters 3 and 4.

### Contents

---

<b>A.1 Implementation of the estimation procedure of Chapter 3 . . . . .</b>	<b>153</b>
A.1.1 Preliminary functions . . . . .	154
A.1.2 Computation of $S(\varphi_\lambda)$ and $B(\varphi_\lambda)$ . . . . .	155
A.1.3 Cascade algorithm for the Haar basis . . . . .	158
<b>A.2 Implementation of the testing procedure of Chapter 4 . . . . .</b>	<b>160</b>
A.2.1 Simulation of the data . . . . .	160
A.2.2 Cascade algorithm and functions . . . . .	160
A.2.3 Computation of the empirical rate of type I error . . . . .	163

---

### A.1 Implementation of the estimation procedure of Chapter 3

The goal of this section is the computation of the quantity

$$S(\varphi_\lambda)(t) = \sum_{i=1}^n \left[ \varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right], \quad \text{for all } t \in \mathbb{R}$$

defined by (3.15), via a cascade algorithm, in order to implement the wavelet coefficient estimates  $\hat{\beta}_\lambda = \sum_{V \in \mathcal{N}} \frac{1}{n} S(\varphi_\lambda)(V)$  and the following quantities that appear in the wavelet thresholding procedure:

- $B(\varphi_\lambda) = \sup_{t \in \mathbb{R}} S(\varphi_\lambda)(t)$ , and
- $\hat{V}(\varphi_\lambda) = \sum_{V \in \mathcal{N}} \frac{1}{n} (S(\varphi_\lambda)(V))^2$ , an estimator of the variance of  $n\hat{\beta}_\lambda$ .

We recall that  $S(\varphi_\lambda)$  is the sum of a random "piecewise constant" part  $S_r(\varphi_\lambda) = \sum_{i=1}^n \varphi_\lambda(\cdot - U_i)$  and a deterministic (piecewise affine) part  $(n-1)\mathbb{E}_\pi(\varphi_\lambda(\cdot - U))$ .

### A.1.1 Preliminary functions

The function `sumpwc` corresponds to the sum of two piecewise constant functions defined by their partitions and the values on its respective partition and the function `evalpwc` evaluates a piecewise constant function in a given point.

```
function y=sumpwc(m1, valm1, m2, valm2)

// m1, m2 are partitions
// valm1, valm2 are the values on the partitions m1 and m2 respectively

n1=size(m1, '*');
// Initialization of the partition
[m, k]=gsort([m1 m2], 'g', 'i');
// Initialization of the values on the partition
valm=zeros(m);

// Computation of the first coefficient
if k(1)<=n1 then
    valm(1)=valm1(1); valm1current=valm1(1); valm2current=0;
else valm(1)=valm2(k(1)-n1); valm1current=0; valm2current=valm2(k(1)-n1);
end
// the variables valm1current and valm2current are created
// in order to remember where we are

// Computation of the other coefficients
i=2; n=size(m, '*');

while i<n
    if m(i)==m(i-1) then
        m(i)={}; valm(i)={}; n=n-1;
        if k(i)>n1 then
            valm(i-1)=valm1current+valm2(k(i)-n1); valm2current=valm2(k(i)-n1);
        else valm(i-1)=valm2current+valm1(k(i)); valm1current=valm1(k(i));
        end
        k(i)={};
    else if k(i)>n1 then
        valm(i)=valm1current+valm2(k(i)-n1); valm2current=valm2(k(i)-n1);
        else valm(i)=valm2current+valm1(k(i)); valm1current=valm1(k(i));
        end
        i=i+1;
    end
end

if m(n)==m(n-1) then
    m(n-1)={}; valm(n-1)={};
end

y=[m; valm];
endfunction
```

```

function y=evalpwc(t,m,valm,cont)

// m is a partition and valm is the values on the partition m
// cont='right' if we want a right continuous function
// (with intervals of the form [a;b[)
// cont='left' if we want a left continuous function
// (with intervals of the form ]a;b])

y=zeros(t);
n=length(m);

if cont=='right' then
    for i=1:(n-1)
        ki=find((t>=m(i))&(t<m(i+1)));
        y(ki)=valm(i);
    end
elseif cont=='left' then
    for i=1:(n-1)
        ki=find((t>m(i))&(t<=m(i+1)));
        y(ki)=valm(i);
    end
end
endfunction

```

### A.1.2 Computation of $S(\varphi_\lambda)$ and $B(\varphi_\lambda)$

First we propose the computation of  $S(f)$  for  $f = \mathbf{1}_{[0;2^{-j}]}$ , for  $j \geq 0$  via the computation of  $S_r(f)$  and the associated deterministic part.

```

// the random "piecewise constant" part : S_r(f)
function x=Srandom(U,j)

// U is a vector whose coordinates are in [0;T] (typically the parents)
// j>=0

par=[0]; valpar=[0];
for p=1:length(U)
    parnew=[U(p) U(p)+2^(-j)]; valparnew=[1 0]; // increment of the sum
    s=sumpwc(par,valpar,parnew,valparnew);
    par=s(1,:); valpar=s(2,:);
end
x=[par;valpar];
endfunction

// the deterministic (piecewise affine) part
function y=Sdeterm(t,j,T)

// t is a vector whose coordinates are in [0;T+2^(-j)] (typically the children)

```

```

// j>=0 and T>0

y=zeros(1,length(t));
y(find(0<t & t<=2^(-j)))=t(find(0<t & t<=2^(-j)));
y(find(2^(-j)<t & t<=T))=2^(-j);
y(find(T<t & t<=T+2^(-j)))=-t(find(T<t & t<=T+2^(-j)))+T+2^(-j);
y=[t;y];

// Warning! To obtain the complete deterministic part,
// we have to multiply by (n-1)/T, with n the number of parents
endfunction

```

This implementation allows the computation of  $S(f)$  for  $f = \mathbf{1}_{[2^{-j}k; 2^{-j}(k+1)]}$ , for  $j \geq 0$  and  $k \in \mathbb{Z}$ . This computation only consists in a shift by  $+2^{-j}k$  of the partition.

```

function S=Srd(n,x,t,j,k,T)

// typically n=length(U); x=Srandom(U,j);

y=Sdeterm(t-2^(-j)*k,j,T);
S=[t;evalpwc(t,x(1,:)+2^(-j)*k,x(2,:), 'left')-(n-1)/T*y(2,:)];
// that gives S(f) evaluated at the coordinates of t
endfunction

```

The following functions `Bfather` and `Bmother` correspond to the computation of  $B(f)$  for  $f = \mathbf{1}_{[0; 2^{-j}]}$  (with  $j \geq 0$ ) and for  $f = \mathbf{1}_{[2^{-J}; 2^{-j}]} - \mathbf{1}_{[0; 2^{-J}]}$  (with  $j \geq 0$  and  $J = j + 1$ ) respectively. We precise that for all  $k \in \mathbb{Z}$ , we have

$$B(\mathbf{1}_{[2^{-j}k; 2^{-j}(k+1)]}) = B(\mathbf{1}_{[0; 2^{-j}]})$$

and

$$B(\mathbf{1}_{[2^{-J}(2k+1); 2^{-j}(k+1)]} - \mathbf{1}_{[2^{-j}k; 2^{-J}(2k+1)]}) = B(\mathbf{1}_{[2^{-J}; 2^{-j}]} - \mathbf{1}_{[0; 2^{-J}]}),$$

that is to say the quantity  $B(f)$  does not depend on  $k$ .

```

function B=Bfather(n,x,j,T)

// typically n=length(U); x=Srandom(U,j);

// We are interested in larger differences between the random "pwc" part S_r
// and the deterministic part on intervals
b=[];

// On [0; 2^(-j)]
n1=length(find(x(1,:)<=2^(-j)));
// = "number of jumps of the function S_r before 2^(-j)" - 1
// (0 being the first element of the partition by default)

for i=2:n1
    bl=x(2,i-1)-(n-1)/T*x(1,i); // comparison at the left
    br=x(2,i)-(n-1)/T*x(1,i); // comparison at the right

```

```

    b=[b max(abs(bl),abs(br))];
end

// On [2-j];T]
n2=length(find(x(1,*)<=T)); // = "number of jumps of the function S_r before T" - 1
for i=n1:n2
    b=[b abs(x(2,i)-(n-1)/T*2-j)];
end

// On [T;T+2-j]
L=length(x(1,*)>T);
for i=n2+1:L // this is the case where there are jumps of the function S_r
    // between T-2-j and T
    bl=x(2,i-1)-(n-1)/T*(-x(1,i)+T+2-j);
    br=x(2,i)-(n-1)/T*(-x(1,i)+T+2-j);
    b=[b max(abs(bl),abs(br))];
end

B=max(b);
endfunction

function B=Bmother(n,x,j,T)

// typically n=length(U); x=Srandom(U,j);

J=j+1;

// I0 = [0;2-J] and I1 = ]2-J;2-j]
x0=[x(1,*)>0;x(2,*)>0];
x1=[x(1,*)>2-J;x(2,*)>2-j];
x=sumpwc(x0(1,*)>0,x0(2,*)>0,x1(1,*)>0,x1(2,*)>0);

// We are interested in larger differences between the random "pwc" part S_r
// and the deterministic part on intervals
b=[];

// On [0;2-J]
n1=length(find(x(1,*)<=2-J));
// = "number of jumps of the function S_r before 2-J" - 1
for i=2:n1
    bl=x(2,i-1)-(n-1)/T*(-x(1,i)); // comparison at the left
    br=x(2,i)-(n-1)/T*(-x(1,i)); // comparison at the right
    b=[b max(abs(bl),abs(br))];
end
b=[b max(abs(evalpwc(2-J,x(1,*)>0,x(2,*)>0,'left')+(n-1)/T*2-J),
    abs(evalpwc(2-J,x(1,*)>0,x(2,*)>0,'right')+(n-1)/T*2-J))];

// On [2-J;2-j]
n2=length(find(x(1,*)<=2-j));
// = "number of jumps of the function S_r before 2-j" - 1

```

```

for i=n1+1:n2
    bl=x(2,i-1)-(n-1)/T*(x(1,i)-2^(-j));
    br=x(2,i)-(n-1)/T*(x(1,i)-2^(-j));
    b=[b max(abs(bl),abs(br))];
end

// On [2^(-j);T]
n3=length(find(x(1,*)<=T)); // = "number of jumps of the function S_r before T" - 1
for i=n2:n3
    b=[b abs(x(2,i))];
end

// On [T;T+2^(-J)]
n4=length(find(x(1,*)<=T+2^(-J)));
// = "number of jumps of the function S_r before T+2^(-J)" - 1
for i=n3+1:n4
    bl=x(2,i-1)-(n-1)/T*(x(1,i)-T);
    br=x(2,i)-(n-1)/T*(x(1,i)-T);
    b=[b max(abs(bl),abs(br))];
end
b=[b max(abs(evalpwc(T+2^(-J),x(1,:),x(2,:),'left')-(n-1)/T*2^(-J)),
    abs(evalpwc(T+2^(-J),x(1,:),x(2,:),'right')-(n-1)/T*2^(-J)))];

// On [T+2^(-J);T+2^(-j)]
L=length(x(1,:));
for i=n4+1:L
    bl=x(2,i-1)-(n-1)/T*(-x(1,i)+T+2^(-j));
    br=x(2,i)-(n-1)/T*(-x(1,i)+T+2^(-j));
    b=[b max(abs(bl),abs(br))];
end

B=max(b);
endfunction

```

### A.1.3 Cascade algorithm for the Haar basis

We propose the implementation of the cascade algorithm described in Section 3.3.2 of Chapter 3. We precise that a cascade algorithm has been also performed for the particular Spline basis studied at Section 3 of Chapter 3 (see Figure 3.1).

```

// j0 = maximal resolution level (=5 in practice)
// T = normalized length of the studied genome (=10000 in practice)
// A = maximal memory (=10 in practice)
// We notice that we choose A an integer in the algorithm

function y=cascade(U,t,j0,T,A)

// typically U=Npar; t=Nchi; (respectively the parents and the children)

n=length(U);
y=list();

```

```

for j=1:j0+1
    y(j)=list(); // y(j) will be the list devoted to the resolution level j0-j
end

// Level j0
x=Srandom(U,j0); // the only S_r which is calculated
l=list();
for k=-2^j0*A:2^j0*A-1
    l($+1)=[x(1,:)+2^(-j0)*k;x(2,:)]; // the list l corresponds to the [par;valpar]
                                     // of S_r for k = -2^j0*A, ..., 2^j0*A-1;
    // namely l(i) gives the [par;valpar] of S_r for k=-2^j0*A-1+i
end

// Levels j0-1 to 0 via a cascade algorithm
for j=j0-1:-1:0
    J=j+1;
    y(j0-j)(1)=j; y(j0-j)(2)="mother"; Y=[]; lnew=list();
    B=2^(j/2)*Bmother(n,l(2^J*A+1),j,T);

    for k=-2^j*A:2^j*A-1
        // The corresponding mother wavelet is: 2^(j/2) * (I1 - I0)
        // with I0 = I_-[2^(-J)2k;2^(-J)(2k+1)] and I1 = I_-[2^(-J)(2k+1);2^(-J)(2k+2)]
        // Computations of S(I0) and S(I1)
        k0=2*k; i0=2^J*A+1+k0; // this is the corresponding index k
                               // of the father wavelets of the previous level
        Sd0=Sdeterm(t-2^(-J)*k0,J,T);
        S0=evalpwc(t,l(i0)(1,:),l(i0)(2,:),'left')-(n-1)/T*Sd0(2,:);
        k1=2*k+1; i1=2^J*A+1+k1;
        Sd1=Sdeterm(t-2^(-J)*k1,J,T);
        S1=evalpwc(t,l(i1)(1,:),l(i1)(2,:),'left')-(n-1)/T*Sd1(2,:);
        // Computation of S
        S=2^(j/2)*(S1-S0);
        // Computation of G
        G=sum(S);
        //Computation of hatV
        hatV=sum(S^2);
        // Increment of Y
        Y=[Y [k;G;B;hatV]];
        // We notice that the difference between two father wavelets of the previous
        // level allows to obtain the mother wavelet of the corresponding level j
        // Now, we do the sum of the two father wavelets
        // in order to obtain the father wavelet of the corresponding level j
        // (for the computations of the mother wavelet of the next level)
        s=sumpwc(l(i0)(1,:),l(i0)(2,:),l(i1)(1,:),l(i1)(2,:));
        lnew($+1)=[s(1,:);s(2,:)];
    end

    y(j0-j)(3)=Y; l=lnew;
end

// Level -1 (corresponding at the father wavelet of level 0

```



```

//          which is used for the reconstruction of the signal)
y(j0+1)(1)=-1; y(j0+1)(2)="father"; Y=[];
B=Bfather(n,l(A+1),0,T);
for k=-A:A-1
    i=A+1+k;
    Sd=Sdeterm(t-k,0,T);
    S=evalpwc(t,l(i)(1,:),l(i)(2:),'left')-(n-1)/T*Sd(2,:);
    // Computation of G
    G=sum(S);
    // Computation of hatV
    hatV=sum(S^2);
    // Increment of Y
    Y=[Y [k;G;B;hatV]];
end

y(j0+1)(3)=Y;
endfunction

```

Consequently the function `cascade` gives the quantities of interest:  $\hat{\beta}_\lambda = G(\varphi_\lambda)/n$ ,  $B(\varphi_\lambda)$  and  $\hat{V}(\varphi_\lambda)$  for the index  $\lambda = (j, k)$  for  $j \leq j_0$  (the maximal resolution level) and so, the threshold  $\eta_\lambda$  defined by (3.14) can be exactly computed with a fixed couple of parameters  $(\gamma, \delta)$ .

## A.2 Implementation of the testing procedure of Chapter 4

This section gives some algorithmic elements of our practical testing procedure of Chapter 4 (see Section 4.4.4). We present here the script related on the evaluation of the empirical rate of type I error, namely the level of our test.

### A.2.1 Simulation of the data

We simulate 5000 independent realizations of  $Data_0$  with  $T = 2$  (time of record in seconds): we simulate processes  $N_p$  (the parents) and  $N_c$  (the children) whose respective intensities are  $\tilde{\lambda}_p = 50$  and  $\tilde{\lambda}_c = 50$ . As explained in Section 4.4.4, the data are multiplied by 50 before being treated, namely in the sequel  $T = 100$ .

### A.2.2 Cascade algorithm and functions

The function `cascadetest` consists in the computation of the testing statistics  $\hat{T}_\lambda = |\hat{\beta}_\lambda|$ , where

$$\hat{\beta}_\lambda = \frac{1}{n} \sum_{U \in N_p} \sum_{V \in N_c} \left[ \varphi_\lambda(V - U) - \frac{n-1}{n} \mathbb{E}_{W \sim \pi}(\varphi_\lambda(V - W)) \right],$$

for  $\lambda = (j, k)$  with  $j \leq j_0$  ( $= 3$  in practice), with respect to a given couple of processes  $N_p$  and  $N_c$ .

```
function y=cascadetest(U,V,n,m,B,T,A,j0)
```

```

// U=parents: matrix 1 x n
// V=children: matrix m x B

```

```

y=list();
for j=1:j0
    y(j)=[]; // y(j) will be the matrix for the level j0-j
             // and its corresponding indices k
end
// We don't keep the level -1 corresponding to the father wavelets.

// Initialization: computation of the (non-renormalized) father wavelets at level j0
MatU=U.*ones(m,1);
c=[-A:2^(-j0):A]; l=length(c);
Matc=c(1:l-1).*ones(m,1);

S=zeros(l-1,B);
for b=1:B
    [ind,occ,info]=dsearch(V(:,b).*ones(1,n)-MatU,c);
    X=V(:,b).*ones(1,l-1)-Matc;
    S(:,b)=( occ - (n-1)/T * sum(X.* (X>0 & X<2^(-j0)) + 2^(-j0).*( X>=2^(-j0) & X<=T
        + (T-X+2^(-j0)).*( X>T & X<=T+2^(-j0)),1) )');
end

Sm=[]; Snew=[]; lnew=(l-1)/2;
for j=j0-1:-1:0
    Wf=eye(lnew,lnew).*sysdiag([1 1]);
    Wm=eye(lnew,lnew).*sysdiag([-1 1]);
    Sm=2^(j/2)*Wm*S; y(j0-j)=abs(Sm);
    Snew=Wf*S; S=Snew;
    lnew=lnew/2; // Remark: l-1 = A*2^(j0+1)
end
// Warning! The coefficients are not divided by n here
endfunction

```

The function `QuantileT` consists in the computation of  $u_{\alpha}^{[U_1, \dots, U_n; m]}$  defined by (4.5), obtained by dichotomie. This function needs the following preliminary function `testprobathresh`.

```

function [Diff,quant] = testprobathresh(ualpha,yBquantsort,yBu,alpha,Nbessai)

// ualpha in [0;1]

// yBquantsort is a list of size j0 where the j-th list is a matrix
// of size A*2^(j0-j+1) x B/2 which contains the quantiles
// putting in ascending order (by row)

// yBu is a list of size j0 where the j-th list is a matrix
// of size A*2^(j0-j+1) x B/2 which contains the testing statistics
// estimated under the hypothesis H0 (to estimate ualpha later)

// alpha is the level test: typically alpha = 0.05

// Nbessai=B/2 typically

quant=list();

```

```

Bilanniveau=zeros(j0,Nbessai);
for j=1:j0 // weights depending on the level j: 1/(j0*A*2^(j0-j+1))
    ntemp = max(min(floor(Nbessai*(1-ualpha/(j0*A*2^(j0-j+1))))+1, Nbessai), 1);
    // the functions 'max' and 'min' force ualpha to be in [0;1]

    quant(j)=yBquantsort(j)(:,ntemp);

    Bilanniveau(j,:)=sum((yBu(j)>=quant(j))*ones(1,Nbessai),1);
    // this is the number of simple tests which have rejected H0 for each simulation
end

Diff=mean(1*(sum(Bilanniveau,1)>0))-alpha;
// the first term is an estimator of ualpha

// Remark: we want Diff < 0
//          i.e. an estimator of the level test < alpha but as close as possible
endfunction

```

Now, we define the function `QuantileT`.

```

function [quant,ualphamin,D]=QuantileT(yBquantsort,yBu,alpha,Nbessai)

// yBquantsort is a list of size j0 where the j-th list is a matrix
// of size A*2^(j0-j+1) x B/2 which contains the quantiles
// putting in ascending order (by row)

// yBu is a list of size j0 where the j-th list is a matrix
// of size A*2^(j0-j+1) x B/2 which contains the testing statistics
// estimated under the hypothesis H0 (to estimate ualpha later)

// alpha is the level test: typically alpha = 0.05

// Nbessai=B/2 typically

// Initialization
ualphamin=0; ualphamax=1;
[Diffmin,quantmin] = testprobathresh(ualphamin,yBquantsort,yBu,alpha,Nbessai);
D=Diffmin;
[Diffmax,quantmax] = testprobathresh(ualphamax,yBquantsort,yBu,alpha,Nbessai);

// Dichotomie

while Diffmax<0
    ualphamax=ualphamax*2;
    [Diffmax,quantmax] = testprobathresh(ualphamax,yBquantsort,yBu,alpha,Nbessai);
end
// This step allows us to obtain ualphamax large enough

if Diffmin<0 then
    while ualphamax-ualphamin>1/Nbessai
        ualphatemp=(ualphamin+ualphamax)/2;

```

```

    [Difftemp,quanttemp]= testprobathresh(ualphatemp,yBquantsort,yBu,alpha,Nbessai);
    if Difftemp>0 then ualphamax=ualphatemp;
    else ualphamin=ualphatemp; quantmin=quanttemp;
    end
end
end
quant=quantmin;
endfunction

```

### A.2.3 Computation of the empirical rate of type I error

This section gives the steps to evaluate the empirical rate of type I error. We consider a test of level  $\alpha = 0.05$ .

```

alpha=0.05;

// Step 1: Simulation of the real data

T=100; A=2; j0=3;
Nsimul=5000;
Z=zeros(Nsimul,3);
for nsimul=1:Nsimul
Npar=N_p(nsimul,:); Nchi=N_c(nsimul,:); // the nsimul-th simulation of Data0
n=length(Npar); m=length(Nchi);

// Step 2: Computation of the testing statistic

yvrai=cascadetest(Npar,Nchi,n,m,1,T,A,j0);

// Step 3: B simulations (to estimate ualpha and the quantiles by Monte Carlo methods)

B=20000;
U=Npar; // the sequel is conditionally on the parents
V=grand(m,B,'unf',-A,T+A); // distribution of the children under the hypothesis H0
// the sequel is conditionally on the number of children
yB=cascadetest(U,V,n,m,B,T,A,j0);

// Step 4: Creation of two samples

yBquant=list(); yBu=list();
for j=1:j0
    yBquant(j)=yB(j)(:,1:B/2); // yBquant(j) will be the matrix devoted to the
// level j0-j with the first half of the sample
// allowing the estimation of the quantiles
    yBu(j)=yB(j)(:,B/2+1:B); // yBu(j) will be the matrix devoted to the level j0-j
// with the second half of the sample
// allowing the estimation of ualpha
end

```

```

// Step 5: Computation of the quantiles (yBquant)

yBquantsort=list();
for j=1:j0
    yBquantsort(j)=gsort(yBquant(j),'c','i');
end
// Each row is put in ascending order

// Step 6: Computation of ualpha by dichotomie (yBquant)

Nbessai=B/2;
[quant,ualphamin,D]=QuantileT(yBquantsort,yBu,alpha,Nbessai);

// Step7: We "test" the testing statistic yvrai:
//     if there exists one lambda=(j,k) s.t. yvrai > associated estimated quantile,
//     then we reject H0

test=list(); testagg=0;
for j=1:j0
    test(j)=(yvrai(j)>quant(j))*1;
    testagg=testagg+sum(test(j));
end
if testagg<>0 then decision=1;
else decision=0;
end

Z(nsimul,:)=[ualphamin D testagg];
end

// Evaluation of the empirical rate of type I error

length(find(Z(:,3)<>0))/5000 // the number of simulations rejecting H0
                          // divided by the number of simulations 5000

```

We precise that similar scripts with the other data sets ( $Data_{10}, \dots$ ) provide the empirical powers of our tests.

# Bibliographie

- [1] ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. AND KEIDING, N. *Statistical models based on counting processes*. Springer-Verlag, New York, 1993.
- [2] BACRY, E., DELATTRE, S., HOFFMANN, M. AND MUZY, J. F. Modelling microstructure noise by mutually exciting point processes. *Quantitative Finance*, 13:65–77, 2013.
- [3] BAIN, L. J., ENGELHARDT, M. AND WRIGHT, F. T. Tests for an increasing trend in the intensity of a Poisson process: a power study. *J. Amer. Statist. Assoc.*, 80(390):419–422, Jun. 1985.
- [4] BARAUD, Y. Non asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8:577–606, 2002.
- [5] BARAUD, Y., HUET, S. AND LAURENT, B. Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251, 2003.
- [6] BARLOW, H. B. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.
- [7] BARRON, A. R., BIRGÉ, L. AND MASSART, P. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113:301–413, 1999.
- [8] BENJAMINI, Y. AND HOCHBERG, Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 57(1):289–300, 1995.
- [9] BERTIN, K., LE PENNEC, E. AND RIVOIRARD, V. Adaptive Dantzig density estimation. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):43–74, 2011.
- [10] BHATTACHARJEE, M., DESHPANDE, J. V. AND NAIK-NIMBALKAR, U. V. Unconditional tests of goodness of fit for the intensity of time-truncated nonhomogeneous Poisson processes. *Technometrics*, 46(3):330–338, Aug. 2004.
- [11] BIRGÉ, L. AND MASSART, P. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, 1997.
- [12] BOUCHERON, S., BOUSQUET, O., LUGOSI, G. AND MASSART, P. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, Mar. 2005.
- [13] BRÉMAUD, P. AND MASSOULIÉ, L. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3):1563–1588, 1996.
- [14] BRÉMAUD, P. AND MASSOULIÉ, L. Hawkes branching point processes without ancestors. *J. Appl. Prob.*, 38:122–135, 2001.
- [15] BUNEA, F. Consistent selection via the Lasso for high dimensional approximating regression models. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, pages 122–138. Inst. Math. Statist. Collect., 2008.

- [16] BUNEA, F., TSYBAKOV, A. B. AND WEGKAMP, M. H. Sparse density estimation with  $\ell_1$  penalties. *Lecture Notes in Artificial Intelligence*, 4539:530–543, 2007.
- [17] BUNEA, F., TSYBAKOV, A. B. AND WEGKAMP, M. H. Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.*, 1:169–194, 2007.
- [18] CANDÈS, E. J. AND PLAN, Y. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.
- [19] CANDÈS, E. J. AND TAO, T. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, Dec. 2005.
- [20] CANDÈS, E. J. AND TAO, T. The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.*, 35(6):2313–2351, 2007.
- [21] CARSTENSEN, L., SANDELIN, A., WINTHER, O. AND HANSEN, N. R. Multivariate Hawkes process models of the occurrence of regulatory elements and an analysis of the pilot ENCODE regions. *BMC Bioinformatics*, 11(456), 2010.
- [22] CAVALIER, L. AND KOO, J.-Y. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Trans. Inform. Theory*, 48:2794–2802, 2002.
- [23] COHEN, A., DAUBECHIES, I. AND FEAUVEAU, J.-C. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.
- [24] COHEN, A., DEVORE, R. A. AND HOCHMUTH, R. Restricted nonlinear approximation. *Constr. Approx.*, 16:85–113, 2000.
- [25] COHEN, A. AND SACKROWITZ, H. B. Evaluating tests for increasing intensity of a Poisson process. *Technometrics*, 35(4):446–448, Nov. 1993.
- [26] COMTE, F., GAÏFFAS, S. AND GUILLOUX, A. Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1171–1196, 2011.
- [27] COX, D.R. Some statistical methods connected with series of events (with Discussion). *J. Roy. Statist. Soc. Ser. B*, 17:129–164, 1955.
- [28] DACHIAN, S. AND KUTOYANTS, Y. A. Hypotheses testing: Poisson versus self-exciting. *Scand. J. Stat.*, 33(2):391–408, 2006.
- [29] DALEY, D. J. AND VERE-JONES, D. *An Introduction to the Theory of Point Processes - Volume I: Elementary Theory and Methods*. Probab. Appl. (N. Y.). Springer, 2003.
- [30] DARLING, D. A. The Kolmogorov-Smirnov, Cramér-von Mises tests. *Ann. Math. Statist.*, 28(4):823–838, Dec. 1957.
- [31] DAUBECHIES, I. *Ten Lectures on Wavelets*, volume 61. CBMS-NSF Regional Conf. Ser. in Appl. Math., 1992.
- [32] DE LA PEÑA, V. H. AND GINÉ, E. *Decoupling: From Dependence to Independence*. Probab. Appl. (N. Y.). Springer, 1999.
- [33] DE LA PEÑA, V. H. AND MONTGOMERY-SMITH, S. J. Decoupling inequalities for the tail probabilities of multivariate  $U$ -statistics. *Ann. Probab.*, 23(2):806–816, 1995.
- [34] DEVORE, R. A. AND LORENTZ, G. G. *Constructive Approximation*. Springer-Verlag, 1993.

- [35] DONOHO, D. L. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics*, volume 47, pages 173–205. Amer. Math. Soc., 1993.
- [36] DONOHO, D. L. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, Apr. 2006.
- [37] DONOHO, D. L. AND JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- [38] DONOHO, D. L. AND JOHNSTONE, I. M. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90:1200–1224, 1995.
- [39] DONOHO, D. L. AND JOHNSTONE, I. M. Neo-classical minimax problems, thresholding and adaptation. *Bernoulli*, 2:39–62, 1996.
- [40] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AND PICARD, D. Wavelet shrinkage: asymptopia? (with discussion). *J. R. Stat. Soc. Ser. B*, 57:301–369, 1995.
- [41] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AND PICARD, D. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996.
- [42] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AND PICARD, D. Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*, pages 183–218. Springer, 1997.
- [43] FAZLI, K. Second order efficient test for inhomogeneous Poisson processes. *Stat. Inference Stoch. Process.*, 2:181–208, 2007.
- [44] FAZLI, K. AND KUTOYANTS, Y. A. Two simple hypotheses testing for Poisson process. *Far East J. Theor. Stat.*, 15(2):251–290, Mar. 2005.
- [45] FROMONT, M., LAURENT, B. AND REYNAUD-BOURET, P. Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):176–213, 2011.
- [46] FROMONT, M., LAURENT, B. AND REYNAUD-BOURET, P. The two-sample problem for Poisson processes: adaptive tests with a non-asymptotic wild bootstrap approach. *arXiv:1203.3572v2*, Nov. 2012.
- [47] GERSTEIN, G. L. Searching for significance in spatio-temporal firing patterns. *Acta Neurol Exp (Wars)*, 64(2):203–207, 2004.
- [48] GINÉ, E., LATALA, R. AND ZINN, J. Exponential and moment inequalities for  $U$ -statistics. In *High Dimensional Probability II*, volume 47, pages 13–38. Progr. Probab. Birkhäuser Boston, 2000.
- [49] GOLDENSHLUGER, A. AND LEPSKI, O. V. Structural adaptation via  $L_p$ -norm oracle inequalities. *Probab. Theory Related Fields*, 143:41–71, 2009.
- [50] GRAMMONT, F. AND RIEHLE, A. Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp Brain Res.*, 128(1-2):118–122, 1999.
- [51] GRÜN, S. *Unitary Joint-Events in Multiple-Neuron Spiking Activity: Detection, Significance, and Interpretation*. Reihe Physik, Band 60. Verlag Harri Deutsch, Thun, Frankfurt/Main, 1996.
- [52] GRÜN, S., DIEMANN, M. AND AERTSEN, A. Unitary Event analysis. In *Analysis of Parallel Spike Trains*, volume 7, pages 191–220. Springer Ser. Comput. Neurosci., 2010.



- [53] GRÜN, S., DIEMANN, M., GRAMMONT, F., RIEHLE, A. AND AERTSEN, A. Detecting unitary events without discretization in time. *J. Neurosci. Methods*, 93:67–79, 1999.
- [54] GUSTO, G. AND SCHBATH, S. FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes’ model. *Stat. Appl. Genet. Mol. Biol.*, 4(1), Sep. 2005.
- [55] HALPERN, D., CHIAPELLO, H., SCHBATH, S., ROBIN, S., HENNEQUET-ANTIER, C., GRUSS, A. AND EL KAROUI, M. Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet.*, 3(9), 2007.
- [56] HANSEN, N. R., REYNAUD-BOURET, P. AND RIVOIRARD, V. Lasso and probabilistic inequalities for multivariate point processes. *arXiv:1208.0570v1*, Aug. 2012.
- [57] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. AND TSYBAKOV, A. *Wavelets, Approximation and Statistical Applications*, volume 129 of *Lecture Notes in Statist.* Springer New York, 1998.
- [58] HARMANY, Z., MARCIA, R. AND WILLETT, R. Spatio-temporal compressed sensing with Coded Apertures and Keyed Exposures. *arXiv:1111.7247v2*, Jan. 2011.
- [59] HAWKES, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [60] HOEFFDING, W. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, 19(3):293–325, 1948.
- [61] HOUDRÉ, C. AND REYNAUD-BOURET, P. Exponential inequalities, with constants, for  $U$ -statistics of order two. In *Stochastic Inequalities and Applications*, volume 56, pages 55–69. Progr. Probab. Birkhäuser Verlag, 2003.
- [62] HUANG, J., MA, S. AND ZHANG, C.-H. Adaptive Lasso for sparse high-dimensional regression models. *Statist. Sinica*, 18:1603–1618, 2008.
- [63] INGSTER, Y. I. AND KUTOYANTS, Y. A. Nonparametric hypothesis testing for intensity of the Poisson process. *Math. Methods Statist.*, 16(3):217–245, 2007.
- [64] JOHNSTONE, I. M. Minimax Bayes, asymptotic minimax and sparse wavelet priors. In *Statistical Decision Theory and Related Topics*, pages 303–326. Springer (N. Y.), 1994.
- [65] JUDITSKY, A. AND LAMBERT-LACROIX, S. On minimax density estimation on  $R$ . *Bernoulli*, 10(2):187–220, Apr. 2004.
- [66] KERKYACHARIAN, G. AND PICARD, D. Thresholding algorithms, maxisets and well-concentrated bases. With discussion and a rejoinder by the authors. *Test*, 9(2):283–345, 2000.
- [67] KINGMAN, J. F. C. *Poisson Processes*. Oxford Stud. Probab. Oxford University Press, New York, 1993.
- [68] KOLACZYK, E. D. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica*, 9:119–135, 1999.
- [69] KRUMIN, M., REUTSKY, I. AND SHOHAM, S. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front Comput Neurosci.*, 4(147), 2010.
- [70] LEPSKI, O. V. Asymptotic minimax adaptive estimation. 1. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, 36:682–697, 1991.

- [71] LEPSKI, O. V. Asymptotic minimax adaptive estimation. 2. Statistical models without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.*, 37:433–448, 1992.
- [72] LEPSKI, O. V. On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation*, volume 12, pages 87–106. Adv. Soviet Math., 1992.
- [73] LESTIENNE, R. Determination of the precision of spike timing in the visual cortex of anaesthetised cats. *Biol. Cybern.*, 74:55–61, 1996.
- [74] MALLAT, S. Multiresolution approximations and wavelet orthonormal bases of  $L^2(\mathbb{R})$ . *Trans. Amer. Math. Soc.*, 315(1):69–87, Sep. 1989.
- [75] MALLAT, S. *A Wavelet Tour of Signal Processing, 3rd ed.* Academic Press, 2008.
- [76] MARCIA, R. AND WILLETT, R. Compressive coded aperture superresolution image reconstruction. *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008.
- [77] MASSART, P. *Concentration Inequalities and Model Selection - Éc. Été Probab. St.-Flour 2003*, volume 1896 of *Lecture Notes in Math.* Springer Berlin Heidelberg, 2007.
- [78] MEYER, Y. *Wavelets and Operators.* Cambridge University Press, 1995. *Translator: Salingier, D. H.*
- [79] MITCHELL, L. AND CATES, M. E. Hawkes process as a model of social interactions: a view on video dynamics. *J. Phys. A*, 2010.
- [80] MØLLER, J., SYVERSVEEN, A.-R. AND WAAGEPETERSEN, R. P. Log Gaussian Cox processes. *Scand. J. Stat.*, 25(3):451–482, Sep. 1998.
- [81] NASON, G. P. Wavelet shrinkage using cross-validation. *J. R. Stat. Soc. Ser. B*, 58:463–479, 1996.
- [82] OGATA, Y. AND AKAIKE, H. On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 44(1):102–107, 1982.
- [83] OZAKI, T. Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Ann. Inst. Statist. Math.*, 31(1):145–155, 1979.
- [84] PERNICE, V., STAUDE, B., CARDANOBILO, S. AND ROTTER, S. How structure determines correlations in neuronal networks. *PLoS Comput Biol*, 7(5), 2011.
- [85] PERNICE, V., STAUDE, B., CARDANOBILO, S. AND ROTTER, S. Recurrent interactions in spiking networks with arbitrary topology. *Phys. Rev. E*, 85(3), 2012.
- [86] PICARD, F., SCHBATH, S., LEBARBIER, E., NEUVIAL, P. ET CHIQUET, J. Statistiques et génome. Dans la *Gazette des Mathématiciens*, volume 130, pages 51–82. SMF, Oct. 2011.
- [87] RASMUSSEN, J. G. *Aspects of temporal and spatio-temporal processes.* PhD thesis, Aalborg University - Denmark, 2006.
- [88] REINERT, G. AND SCHBATH, S. Large compound Poisson approximations for occurrences of multiple words. In *Statistics in Molecular Biology and Genetics*, volume 33, pages 257–275. IMS Lecture Notes Monogr. Ser., 1999.
- [89] REINERT, G., SCHBATH, S. AND WATERMAN, M. S. Statistics on words with applications to biological sequences. In *Applied Combinatorics on Words*, volume 105. Cambridge University Press, 2005.

- [90] REYNAUD-BOURET, P. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields*, 126:103–153, 2003.
- [91] REYNAUD-BOURET, P. AND RIVOIRARD, V. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electron. J. Stat.*, 4:172–238, 2010.
- [92] REYNAUD-BOURET, P., RIVOIRARD, V. AND TULEAU-MALOT, C. Adaptive density estimation: a curse of support? *J. Statist. Plann. Inference*, 141:115–139, 2011.
- [93] REYNAUD-BOURET, P. AND SCHBATH, S. Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.*, 38(5):2781–2822, 2010.
- [94] RIVOIRARD, V. Maxisets for linear procedures. *Statist. Probab. Lett.*, 67(3):267–275, 2004.
- [95] RIVOIRARD, V. Thresholding procedure with priors based on Pareto distributions. *Test*, 13(1):213–246, 2004.
- [96] ROBIN, S., RODOLPHE, F. ET SCHBATH, S. *ADN, mots et modèles*. Belin, 2003.
- [97] ROBIN, S., RODOLPHE, F. AND SCHBATH, S. *DNA, Words and Models*. Cambridge Univ. Press, 2005.
- [98] ROMBERG, J. Compressive sensing by random convolution. *SIAM J. Imaging Sci.*, 2(4):1098–1128, 2009.
- [99] RUDELSON, M. AND VERSHYNIN, R. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
- [100] RUDEMO, M. Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.*, 9:65–78, 1982.
- [101] SANDHU, K. S., LI, G., POH, H. M., QUEK, Y. L., SIA, Y. Y., PEH, S. Q., MULAWADI, F. H., LIM, J., SIKIC, M., MENGHI, F., THALAMUTHU, A., SUNG, W. K., RUAN, X., FULLWOOD, M. J., LIU, E., CSERMELY, P. AND RUAN, Y. Large-scale functional organization of long-range chromatin interaction networks. *Cell Rep.*, 2(5):1207–1219, Oct. 2012.
- [102] SCHBATH, S. Compound Poisson approximation of word counts in DNA sequences. *ESAIM Probab. Stat.*, 1:1–16, 1995.
- [103] SCHBATH, S. AND ROBIN, S. How pattern statistics can be useful for DNA motif discovery? In *Scan Statistics: Methods and Applications (Statistics for Industry and Technology)*. Birkhäuser, 2009.
- [104] TANAKA, U., OGATA, Y. AND STOYAN, D. Parameter estimation and model selection for Neyman-Scott point processes. *Biom. J.*, 50(1):43–57, Feb. 2008.
- [105] TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 58(1):267–288, 1996.
- [106] TOUZAIN, F., PETIT, M.-A., SCHBATH, S. AND EL KAROUI, M. DNA motifs that sculpt the bacterial chromosome. *Nature Reviews Microbiology*, 9:15–26, Jan. 2011.
- [107] TRIEBEL, H. *Theory of Function Spaces*, volume 78 of *Monogr. Math.* Birkhäuser Verlag, Switzerland, 1983.
- [108] TSYBAKOV, A. B. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications*. Springer, 2004.

- 
- [109] TULEAU-MALOT, C., ROUIS, A., REYNAUD-BOURET, P. AND GRAMMONT, F. Multiple tests based on a Gaussian approximation of the Unitary Events method. *hal-00757323*, Nov. 2012.
- [110] VAN DE GEER, S. High-dimensional generalized linear models and the Lasso. *Ann. Statist.*, 36(2):614–645, 2008.
- [111] VAN DE GEER, S., BÜHLMANN, P. AND ZHOU, S. The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso). *Electron. J. Stat.*, 5:688–749, 2011.
- [112] VAN DER VAART, A. W. *Asymptotic Statistics*, volume 3 of *Camb. Ser. Stat. Probab. Math.* Cambridge University Press, 1998.
- [113] VERE-JONES, D. AND OZAKI, T. Some examples of statistical estimation applied to earthquake data. *Ann. Inst. Statist. Math.*, 34(1):189–207, 1982.
- [114] WILLETT, R. M. AND NOWAK, R. D. Multiscale Poisson intensity and density estimation. *IEEE Trans. Inform. Theory*, 53(9):3171–3187, Sep. 2007.
- [115] YING, Y. AND CAMPBELL, C. Rademacher chaos complexities for learning the kernel problem. *Neural Comput.*, 22(11):2858–2886, 2010.
- [116] ZHANG, B., FADILI, J. M. AND STARCK, J.-L. Wavelets, ridgelets, and curvelets for Poisson noise removal. *IEEE Trans. Image Process.*, 17(7):1093–1108, Jul. 2008.
- [117] ZOU, H. The adaptive Lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476), Dec. 2006.





