



Quelques Problèmes de Statistique autour des processus de Poisson

Gaspar Massiot

► To cite this version:

Gaspar Massiot. Quelques Problèmes de Statistique autour des processus de Poisson. Statistiques [math.ST]. École normale supérieure de Rennes, 2017. Français. ⟨NNT : 2017ENSR0006⟩. ⟨tel-01591254⟩

HAL Id: tel-01591254

<https://theses.hal.science/tel-01591254v1>

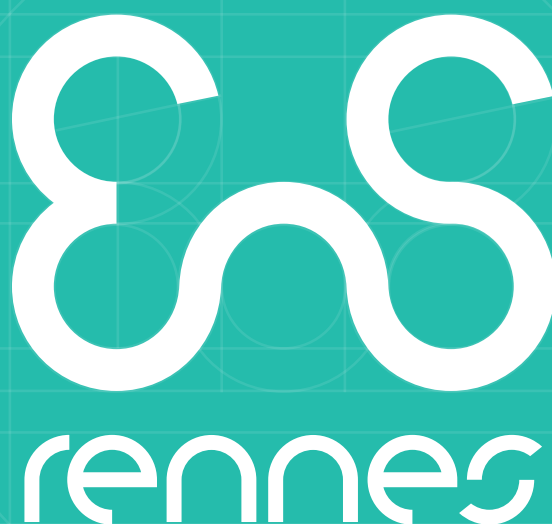
Submitted on 21 Sep 2017

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.



HAL Authorization



THÈSE / ENS RENNES

Université Bretagne Loire

pour obtenir le titre de

DOCTEUR DE L'ÉCOLE NORMALE SUPÉRIEURE DE RENNES

Mention : Mathématiques et Applications

École doctorale MATISSE

présentée par

Gaspar Massiot

Préparée à l'unité mixte de recherche 6625

Institut de recherche mathématiques de Rennes

Quelques Problèmes de Statistique autour des processus de Poisson

Thèse soutenue le 7 juillet 2017

devant le jury composé de :

Fabienne Comte / *rapporteur*

Professeure, Université Paris Descartes

André Mas / *rapporteur*

Professeur, Université Montpellier II

Anne Gégout-Petit / *examinateur*

Professeure, Université de Lorraine

Magalie Fromont / *examinateur*

Maître de conférences, Université Rennes 2

Benoît Cadre / *directeur de thèse*

Professeur, ENS Rennes

Nicolas Klutchnikoff / *co-directeur de thèse*

Maître de conférences, Université Rennes 2

Quelques Problèmes de Statistique autour des processus de Poisson

Gaspar Massiot
ENS Rennes

Ph.D. Dissertation

Remerciements

Je tiens dans un premier temps à remercier mes deux directeurs de thèse, Benoît Cadre et Nicolas Klutchnikoff, pour m'avoir permis de travailler sur un sujet de recherche intéressant, mêlant probabilités et statistiques et pour m'avoir encadré au cours de ces quatre années. Nicolas, je te remercie pour le soutien moral aussi bien que logistique et technique que tu m'as apporté durant cette thèse. Benoît, tu as toujours su te montrer disponible et efficace. Je te remercie pour ton indispensable rigueur.

Merci à mes rapporteurs André Mas et Fabienne Comte pour le temps que vous avez consacré à mon travail et les conseils donnés pour l'amélioration du manuscrit. Merci également à mes examinateurs Magalie Fromont et Anne Gégout-Petit d'avoir accepté de faire partie de mon jury de thèse.

Je tiens à remercier l'ensemble des membres de l'équipe de statistique de l'Ensaï pour leurs qualités d'enseignants dont j'ai profité durant les 3 années qui ont précédé ma thèse et leurs qualités en tant que collègues avec qui j'ai partagé mon quotidien pendant les 3 premières années de ma thèse. Merci Lionel, Myriam, Valentin, Marian, François, Guillaume, Jocelyn et Brigitte et plus récemment, Adrien, Fabien, Salima, Pierre, et Pavlo.

Merci aussi à l'équipe de statistiques de Rennes 2 qui m'a accueilli pour mon année d'ATER. Merci à Jean-Louis d'avoir répondu à mes nombreuses questions et d'être toujours disponible. Merci à Nicolas, Laurent, Pierre-André, Dominique et Madison pour leur bonne humeur.

Je n'oublie pas de remercier mes acolytes de thèse : Samuel et Cyril avec qui j'ai passé une grande partie de mon temps à l'Ensaï, ainsi qu'Audrey, Vincent et Tam pour les bonnes soirées passées ensemble. J'adresse aussi une pensée à mes amis d'étude montés pour la plupart sur paris : François, Matthieu, Alexandre, Adrien, Thomas, Sébastien et à tous ceux que je n'ai pas cités dans ces remerciements et avec qui j'ai passé de bons moments

qu'ils aient été de passage à Rennes ou bien rencontrés en conférence ou ailleurs.

Je remercie aussi bien entendu mes parents pour leur soutien et mon frère aujourd'hui aux Etats-Unis.

Pour finir, merci à Emeline.

QUELQUES PROBLÈMES DE STATISTIQUE AUTOUR DES PROCESSUS DE POISSON

Résumé : L'objectif principal de cette thèse est de développer des méthodologies statistiques adaptées au traitement de données issues de processus stochastiques et plus précisément de processus de Cox. Les problématiques étudiées dans cette thèse sont issues des trois domaines statistiques suivants : les tests non paramétriques, l'estimation non paramétrique à noyaux et l'estimation minimax. Dans un premier temps, nous proposons, dans un cadre fonctionnel, des statistiques de test pour détecter la nature Poissonienne d'un processus de Cox. Nous étudions ensuite le problème de l'estimation minimax de la régression sur un processus de Poisson ponctuel. En se basant sur la décomposition en chaos d'Itô, nous obtenons des vitesses comparables à celles atteintes pour le cas de la régression Lipschitz en dimension finie. Enfin, dans le dernier chapitre de cette thèse, nous présentons un estimateur non-paramétrique de l'intensité d'un processus de Cox lorsque celle-ci est une fonction déterministe d'un co-processus.

Mots-Clés : Statistique fonctionnelle, processus de Cox, tests statistiques, théorie Martingale, processus ponctuels de Poisson, estimation de la régression, estimation Minimax, estimation de l'intensité, lissage à noyaux.

SOME STATISTICAL PROBLEMS AROUND POISSON PROCESSES

Abstract: The main purpose of this thesis is to develop statistical methodologies for stochastic processes data and more precisely Cox process data. The problems considered arise from three different contexts: nonparametric tests, nonparametric kernel estimation and minimax estimation. We first study the statistical test problem of detecting whether a Cox process is Poisson or not. Then, we introduce a semiparametric estimate of the regression over a Poisson point process. Using Itô's famous chaos expansion for Poisson functionals, we derive asymptotic minimax properties of our estimator. Finally, we introduce a nonparametric estimate of the intensity of a Cox process whenever it is a deterministic function of a known coprocess.

Keywords: Functional Statistic, Cox process, test Statistic, Martingale theory, Poisson point process, regression estimate, Minimax estimation, intensity estimation, Kernel smoothing.

Table des matières

1	Introduction	11
1.1	Processus de Poisson et de Cox	12
1.1.1	Processus sur un espace mesurable σ -fini	13
1.1.2	Processus en temps continu	16
1.2	Tests non-paramétriques pour les processus de Cox	21
1.2.1	Tests pour les processus de Cox	21
1.2.2	Principe du test	22
1.2.3	Alternative locale	25
1.3	Estimation minimax de la régression	27
1.3.1	Estimation de la régression dans un cadre fonctionnel	27
1.3.2	Régression minimax pour une covariable issue d'un processus de Poisson	29
1.3.3	Développement en chaos de Wiener-Itô dans l'espace de Poisson	30
1.3.4	Modèle	31
1.3.5	Estimateur des chaos	33
1.3.6	Propriétés minimax lorsque δ est connu	34
1.3.7	Propriétés adaptatives lorsque δ est inconnu	35
1.4	Estimation non paramétrique à noyaux	36
1.4.1	Estimation de l'intensité du processus de Cox	36
1.4.2	Présentation du modèle	38
1.4.3	Stratégie d'estimation	39
1.4.4	Résultats	40
1.4.5	Validation Croisée	43
2	Nonparametric test	45
2.1	Introduction	46
2.2	Tests for Cox processes	47

2.2.1	Principle of the test	47
2.2.2	Testing $\mathbf{H}_0$ vs $\mathbf{H}_1$	49
2.2.3	Local alternative	50
2.2.4	Non-asymptotic test	53
2.3	Simulation study	56
2.3.1	Asymptotic level study	56
2.3.2	Rejection power study	58
2.4	Application to real data	60
2.4.1	Analysis of some arrival times in a call center	60
2.4.2	Analysis of the scoring times of a soccer team	61
2.5	Proof of Theorem 2.1	62
2.5.1	Auxiliary results	62
2.5.2	Proof of Theorem 2.1	65
2.6	Proof of corollary 2.2	66
2.7	Proof of Theorem 2.3	68
2.7.1	Auxiliary results	68
2.7.2	Proof of Theorem 2.3	72
2.8	Proofs of Theorems 2.4 and 2.5	73
2.8.1	Auxiliary results	73
2.8.2	Proof of Theorem 2.4	76
2.8.3	Proof of Theorem 2.5	77
3	Minimax regression estimation	79
3.1	Introduction	80
3.1.1	Regression estimation	80
3.1.2	Minimax regression for Poisson coprocess	80
3.1.3	Chaotic decomposition in the Poisson space	81
3.1.4	Organization of the chapter	82
3.2	Model	82
3.3	Minimax properties for known δ	84
3.3.1	Chaos estimator	84
3.3.2	Results	86
3.4	Adaptive properties for unknown δ	87
3.5	Proof of Theorem 3.1	88
3.5.1	Technical results	88
3.5.2	Proof of Theorem 3.1	93
3.6	Proof of Theorem 3.2	93
3.6.1	Technical results	94

3.6.2	Proof of Theorem 3.2	95
3.7	Proof of Theorem 3.3	98
3.7.1	Technical results	99
3.7.2	Proof of Theorem 3.3	105
4	Kernel estimation	107
4.1	Introduction	107
4.2	Estimation strategy and results	109
4.2.1	Estimation strategy	109
4.2.2	Results	111
4.2.3	Examples	113
4.3	Simulation study	116
4.3.1	Experimental design	116
4.3.2	Results	117
4.3.3	Cross-Validation	121
4.4	Application to real data	123
4.5	Proofs	125
4.5.1	Proof of Theorem 4.1	126
4.5.2	Proof of Proposition 4.2	131
4.5.3	Proof of Theorem 4.3	132
4.6	Auxiliary results	134
	Bibliography	141

Introduction

Sommaire

1.1	Processus de Poisson et de Cox	12
1.1.1	Processus sur un espace mesurable σ -fini	13
1.1.2	Processus en temps continu	16
1.2	Tests non-paramétriques pour les processus de Cox	21
1.2.1	Tests pour les processus de Cox	21
1.2.2	Principe du test	22
1.2.3	Alternative locale	25
1.3	Estimation minimax de la régression	27
1.3.1	Estimation de la régression dans un cadre fonctionnel	27
1.3.2	Régression minimax pour une covariable issue d'un processus de Poisson	29
1.3.3	Développement en chaos de Wiener-Itô dans l'espace de Poisson	30
1.3.4	Modèle	31
1.3.5	Estimateur des chaos	33
1.3.6	Propriétés minimax lorsque δ est connu	34
1.3.7	Propriétés adaptatives lorsque δ est inconnu	35
1.4	Estimation non paramétrique à noyaux	36
1.4.1	Estimation de l'intensité du processus de Cox	36
1.4.2	Présentation du modèle	38
1.4.3	Stratégie d'estimation	39
1.4.4	Résultats	40
1.4.5	Validation Croisée	43

1.1 Processus de Poisson et de Cox

Dans cette section, nous allons présenter les outils probabilistes utilisés dans la thèse ainsi que l’approche choisie pour répondre aux problématiques rencontrées.

Les processus ponctuels forment un domaine d’étude particulièrement vaste. Ces processus permettent de modéliser à titre d’exemple les phénomènes suivants : instants de mort dans une population, émission de particules radioactives, appels dans un central téléphonique, magnitude de tremblements de terre, ou bien encore position de champs de pétrole. Pour tous ces phénomènes, on observe une suite d’instants aléatoires où se produisent certains événements spécifiques. C’est ce qui les caractérise du point de vue de la modélisation statistique.

Une des manières les plus usuelles de décrire ces processus est d’introduire la notion d’*intensité*. Cet objet représente le potentiel du processus à générer un nouvel événement dans un futur proche à un instant donné, sachant certaines informations sur le passé du processus dont en particulier tous les instants d’arrivée d’événements passés. La notion d’intensité vient à l’origine de la théorie des *processus de Poisson* mais peut être définie pour une classe bien plus large de processus ponctuels qui « contient presque tous les processus ponctuels qui présentent intérêt pratique » [*traduction*] comme l’indique Pierre Brémaud dans l’introduction de son livre *Point Processes and Queues: Martingale Dynamics* (voir Brémaud, 1981).

La théorie des martingales, en particulier via l’introduction de la notion de filtration, permet de décrire rigoureusement l’évolution d’un phénomène aléatoire en fonction du temps. C’est un des cadres mathématiques qui permet de donner une définition rigoureuse de l’intensité.

Une autre approche consiste à s’appuyer sur la théorie de la mesure pour définir les processus ponctuels. Un processus ponctuel est alors défini comme une mesure aléatoire discrète. Cette idée s’avère mathématiquement très riche et a donné lieu à de nombreux travaux présentés par exemple dans l’ouvrage de Kerstan, Matthes et Mecke (1978) ou plus récemment dans les œuvres de Last (2016) et Last et Penrose (2016). Le Chapitre 3 de ce manuscrit présente un travail de recherche qui se place dans la continuité des travaux présentés dans ces deux derniers ouvrages.

Nous allons dans un premier temps introduire et donner quelques propriétés

simples des processus ponctuels et de Poisson dans le cadre de la théorie de la mesure. Dans le but de modéliser les phénomènes aléatoires qui évoluent dans le temps, nous présenterons ensuite les processus de Poisson à valeurs réelles et indexés par $\mathbb{R}^+$ dans le cadre de la théorie martingale. Enfin nous présenterons la notion de processus de Cox sur la demi-droite réelle.

1.1.1 Processus sur un espace mesurable σ -fini

Processus ponctuels

Un *processus ponctuel* est un outil qui permet de modéliser une collection aléatoire Z , au plus dénombrable, de points d'un *espace d'états* $\mathbb{X}$. Considérer par exemple $\mathbb{X} = \mathbb{R}^d$ en donne une bonne représentation. Sans se soucier des différents problèmes de mesurabilité, on peut se représenter Z comme une fonction qui à tout élément $\omega \in \Omega$ associe un sous-ensemble $Z(\omega)$ dénombrable de $\mathbb{X}$, où on a noté $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

De manière plus formelle, soit $(\mathbb{X}, \mathcal{X})$ un espace mesurable. Soit $\mathbf{N}(\mathbb{X}) \equiv \mathbf{N}$ l'espace contenant les mesures σ -finies χ sur $\mathbb{X}$ telles que, pour tout $B \subset \mathcal{X}$, $\chi(B) \in \mathbb{N}^+ \cup \{\infty\}$. Soit de plus $\mathcal{N}(\mathbb{X}) \equiv \mathcal{N}$ la plus petite tribu engendrée $\mathbf{N}$ telle que $\chi \mapsto \chi(B)$ soit mesurable pour tout $B \subset \mathcal{X}$. On peut alors donner la définition suivante d'un processus ponctuel.

Définition 1. *Un processus ponctuel sur $\mathbb{X}$ est une variable aléatoire η à valeurs dans $(\mathbf{N}, \mathcal{N})$, c'est-à-dire une fonction mesurable $\eta : \Omega \rightarrow \mathbf{N}$.*

Si η est un processus ponctuel sur $\mathbb{X}$ et $B \in \mathcal{X}$, on note $\eta(B)$ la fonction définie par $\omega \mapsto \eta(\omega, B) := \eta(\omega)(B)$. On déduit de la définition de η et de la tribu $\mathcal{N}$ que ces variables aléatoires sont telles que

$$\{\eta(B) = k\} = \{\omega \in \Omega : \eta(\omega, B) = k\} \in \mathcal{A}, \quad B \in \mathcal{X}, k \in \mathbb{N}. \quad (1.1)$$

À l'inverse, toute fonction $\eta : \Omega \rightarrow \mathbf{N}$ est un processus ponctuel si l'équation (1.1) est vérifiée. Dans ce cas, $\eta(B)$ représente le *nombre de points* de η qui tombent dans B .

Exemple. On remarquera que si X est un élément aléatoire de $\mathbb{X}$, alors la mesure de Dirac δ_X définie par $\delta_X(B) := \mathbf{1}_B(X)$ est un processus ponctuel sur $\mathbb{X}$. Si on note de plus $\mathbb{Q}$ une mesure de probabilité sur $\mathbb{X}$ et qu'on se

donne $X_1, \dots, X_m$ des éléments aléatoires de $\mathbb{X}$ et de loi $\mathbb{Q}$, alors

$$\eta := \delta_{X_1} + \dots + \delta_{X_m},$$

est un processus ponctuel sur $\mathbb{X}$.

Dans cet exemple, la mesure aléatoire η s'écrit comme une somme de mesures de Dirac. On peut formaliser la classe des processus ponctuels ayant cette propriété de la manière suivante :

Définition 2. *On dira qu'un processus ponctuel η sur $\mathbb{X}$ est un processus ponctuel propre s'il existe une suite d'éléments aléatoires $X_1, X_2, \dots$ de $\mathbb{X}$ et une variable aléatoire κ à valeurs dans $\mathbb{N} \cup \{\infty\}$ telles que*

$$\eta = \sum_{n=1}^{\kappa} \delta_{X_n}, \quad p.s.$$

Un processus ponctuel propre s'interprète comme un ensemble aléatoire dénombrable de points de $\mathbb{X}$ (potentiellement répétés) ce qui était la première intuition de ce que devait représenter un processus ponctuel, plutôt qu'une mesure à valeur entière comme le laisse entendre la Définition 2. Pour un processus ponctuel (localement fini) sans valeur répétée, il est alors naturel de penser à définir le processus de comptage associé au processus ponctuel propre, ce que nous ferons dans la sous-section 1.1.2 dans le cas où $\mathbb{X} = \mathbb{R}$ et où le processus est indexé sur $\mathbb{R}^+$.

L'objet le plus naturel permettant de caractériser un processus ponctuel est la moyenne du nombre de ses points tombant dans un ensemble mesurable quelconque. C'est ce qu'on appelle la *mesure moyenne* du processus :

Définition 3. *La mesure moyenne d'un processus ponctuel η sur $\mathbb{X}$ est la mesure Λ définie par*

$$\Lambda(B) := \mathbb{E}[\eta(B)], \quad B \in \mathcal{X}.$$

Soit $\mathbb{R}(\mathbb{X})$ l'ensemble des fonctions mesurables $u : \mathbb{X} \rightarrow \mathbb{R}$.

Proposition 1.1 (Formule de Campbell). *Soit η un processus ponctuel sur $(\mathbb{X}, \mathcal{X})$ de mesure moyenne Λ . Soit $u \in \mathbb{R}(\mathbb{X})$. Alors $\int u(x)\eta(dx)$ est une variable aléatoire, telle que*

$$\mathbb{E}\left[\int u(x)\eta(dx)\right] = \int u(x)\Lambda(dx),$$

lorsque $u \geq 0$ ou $\int |u(x)|\Lambda(dx) < \infty$.

Nous aurons l'occasion de revenir sur cette proposition dans la sous-section suivante lorsque nous définirons le compensateur d'un processus de Poisson sur $\mathbb{X} = \mathbb{R}$.

Processus de Poisson

Un *processus de Poisson* est un processus ponctuel pour lequel le nombre de points qui tombent dans un ensemble donné suit une *loi de Poisson* et le nombre de points qui tombent dans des ensembles disjoints sont stochastiquement indépendants. Ces processus sont définis sur des espaces de *mesures s-finies*, i.e. des mesures qui s'écrivent comme une somme dénombrable de mesures finies.

Soit η un processus ponctuel propre tel que κ suit une loi de Poisson de paramètre $\lambda \geq 0$ (voir (1.3) ci-après). On peut alors aisément montrer que pour toute fonction mesurable $u : \mathbb{X} \rightarrow \mathbb{R}^+$,

$$\mathbb{E} \exp \left[- \int u(x) \eta(dx) \right] = \exp \left[- \int (1 - e^{-u(x)}) \Lambda(dx) \right], \quad (1.2)$$

où $\Lambda := \lambda \mathbb{Q}$ est la mesure moyenne de η . Des arguments classiques permettent de montrer que η est alors un processus à *accroissements indépendants*, i.e. les variables aléatoires $\eta(B_1), \dots, \eta(B_m)$ sont stochastiquement indépendants lorsque $B_1, \dots, B_m \in \mathcal{X}$ sont deux à deux disjoints. De plus, $\eta(B)$ suit une *loi de Poisson* de paramètre $\Lambda(B)$, i.e.

$$\mathbb{P}(\eta(B) = k) = \frac{\Lambda(B)^k}{k!} e^{-\Lambda(B)}, \quad k \in \mathbb{N}. \quad (1.3)$$

Il est donc naturel de définir les processus de Poisson de la manière suivante :

Définition 4. Soit Λ une mesure s-finie sur $\mathbb{X}$. Un processus de Poisson de mesure moyenne Λ est un processus ponctuel η sur $\mathbb{X}$ qui vérifie les propriétés suivantes :

- (i) Pour tout $B \in \mathcal{X}$, $\eta(B)$ suit une loi de Poisson de paramètre $\Lambda(B)$.
- (ii) Les variables aléatoires $\eta(B_1), \dots, \eta(B_m)$ sont indépendantes lorsque les ensembles $B_1, \dots, B_m \in \mathcal{X}$ sont deux à deux disjoints.

Une propriété classique des processus de Poisson ainsi définis est la formule de Mecke (voir Mecke, 1967) :

$$\mathbb{E} \int h(\eta, x) \eta(dx) = \mathbb{E} \int h(\eta + \delta_x, x) \mu(dx),$$

pour toute fonction mesurable $h : \mathbf{N} \times \mathbb{X} \rightarrow [0, \infty]$. Cette formule permet, entre autres, de construire une isométrie entre l'espace des fonctions mesurables et de carré intégrable d'un processus de Poisson η et l'espace de Fock associé à la mesure moyenne de η . Cette remarque est détaillée dans l'article de Last et Penrose (2011). Le Chapitre 3 de cette thèse se base en grande partie sur celle-ci pour construire un estimateur minimax de la régression sur un processus de Poisson.

1.1.2 Processus en temps continu

Nous nous intéressons maintenant à l'étude d'un phénomène aléatoire qui évolue avec le temps. Un tel phénomène est modélisé par un *processus stochastique* et une *filtration*. Le processus stochastique décrit l'évolution du phénomène en fonction du temps. Il est donc représenté par une famille de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Le plus souvent, un processus stochastique sera écrit sous la forme $X = (X_t)_{t \in \mathbb{R}^+}$. La filtration exprime l'information détenue à chaque instant par le modélisateur. Elle est donc représentée par une suite croissante de sous-tribus et écrite sous la forme $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}^+}$, où chaque sous-tribu $\mathcal{F}_t$ recense l'ensemble des événements avant l'instant $t \in \mathbb{R}^+$.

Processus ponctuels

On a vu dans la sous-section précédente qu'on peut définir un *processus ponctuel* comme une mesure aléatoire discrète (voir Définition 2). On peut aussi imaginer une définition de ces processus via le processus de comptage qui leur est naturellement associé. C'est le point de vue que nous développons ici.

Définition 5. Une suite $(T_n)_{n \geq 0}$ de variables aléatoires sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^+$ définit un processus ponctuel si $N_B = \sum_n \mathbf{1}\{T_n \in B\}$ est presque sûrement fini pour tout borélien borné B de $\mathbb{R}^+$.

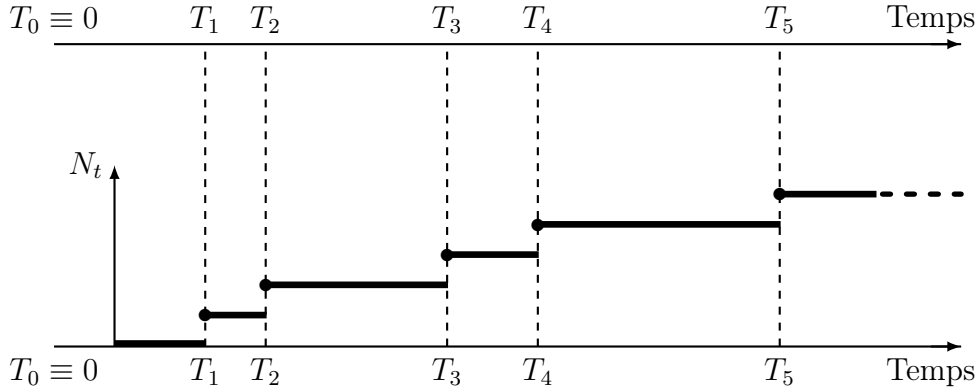


Figure 1.1: Une réalisation d'un processus ponctuel et la fonction de comptage associée.

Une réalisation d'un processus ponctuel sur $\mathbb{R}^+$ peut donc être décrite par une suite $(T_n)_{n \geq 0}$ à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$ et telle que $T_0 = 0$. De plus, si, pour tout $n \geq 0$, lorsque $T_n < \infty$, on a $T_n < T_{n+1}$, il est naturel de définir la fonction de comptage N_t associée à la réalisation $(T_n)_{n \geq 0}$ de la manière suivante

$$N_t = \begin{cases} n & \text{si } t \in [T_n, T_{n+1}[, n \geq 0, \\ +\infty & \text{sinon.} \end{cases}$$

On a alors défini une fonction en escalier continue à droite telle que $N_0 = 0$, et ses sauts sont d'amplitude 1 (voir Figure 1.1).

Nous pouvons utiliser cette remarque pour définir les processus de comptage de la manière suivante :

Définition 6. Soit N un processus stochastique. On dit que N est un processus de comptage sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ si, pour $\mathbb{P}$ -presque tout $\omega \in \Omega$, la trajectoire $t \mapsto N_t(\omega)$ est croissante par sauts d'amplitude 1, continue à droite et telle que $N_0(\omega) = 0$.

Du point de vue du processus de comptage N , le processus $(T_n)_{n \geq 0}$ est la suite de ses *instants de sauts*. Elle vérifie $0 = T_0 < T_1 < T_2 < \dots$ p.s. du fait que les sauts de N sont d'amplitude 1. On en déduit la représentation

suivante du processus de comptage N , pour tout $t \in \mathbb{R}^+$:

$$N_t = \sum_{n \geq 1} \mathbf{1}\{T_n \leq t\}.$$

Dans cette représentation, la variable T_n est appelée *n-ème instant d'arrivée* ou *de saut* de N , et l'accroissement $T_n - T_{n-1}$ est son *n-ème instant d'inter-arrivée* ou *d'inter-saut*. La relation fondamentale

$$N_t \geq n \Leftrightarrow T_n \leq t, \quad (1.4)$$

nous indique que connaître le processus ponctuel revient à connaître son processus de comptage associé. Les processus N et $(T_n)_{n \geq 0}$ portent donc la même information.

Notons $\mathcal{F}$ la *filtration naturelle* associée à N , i.e. $\mathcal{F}_t = \sigma(N_s, s \leq t)$ pour tout $t \in \mathbb{R}^+$. On déduit de la relation fondamentale (1.4) que cette filtration peut s'exprimer en faisant intervenir le processus ponctuel $(T_n)_{n \geq 0}$ de la manière suivante :

$$\mathcal{F}_t = \sigma(T_n \wedge t, \mathbf{1}\{T_n = t\}; n \geq 1).$$

On peut de plus, pour tout $t \in \mathbb{R}^+$, exprimer la loi de N_t en fonction de celle des instants d'arrivée :

$$\mathbb{P}(N_t = n) = F_n(t) - F_{n+1}(t),$$

où F_n désigne la fonction de répartition du n -ème instant de saut T_n .

Pour décrire la dynamique d'un processus de comptage N , il est utile de définir son intensité. Il s'agit d'un objet sensé représenter le potentiel du processus à générer un nouvel évènement dans un futur proche à un instant donné, sachant certaines informations sur le passé du processus dont en particulier tous les instants d'arrivée d'évènements passés.

Définition 7. *Soit N un processus de comptage. On appelle compensateur de N un processus nul en 0, croissant, et prévisible Λ tel que $N - \Lambda$ est une martingale locale. Si $\mathbb{P}$ -presque toute trajectoire de Λ est absolument continue, on appelle intensité de N sa dérivée λ .*

La terminologie de compensateur tire son origine du fait qu'il stabilise, ou compense le processus de comptage, en extrayant sa partie non martingale. Il est possible de montrer un résultat similaire à la Formule de Campbell (voir Proposition 1.1) dans le cadre martingale :

Théorème 1.2. *Soit N un processus de comptage d'intensité λ et de compensateur Λ . On note $M = N - \Lambda$. Soit ξ un processus prévisible.*

- (i) *Si pour tout $t \geq 0$, $\mathbb{E}[\int_0^t |\xi_s| \lambda_s ds] < \infty$, alors $(\int_0^t \xi_s dM_s)_{t \in \mathbb{R}^+}$ est une martingale;*
- (ii) *Si pour tout $t \geq 0$, $\int_0^t |\xi_s| \lambda_s ds < \infty$, alors $(\int_0^t \xi_s dM_s)_{t \in \mathbb{R}^+}$ est une martingale locale.*

Processus de Cox et de Poisson

Un *processus de Poisson* est un processus de comptage dont les accroissements sont indépendants et suivent une loi de Poisson. Un *processus de Cox* est aussi appelé *processus de Poisson doublement stochastique*. La terminologie de processus de Poisson doublement stochastique a été d'abord introduite par Cox (1955) et tire son origine du fait que le processus de Cox peut être vu comme issu d'une procédure de double randomisation.

Définition 8. *Soit N un processus de comptage, $\mathcal{F}$ la filtration naturelle associée à N et Λ son compensateur. On dit que N est un processus de Poisson doublement stochastique, aussi appelé processus de Cox si, conditionnellement à Λ , la loi de N est celle d'un processus de Poisson de mesure moyenne Λ .*

Dans le cas où le processus Λ peut s'écrire comme l'intégrale d'une intensité λ , on dit que N est un processus de Cox d'intensité λ . Si de plus λ est déterministe, le processus N est un *processus de Poisson* (inhomogène) d'intensité λ . Le cas où λ est un processus stochastique non dégénéré est porteur de nombreuses améliorations dans les applications : par exemple, une compagnie d'assurance voulant modéliser le processus d'arrivée des sinistres doit prendre en compte des facteurs observés supplémentaires, de natures géographiques, climatiques ou bien des données personnelles de ses assurés, ceci pour améliorer les performances de son modèle. De même, dans le domaine des essais cliniques, l'introduction de covariables telles que l'âge, le sexe ou certaines caractéristiques physiologiques interviennent aussi, au titre de variables explicatives, pour améliorer le modèle.

Un exemple d'intensité souvent rencontré dans la littérature est l'intensité dans le modèle à taux proportionnel. Dans ce modèle, l'intensité semi-

paramétrique est du type

$$\lambda(t) = \lambda_0(t)e^{\beta^\top Z_t}, \quad (1.5)$$

où λ_0 est une fonction déterministe et $\beta \in \mathbb{R}^d$ est un paramètre associé aux effets des covariables Z_t . Dans ce modèle, le paramètre β décrit de quelle manière l'intensité réagit face à la variable explicative; elle est donc une caractéristique essentielle du modèle. L'un des intérêts de ce modèle à risque proportionnel est qu'il est possible d'estimer alors β indépendamment de λ_0 . Cette observation, dont la paternité revient à Cox (1955), est à l'origine de la popularité de ce modèle; on trouvera par exemple dans l'ouvrage de Bening et Korolev (2002) une synthèse relativement exhaustive de la littérature particulièrement abondante qui traite de ce sujet.

Le résultat suivant, introduit par Watanabe en 1964, donne une caractérisation des processus de Poisson par des propriétés martingales.

Théorème 1.3 (Watanabe). *Soit N un processus de comptage, $\mathcal{F}$ la filtration naturelle associée à N et λ son intensité. Alors, $(N_t - \int_0^t \lambda_s ds)_{t \geq 0}$ est une martingale si et seulement si N est un processus de Poisson d'intensité λ .*

Le théorème suivant peut être vu comme une extension du théorème de Watanabe aux processus de Cox :

Théorème 1.4 (Caractérisation des processus de Cox). *Soit N un processus de comptage, $\mathcal{F}$ la filtration naturelle associée à N et λ son intensité. Supposons que λ est telle que pour tout $t \geq 0$, λ_t est $\mathcal{F}_0$ -mesurable et $\int_0^t \lambda_s ds < \infty$, $\mathbb{P}$ -p.s. Si on a*

$$\mathbb{E} \left[\int_0^\infty C_s dN_s \right] = \mathbb{E} \left[\int_0^\infty C_s \lambda_s ds \right],$$

pour tout processus prévisible positif C , alors N est un processus de Cox d'intensité λ .

Enfin, le résultat suivant illustre la position centrale qu'occupe le processus de Poisson standard dans la théorie des processus de comptage.

Théorème 1.5. *Soit N un processus ponctuel d'intensité λ par rapport à la filtration $\mathcal{F}$ et d'intensité $\tilde{\lambda}$ par rapport à la filtration $\mathcal{G}$, où les filtrations $\mathcal{F}$ et $\mathcal{G}$ sont telles que $\mathcal{F}^N \subseteq \mathcal{G} \subset \mathcal{F}$, où $\mathcal{F}^N$ désigne la filtration naturelle de N . Supposons que $N_\infty = \infty$, p.s. Si on note pour tout t , $\tau(t)$ le $\mathcal{G}_t$ -temps d'arrêt tel que*

$$\int_0^{\tau(t)} \tilde{\lambda}_s ds = t,$$

alors le processus ponctuel $\tilde{N}$ défini par $\tilde{N}_t = N_{\tau(t)}$ est un processus de Poisson standard (d'intensité 1).

Les preuves des Théorèmes 1.4 et 1.5 sont disponibles dans l'ouvrage de Brémaud (1981).

1.2 Tests non-paramétriques pour les processus de Cox

1.2.1 Tests pour les processus de Cox

Considérons l'exemple d'une compagnie d'assurance qui recense les sinistres subis par ses assurés. Alors que le principe de parcimonie nous encouragerait à considérer un simple processus de Poisson, il peut être tentant d'ajuster le modèle sur un nombre trop important de covariables : l'âge ou le sexe de l'assuré pour un particulier, le secteur d'entreprise ou le nombre de salarié pour une entreprise, l'âge du véhicule ou sa puissance dans l'assurance automobile, ou encore des informations géographiques comme le revenu moyen dans la commune, la densité de population, etc. Il serait alors intéressant de pouvoir tester l'influence des covariables dans le modèle pour pouvoir sélectionner les variables superflues. De manière plus précise, en se basant sur des copies i.i.d. du processus de Cox N , on va chercher à construire un test d'hypothèse $\mathbf{H}_0 : N$ est un processus de Poisson vs $\mathbf{H}_1 : N$ n'est pas un processus de Poisson.

Dans ce contexte, sans plus de détails, une approche naïve consiste à estimer les fonctions $t \mapsto \mathbb{E}[N_t \mid \Lambda]$ et $t \mapsto \mathbb{E}N_t$ pour ensuite tester leur égalité. On sait en effet que sous l'hypothèse $\mathbf{H}_0$, la première quantité devrait être égale à la seconde, ce qui ne sera pas le cas sous $\mathbf{H}_1$. Cette approche à

deux principaux désavantages : le fléau de la dimension (dès que Λ est à valeurs dans un espace de grande dimension) et la connaissance a priori de Λ . Une autre approche consiste à tester si la loi des temps de saut du processus N est celle d'un processus de Poisson ; c'est l'approche utilisée par exemple dans l'article de Reynaud-Bourret et al. (2014), dans lequel les auteurs construisent une statistique de Kolmogorov-Smirnov modifiée pour répondre à la problématique.

Dans l'exemple classique du modèle à risque proportionnel (1.5), le processus de Cox N dépend de la covariable Z qui est un processus stochastique à valeurs dans $\mathbb{R}^d$. Appelons $Z^1, \dots, Z^d$ ses coordonnées. Tester l'influence de la covariable Z^i , ou plus généralement l'influence d'un sous-modèle $(Z^i)_{i \in \mathcal{I}}$ avec $\mathcal{I} \subset \{1, \dots, d\}$, revient à considérer le test d'hypothèse suivant :

$$\mathbf{H}_0 : \beta_i = 0 \ \forall i \in \mathcal{I} \quad \text{contre} \quad \mathbf{H}_1 : \beta_i \neq 0 \text{ pour un } i \in \mathcal{I}.$$

On peut alors se reporter aux procédures d'estimation du paramètre β introduites par Cox (1955) mentionnées précédemment pour construire les statistiques de test correspondantes.

Dans le Chapitre 2, nous présentons notre contribution sur les tests pour les processus de Cox. Nous proposons deux statistiques de test qui permettent, en se basant sur l'observation de n trajectoires du processus, de révéler la nature Poissonienne d'un processus de Cox. Les procédures de test ainsi construites ont le double avantage pratique d'être très simple à mettre en œuvre et de ne nécessiter aucune connaissance a priori sur la forme de la mesure d'intensité du processus de Cox. On montre de plus que les tests ainsi définis sont de puissance asymptotiquement maximale.

1.2.2 Principe du test

Fixons $T > 0$ la durée (déterministe) des observations, et $N = (N_t)_{t \in [0, T]}$ le processus de Cox de mesure d'intensité $\Lambda = (\Lambda(t))_{t \in [0, T]}$, tel que $\mathbb{E}N_T^4 < \infty$, et, pour un certain $t \in]0, T[$, $\mathbb{E}N_t \neq 0$. On remarquera que, sous ces hypothèses, la fonction $t \mapsto \mathbb{E}N_t = \mathbb{E}\Lambda(t)$ n'est pas nécessairement absolument continue. Le modèle n'est donc pas réduit au cas où il existe une fonction d'intensité λ telle que Λ puisse s'écrire comme l'intégrale de cette intensité.

Notons m et σ^2 les fonctions moyenne et variance du processus N , i.e. pour tout $t \in [0, T]$:

$$m(t) = \mathbb{E}N_t \text{ et } \sigma^2(t) = \text{var}(N_t).$$

Alors pour tout $t \in [0, T]$ (voir p. 66 du livre de Kingman, 1993) :

$$\sigma^2(t) = m(t) + \text{var}(\mathbb{E}[N_t \mid \Lambda]) = m(t) + \text{var}(\Lambda(t)).$$

On en déduit que N est un processus de Poisson si et seulement si $m = \sigma^2$. En se basant sur cette observation, on peut réécrire le problème de test d'hypothèse $\mathbf{H}_0$: N est un processus de Poisson vs $\mathbf{H}_1$: N n'est pas un processus de Poisson de la manière suivante :

$$\mathbf{H}_0 : \sigma^2 = m \text{ vs } \mathbf{H}_1 : \exists t \leq T \text{ tel que } \sigma^2(t) > m(t).$$

Soient les données $N^{(1)}, \dots, N^{(n)}$ des copies indépendantes du processus N . Des statistiques de test naturelles pour le problème de test ci-dessus s'expriment en fonction du processus $\hat{\sigma}^2 - \hat{m} = (\hat{\sigma}^2(t) - \hat{m}(t))_{t \in [0, T]}$ où $\hat{m}$ et $\hat{\sigma}^2$ sont les contreparties empiriques de m et σ^2 :

$$\hat{m}(t) = \frac{1}{n} \sum_{i=1}^n N_t^{(i)} \text{ et } \hat{\sigma}^2(t) = \frac{1}{n-1} \sum_{i=1}^n (N_t^{(i)} - \hat{m}(t))^2.$$

Notre premier résultat permet d'obtenir la loi asymptotique du processus $\hat{\sigma}^2 - \hat{m}$ sous $\mathbf{H}_0$.

Théorème 1.6. *Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un mouvement Brownien standard sur l'ensemble des réels. Sous $\mathbf{H}_0$, $\hat{\sigma}^2 - \hat{m}$ est une martingale et*

$$\sqrt{n} (\hat{\sigma}^2 - \hat{m}) \xrightarrow{(\text{loi})} (B_{2m(t)^2})_{t \leq T}.$$

Ce résultat est démontré dans le Chapitre 2. À notre connaissance, la propriété martingale de $\hat{\sigma}^2 - \hat{m}$ n'avait pas été observée jusqu'alors. Cette propriété, intéressante en elle-même, joue un rôle majeur dans l'étude asymptotique des statistiques de test naturelles que nous présentons dans la suite.

Nous nous concentrons principalement sur deux statistiques de test qui permettent de détecter si un processus de Cox est un processus de Poisson

ou non. Notons $\hat{S}_1$ et $\hat{S}_2$ ces statistiques. Nous les définissons de la manière suivante :

$$\hat{S}_1 = \sup_{t \leq T} (\hat{\sigma}^2(t) - \hat{m}(t)), \text{ et } \hat{S}_2 = \int_0^T (\hat{\sigma}^2(t) - \hat{m}(t)) dt.$$

On reconnaît deux statistiques de test naturelles pour tester la nullité d'une fonction positive. Comme l'hypothèse alternative peut s'écrire $\mathbf{H}_1$: il existe $t \leq T$ tel que $\sigma^2(t) > m(t)$, elles sont de plus compatibles avec la nature unilatérale du problème d'intérêt.

Présentons maintenant les propriétés asymptotiques de $\hat{S}_1$ et $\hat{S}_2$.

Corollaire 1.7. *Soit $\hat{I}^2 = \int_0^T (T-t)\hat{m}(t)^2 dt$.*

(i) *Sous $\mathbf{H}_0$,*

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{(\text{loi})} |\mathcal{N}(0, 2)|, \text{ et } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{(\text{loi})} \mathcal{N}(0, 4).$$

(ii) *Sous $\mathbf{H}_1$,*

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{\text{prob.}} +\infty, \text{ et } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{\text{prob.}} +\infty$$

On déduit de ce corollaire que les statistiques de test $\hat{S}_1/\hat{m}(T)$ et $\hat{S}_2/\hat{I}$ définissent toutes deux des tests de puissance asymptotiquement maximale. Les régions de rejet asymptotiques de ces tests pour un niveau $\alpha \in]0, 1[$ s'écrivent alors:

$$R_1(\alpha) = \left\{ \frac{\hat{S}_1}{\hat{m}(T)} \geq \sqrt{\frac{2}{n}} q_{1-\alpha/2} \right\} \text{ et } R_2(\alpha) = \left\{ \frac{\hat{S}_2}{\hat{I}} \geq \frac{2}{\sqrt{n}} q_{1-\alpha} \right\}, \quad (1.6)$$

où $\alpha \in]0, 1[$ et $q_{1-\alpha}$ est le $\mathcal{N}(0, 1)$ -quantile d'ordre $1 - \alpha$.

Un cadre plus général pourrait être atteint. En effet, la preuve du Corollaire 1.7 n'utilise que l'hypothèse que le processus N appartient à une classe de processus de comptages surdispersés (i.e. tels que $\text{var}(N_t) \geq \mathbb{E}N_t$ pour tout $t \in [0, T]$) et qui vérifient : $\text{var}(N_t) = \mathbb{E}N_t$ pour tout $t \in [0, T]$ si et seulement si N est un processus de Poisson. Le processus de Cox est simplement un archétype de cette classe de processus de comptages. Cette propriété est aussi vérifiée par d'autres classes de processus, comme certaines

sous-classes de processus de Hawkes par exemple. Le test ainsi décrit est donc plus ou moins une version fonctionnelle du test classique de surdispersion, utilisé pour tester la loi de Poisson de données de comptage (voir par exemple les articles de Rao et Chakravarti, 1956 ou Bohning, 1994). Les tests de surdispersion sont également largement utilisés pour l'étude du nombre de sinistres en actuariat (e.g. Denuit et al., 2007).

1.2.3 Alternative locale

Nous avons vu que les statistiques $\hat{S}_1$ et $\hat{S}_2$ définissent des tests de puissance asymptotiquement maximale pour $\mathbf{H}_0$ vs $\mathbf{H}_1$. Nous souhaitons maintenant comparer les deux tests de manière théorique.

Une des façons les plus populaires de caractériser les propriétés asymptotiques de statistiques de test lorsque l'hypothèse nulle $\mathbf{H}_0$ est fausse est de considérer une situation artificielle, représentée par l'*alternative locale* $\mathbf{H}_1^n$, sous laquelle le processus générateur des données change dès lors que la taille d'échantillon augmente. En particulier, il est alors nécessaire que la distance entre l'hypothèse nulle $\mathbf{H}_0$ et l'alternative $\mathbf{H}_1^n$ diminue lorsque la taille de l'échantillon tend vers l'infini. La puissance asymptotique liée à l'alternative locale ainsi définie est alors une quantité très utile pour comparer des tests entre eux. Les œuvres de Engle (1984) et van der Vaart (2000) sont des ouvrages de référence sur ce sujet.

Dans la suite, nous faisons l'hypothèse supplémentaire suivante sur le processus N : le processus de Cox N admet une intensité $\lambda = (\lambda(t))_{t \in [0, T]}$, i.e. avec probabilité 1, Λ est absolument continu, et

$$\Lambda(t) = \int_0^t \lambda(s) ds, \quad \forall t \in [0, T].$$

Soit $(d_n)_n$ une suite de nombres positifs qui décroît vers 0. Une manière classique de définir l'alternative locale est de considérer un *Pitman drift* (voir Davidson et MacKinnon, 2006, et les références qui s'y trouvent), qui permet de spécifier l'intensité comme une déviation locale de l'hypothèse nulle. L'alternative locale est alors définie de la manière suivante :

$\mathbf{H}_1^n$: Il existe une fonction non-nulle, $\lambda_0 : [0, T] \rightarrow \mathbb{R}_+$ et un processus

$$\Delta = (\Delta_t)_{t \in [0, T]} \text{ avec } \sup_{t \in [0, T]} \mathbb{E} \Delta_t^6 < \infty \text{ et } \text{var} \left(\int_0^{t_0} \Delta_s ds \right) > 0$$

pour un $t_0 \in [0, T]$, et tels que $\lambda = \lambda_0 + d_n \Delta$ avec probabilité 1.

On déduit des deux hypothèses sur le processus Δ que l'alternative locale $\mathbf{H}_1^n$ est contenue dans l'hypothèse $\mathbf{H}_1$.

De plus, pour n grand, lorsque l'hypothèse $\mathbf{H}_1^n$ est vérifiée, le processus N se rapproche d'un processus de Poisson. Ainsi, la suite $(d_n)_n$ peut être vue comme une vitesse de séparation entre l'hypothèse alternative $\mathbf{H}_1^n$ et l'hypothèse nulle $\mathbf{H}_0$. On obtient en particulier, le résultat suivant qui montre que pour pouvoir construire un test consistant de $\mathbf{H}_0$ vs $\mathbf{H}_1^n$, il est nécessaire et suffisant que d_n^2 tende vers 0 plus lentement que $1/\sqrt{n}$.

Théorème 1.8. *Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un mouvement Brownien standard sur l'ensemble des réels. Supposons que $\mathbf{H}_1^n$ est vérifiée, et notons m_0 et v les fonctions définies pour tout $t \in [0, T]$ par*

$$m_0(t) = \int_0^t \lambda_0(s) ds, \text{ et } v(t) = \text{var} \left(\int_0^t \Delta_s ds \right).$$

Notons de plus $I_0^2 = \int_0^T (T-t)m_0(t)^2 dt$.

(i) Si $\sqrt{n} d_n^2 \rightarrow \infty$, alors

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{\text{prob.}} +\infty, \text{ et } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{\text{prob.}} +\infty.$$

(ii) Si $\sqrt{n} d_n^2 \rightarrow d < \infty$, alors

$$\begin{aligned} \sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} &\xrightarrow{(\text{loi})} \frac{1}{m_0(T)} \sup_{t \leq T} (B_{2m_0(t)^2} + dv(t)), \text{ et} \\ \sqrt{n} \frac{\hat{S}_2}{\hat{I}} &\xrightarrow{(\text{loi})} 2\mathcal{N}(0, 1) + \frac{d}{I_0} \int_0^T v(t) dt. \end{aligned}$$

Considérons les tests induits par les régions de rejet définies en (1.6) dans le problème $\mathbf{H}_0$ vs $\mathbf{H}_1^n$ pour un niveau $\alpha \in]0, 1[$. En se basant sur le résultat du Théorème 1.8, l'étude d'un exemple détaillé dans le Chapitre 2 permet de penser que le test induit par la région de rejet $R_1(\alpha)$ est plus puissant que le second. Une étude par simulation et deux applications sur données réelles confirment les résultats ici présentés.

1.3 Estimation minimax de la régression sur une covariable issue d'un processus de Poisson

1.3.1 Estimation de la régression dans un cadre fonctionnel

L'estimation de la régression est une problématique centrale de la statistique. Elle est largement utilisée et fait l'objet de nombreuses études dans la littérature. Parmi les différentes méthodes étudiées pour répondre au problème de l'estimation de la régression, l'estimation non-paramétrique a été examinée de manière très précise (on renvoie par exemple le lecteur à l'ouvrage de Tsybakov, 2009 pour une introduction très complète sur l'estimation non-paramétrique de manière générale et au livre de Györfi et al., 2006 pour un résumé clair des méthodes d'estimation non-paramétrique pour la régression). Plus récemment, la communauté scientifique s'est intéressée au problème d'estimation de la régression sur des covariables fonctionnelles (les livres de Ramsay et Silverman, 2006 et Horváth et Kokoszka, 2012 résument avec précision plusieurs méthodes d'analyse de données fonctionnelles déjà étudiées). Le sujet de ces travaux est l'estimation de la fonction de régression $r(\cdot) = \mathbb{E}(y \mid X = \cdot)$ où $y \in \mathbb{R}$ et X est à valeur dans un espace fonctionnel à partir de l'observation d'un n -échantillon du couple (X, Y) .

Plusieurs approches s'offrent alors aux statisticiens. L'une d'elles consiste à souligner le rôle majeur des probabilités dites de petites boules $\mathbb{P}(\|X - x_0\| < \cdot)$ dans la variance des estimateurs classiques. On notera dans cette idée l'article de Mas (2012) dans lequel l'auteur montre une borne inférieure minimax pour le risque ponctuel au point x_0 lorsque la covariable X appartient à un espace fonctionnel hilbertien $\mathcal{H}$. En définissant

le risque minimax de la manière suivante :

$$\mathcal{R}_n^* = \inf_{T_n} \sup_{r \in \mathcal{E}} \mathbb{E} |T_n - r(x_0)|^2,$$

où T_n est n'importe quel estimateur de $r(x_0)$ et $\mathcal{E}$ est une classe de fonctions régulières de $\mathcal{H}$ dans $\mathbb{R}$, l'auteur montre que pour tout $\tau > 0$, $n^\tau \mathcal{R}_n^* \rightarrow +\infty$, mettant ainsi en évidence le « fléau de la dimension » dans les espaces infini dimensionnels généraux.

Toujours dans cette idée d'exploiter l'importance des probabilités de petites boules dans la variance des estimateurs ponctuels de la régression sur une covariable fonctionnelle, l'article de Chagny et Roche (2016) propose une procédure d'estimation adaptative de la fonction de régression r . En examinant le risque ponctuel d'une collection d'estimateurs à noyaux de type Nadaraya-Watson, les auteures définissent un critère de sélection de la fenêtre entièrement guidé par les données et optimal au sens de l'oracle. Elles définissent pour ce faire pour tout point x_0 , le risque d'un estimateur $\hat{r}(x_0)$ calculé en x_0 par

$$\mathbb{E}(\hat{r}(x_0) - r(x_0))^2,$$

et obtiennent une borne supérieure non asymptotique pour ce risque qui dépend de la régularité de la fonction de régression et de la probabilité de petites boules du processus X de manière usuelle. Les vitesses obtenues, bien que lentes, sont comparables à celles obtenues dans l'article de Mas (2012).

Une autre approche, présentée par exemple dans l'article de Biau, Cérou et Guyader (2010) consiste à considérer l'estimateur non-paramétrique $r_n(x)$ des k -plus proches voisins de la fonction de régression $r(x) = \mathbb{E}(Y \mid X = x)$ lorsque le couple de variables aléatoires (X, Y) prend ses valeurs dans $\mathcal{F} \times \mathbb{R}$ où $\mathcal{F}$ est un espace de Banach séparable. Rappelons que, pour tout x de $\mathcal{F}$, l'estimateur des k_n -plus proches voisins est défini par

$$r_n(x) = \frac{1}{k_n} \sum_{i=1}^{k_n} Y_{(i,n)}(x),$$

où $(X_{(1,n)}(x), Y_{(1,n)}(x)), \dots, (X_{(k_n,n)}(x), Y_{(k_n,n)}(x))$ est le réordonnement des données par ordre croissant des quantités $\|X_i - x\|$. Les auteurs montrent dans un premier temps que le risque quadratique de cet estimateur

$\mathbb{E}(r_n(X) - r(X))^2$ est supérieurement borné par un terme de variance en $1/k_n$ et un terme de biais qui dépend du nombre de ε -recouvrements du support de la distribution de X , i.e. le plus petit nombre de boules ouvertes de rayon ε permettant de recouvrir cet espace. Le plongement du support de la distribution de X dans un sous-espace de Banach de $\mathcal{F}$ permet de contrôler supérieurement ces ε -recouvrements et d'obtenir une borne supérieure minimax pour le risque décrit plus haut.

Dans les travaux précédemment cités les auteurs pointent que les vitesses obtenues en dimension infinie sont toujours plus lentes que les vitesses traditionnelles obtenues en dimension finie.

Dans le Chapitre 3, nous nous intéressons au problème de régression sur une covariable issue d'un processus de Poisson. Le but de notre contribution est de construire un nouvel estimateur asymptotiquement minimax de la fonction de régression r basé sur des copies indépendantes du couple (X, Y) . Dans l'esprit de l'article de Cadre et Truquet (2015), nous proposons une méthode d'estimation qui se base sur la décomposition en chaos des fonctionnelles de Poisson. Sous une hypothèse de finitude du développement en chaos de Wiener-Itô discutée plus tard, nous définissons un estimateur qui s'adapte au nombre de chaos. Nous montrons que, à un facteur logarithmique près, notre estimateur est asymptotiquement minimax lorsque le nombre de chaos est connu. La vitesse de convergence obtenue est comparable à celles obtenues classiquement en dimension finie (voir, e.g., Theorem 1 de Kohler et al., 2009). On retrouve finalement la même borne supérieure pour le risque dans le cas où le nombre de chaos est inconnu.

1.3.2 Régression minimax pour une covariable issue d'un processus de Poisson

Nous commençons par présenter le problème de régression pour une covariable issue d'un processus de Poisson dans un cadre minimax.

Soit X un processus de Poisson sur un domaine compact $\mathbb{X} \subset \mathbb{R}^d$ équipé de sa tribu Borélienne $\mathcal{X}$. Notons $\mathcal{S}$ l'espace des états du processus de Poisson X , $\mathbb{P}_X$ la loi de X et $\mathbb{L}^2(\mathbb{P}_X)$ l'espace de toutes les fonctions mesurables $g : \mathcal{S} \rightarrow \mathbb{R}$ telles que

$$\|g\|_{\mathbb{L}^2(\mathbb{P}_X)}^2 = \mathbb{E}g(X)^2 < +\infty.$$

Soit maintenant $\mathbb{P}$ une distribution de probabilité sur $\mathcal{S} \times \mathbb{R}$ et (X, Y) de loi $\mathbb{P}$. Supposons que $\mathbb{E}|Y| < +\infty$, où $\mathbb{E}$ désigne l'espérance sous $\mathbb{P}$. Notre but est d'estimer la fonction de régression r de $\mathbb{L}^2(\mathbb{P}_X)$, définie de $\mathcal{S}$ dans $\mathbb{R}$ par $r(s) = \mathbb{E}(Y \mid X = s)$ par un estimateur asymptotiquement minimax.

Nous allons maintenant définir ce que nous appelons estimateur asymptotiquement minimax. Nous disposons, pour estimer la fonction de régression r , d'un échantillon i.i.d. tiré de la loi $\mathbb{P}$ de (X, Y) . Dans ce contexte, un estimateur est une application mesurable $\tilde{r} : (\mathcal{S} \times \mathbb{R})^n \rightarrow \mathbb{L}^2(\mathbb{P}_X)$, dont la précision est mesurée par le risque

$$R_n(\tilde{r}, r) = \mathbb{E}^n \|\tilde{r} - r\|_{\mathbb{L}^2(\mathbb{P}_X)}^2,$$

où $\mathbb{E}^n$ désigne l'espérance par rapport à la loi de probabilité $\mathbb{P}^{\otimes n}$.

Le risque maximal d'un estimateur $\tilde{r}$ de r sur l'ensemble de loi $\mathcal{P}$ du couple (X, Y) est alors défini par

$$R_n(\tilde{r}, \mathcal{P}) = \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r).$$

Ainsi, un estimateur $\hat{r}$ est dit asymptotiquement minimax si il est tel que

$$R_n(\hat{r}, \mathcal{P}) \asymp \inf_{\tilde{r}} R_n(\tilde{r}, \mathcal{P}),$$

où l'infimum est pris sur l'ensemble des estimateurs $\tilde{r}$ de r et $u_n \asymp v_n$ signifie $0 < \liminf_n u_n v_n^{-1} \leq \limsup_n u_n v_n^{-1} < +\infty$.

1.3.3 Développement en chaos de Wiener-Itô dans l'espace de Poisson

Nous introduisons maintenant le développement en chaos de Wiener-Itô. Sommairement, le développement en chaos de Wiener-Itô (voir Itô, 1956 et Nualart et Vives, 1990 pour des détails techniques) assure que toute variable aléatoire de carré intégrable et mesurable par rapport à la tribu engendré par le processus de Poisson X peut être décomposée en une somme d'intégrales stochastiques, appelés chaos.

De manière plus précise, rappelons quelques propriétés élémentaires du développement en chaos de Wiener-Itô dans l'espace de Poisson. Soit μ la mesure moyenne du processus de Poisson X , définie par $\mu(A) = \mathbb{E}X(A)$

pour tout $A \in \mathcal{X}$, où $X(A)$ est le nombre de points de X qui sont dans l'ensemble A . On pose $k \geq 1$. Supposons que $g \in \mathbb{L}^2(\mu^{\otimes k})$, on peut alors définir le k -ième chaos $I_k(g)$ associé à g par

$$I_k(g) = \int_{\Delta_k} g d(X - \mu)^{\otimes k}, \quad (1.7)$$

où $\Delta_k = \{x \in \mathbb{X}^k : x_i \neq x_j \text{ pour tous } i \neq j\}$. Dans l'article de Nualart et Vives (1990), les auteurs montrent que toute variable aléatoire de carré intégrable et mesurable par rapport à la tribu engendré par X peut être décomposée en une somme de chaos. En appliquant cette remarque à notre problème de régression, on obtient

$$r(X) = \mathbb{E}Y + \sum_{k \geq 1} \frac{1}{k!} I_k(f_k), \quad (1.8)$$

où l'égalité à lieu dans $\mathbb{L}^2(\mathbb{P}_X)$, pour peu que $\mathbb{E}Y^2 < \infty$. Dans la formule précédente, chaque fonction f_k est un élément de $\mathbb{L}_{\text{sym}}^2(\mu^{\otimes k})$ qui est le sous ensemble des fonctions symétriques de $\mathbb{L}^2(\mu^{\otimes k})$, et la décomposition est définie de manière unique.

1.3.4 Modèle

Soient $\Theta \subset \mathbb{R}^p$ et $\{\varphi_\theta\}_{\theta \in \Theta}$ une famille de fonctions de $\mathbb{X} \rightarrow \mathbb{R}_+$ dont la régularité est détaillée au Chapitre 3. Soit (X, Y) , un couple de variables aléatoires à valeurs dans $\mathcal{S} \times \mathbb{R}$ et de loi $\mathbb{P}$, où $\mathcal{S}$ est l'espace de Poisson défini sur le domaine compact $\mathbb{X} \subset \mathbb{R}^d$. Ici, X est un processus de Poisson sur $\mathbb{X}$ d'intensité paramétrique φ_θ , i.e. pour tout ensemble Borélien $A \in \mathcal{X}$:

$$\mathbb{E}X(A) = \int_A \varphi_\theta d\lambda, \quad (1.9)$$

où $\mathbb{E}$ désigne l'espérance par rapport à $\mathbb{P}$. En d'autres termes, la mesure d'intensité μ de X admet une dérivée de Radon-Nikodým φ_θ par rapport à λ .

On suppose de plus qu'on peut trouver un estimateur suffisamment précis du paramètre θ et que la variable aléatoire Y vérifie une condition sur ses moments exponentiels. Ces deux conditions techniques sont détaillées et discutées au Chapitre correspondant.

Nous avons déjà présenté à l'équation (1.8) la décomposition en une somme de chaos de la fonction de régression. Nous supposons ici que cette somme est finie, i.e. il existe un entier strictement positif δ et des fonctions $f_1 \in \mathbb{L}_{sym}^2(\mu), \dots, f_\delta \in \mathbb{L}_{sym}^2(\mu^{\otimes \delta})$ tels que

$$r(X) = \mathbb{E}Y + \sum_{k=1}^{\delta} \frac{1}{k!} I_k(f_k), \quad (1.10)$$

où les $I_k(f_k)$ ont été défini par l'équation (1.7) et les coefficients des chaos f_k sont supposés appartenir à une famille non-paramétrique de fonctions régulières en un sens détaillé dans le Chapitre 3.

Remarques (Sur le caractère fini du développement en chaos de Wiener-Itô).

1. On notera que l'hypothèse de finitude du développement en chaos de Wiener-Itô de la fonction de regression $r(X)$ semble lui interdire d'être bornée. En effet, considérons pour simplifier le cas où $r(X)$ peut être décomposée sur un seul chaos, i.e.

$$r(X) = \int f d(X - \lambda).$$

Ici, X est un processus de Poisson simple sur le domaine $\mathbb{X} = [0, 1]$, d'intensité la fonction constante égale à 1 et f est une fonction λ -intégrable sur $\mathbb{X}$. Alors, si $f \geq a > 0$, on a

$$r(X) \geq aX([0, 1]) - \int f d\lambda.$$

On en déduit que $r(X)$ ne peut pas être bornée. On peut s'attendre au même type de propriété pour un nombre quelconque de chaos.

2. La Proposition 4.1 de l'article de Last et Penrose (2011), permet de montrer que la décomposition en δ chaos de $r(X)$ est possible si et seulement si la dérivée $\delta + 1$ -ième de Malliavin de r est nulle. Cette hypothèse repose donc entièrement sur la distribution du couple (X, Y) par le biais du calcul de Malliavin.

Dans la suite de cette section, on suppose fixées les constantes de régularité liées aux hypothèses énoncées ci-dessus. En particulier, le nombre de chaos

δ est supposé fixe. On notera $\mathcal{P}$ l'ensemble des lois $\mathbb{P}$ de (X, Y) qui vérifient ces hypothèses. Dans ce cadre, θ désigne implicitement la vraie valeur du paramètre, i.e. φ_θ est l'intensité de X (de mesure d'intensité μ).

1.3.5 Estimateur des chaos

Notre problématique est maintenant la suivante : nous souhaitons construire un estimateur de la régression qui atteigne des vitesses de convergence rapides sur $\mathcal{P}$. Soient $\mathbb{P} \in \mathcal{P}$ et $(X, Y) \sim \mathbb{P}$ où X admet une mesure d'intensité $\mu = \varphi_\theta \cdot \lambda$.

Soient $\overline{W}$ une constante strictement positive, W une densité sur $\mathbb{X}$ telle que $\sup_{\mathbb{X}} W \leq \overline{W}$, et

$$W_{h_k}(\cdot) = \frac{1}{h_k^d} W\left(\frac{\cdot}{h_k}\right),$$

où $h_k = h_k(n) > 0$ est une fenêtre à ajuster. On déduit aisément des propriétés du développement en chaos de Wiener-Itô que :

$$\mathbb{E} Y I_k(W_{h_k}^{\otimes k}(x - \cdot)) = \int_{\mathbb{X}^k} f_k W_{h_k}^{\otimes k}(x - \cdot) \varphi_\theta^{\otimes k} d\lambda^{\otimes k},$$

où, ici et dans la suite, pour toute fonction à valeurs réelles g définie sur $\mathbb{X}$, on a noté $g^{\otimes k}$ la fonction à valeurs réelle sur $\mathbb{X}^k$ telle que

$$g^{\otimes k}(x) = \prod_{i=1}^k g(x_i), \quad x = (x_1, \dots, x_k) \in \mathbb{X}^k.$$

Ainsi, sous les hypothèses de régularités vérifiées sur l'ensemble de lois $\mathcal{P}$, le terme de droite de l'équation précédente converge vers $f_k(x) \varphi_\theta^{\otimes k}(x)$, pour peu que $h_k \rightarrow 0$.

Soit maintenant l'échantillon i.i.d. $(X_1, Y_1), \dots, (X_n, Y_n)$ du couple (X, Y) de loi $\mathbb{P}$. On définit un estimateur semi-paramétrique du k -ième chaos $I_k(f_k)$ défini dans l'équation (1.7) construit à partir de cet échantillon de la manière suivante :

$$\frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} \frac{W_{h_k}^{\otimes k}(x - y)}{\varphi_{\hat{\theta}_i}^{\otimes k}(x)} (X_i - \varphi_{\hat{\theta}_i} \cdot \lambda)^{\otimes k}(dy) (X - \varphi_{\hat{\theta}_i} \cdot \lambda)^{\otimes k}(dx), \quad (1.11)$$

où $T_n > 0$ est un paramètre de troncature à calibrer et les $\hat{\theta}_i$ sont des estimateurs *leave-one-out* définis dans le Chapitre 3.

1.3.6 Propriétés minimax lorsque δ est connu

On définit grâce à l'estimateur semi-paramétrique (1.11) du k -ième chaos, l'estimateur de type moyenne empirique suivant de la fonction de régression r pour tout entier strictement positif l

$$\hat{r}_l(X) = \bar{Y}_n + \sum_{k=1}^l \frac{1}{k!} \hat{I}_{k,h_k}(X), \quad (1.12)$$

où $\bar{Y}_n$ est la moyenne empirique de $Y_1, \dots, Y_n$.

Le résultat suivant donne une borne supérieure du risque de l'estimateur $\hat{r}_\delta$ de la fonction de régression pour un nombre de chaos δ connu.

Théorème 1.9. *Soit $\varepsilon > 0$. On fixe $T_n = (\ln n)^{1+\varepsilon}$ et $h_k = (T_n^2 n^{-1})^{1/(2+dk)}$. Alors,*

$$\limsup_{n \rightarrow +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}} \right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}} R_n(\hat{r}_\delta, r) < \infty.$$

Remarque. On déduit du Théorème 1.9 que la vitesse optimale de convergence sur $\mathcal{P}$ est bornée supérieurement par $((\ln n)^{2+2\varepsilon} n^{-1})^{2/(2+d\delta)}$. On remarquera que, à un facteur logarithmique près, nous retrouvons la vitesse optimale $n^{-2/(2+d\delta)}$ obtenue pour l'estimation de la régression en dimension $d\delta$ lorsque la fonction de régression est Lipschitz (voir, e.g., Theorem 1 de Kohler et al. 2009).

Le prochain théorème donne une borne inférieure pour la vitesse de convergence optimale sous $\mathcal{P}$.

Théorème 1.10. *On a,*

$$\liminf_{n \rightarrow +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r) > 0,$$

où l'infimum est pris sur l'ensemble des estimateurs $\tilde{r}$ de r .

Remarque. Le Théorème 1.10 montre que la vitesse de convergence optimale sous $\mathcal{P}$ est bornée inférieurement par $n^{-2/(2+d\delta)}$, ce qui correspond, à un facteur logarithmique près, à la borne supérieure mise en évidence dans le Théorème 1.9.

Les Théorèmes 1.9 et 1.10 nous permettent de conclure qu'à un facteur logarithmique près, l'estimateur $\hat{r}_\delta$ est asymptotiquement minimax sur $\mathcal{P}$.

1.3.7 Propriétés adaptatives lorsque δ est inconnu

On s'intéresse maintenant au cas où le nombre de chaos δ est inconnu. Notons pour tout $m > 0$

$$\mathcal{P}(m) = \{\mathbb{P} \in \mathcal{P} : \|f_k\| \geq m; k \in 1, \dots, \delta\},$$

où $\|\cdot\|$ désigne la norme dans $\mathbb{L}^2$ relativement à la mesure de Lebesgue. Supposons maintenant que $\mathbb{P} \in \mathcal{P}(m)$, on a

$$\delta = \min(k : \|f_k\| = 0) - 1.$$

Considérons pour simplifier un échantillon i.i.d. $(X_1, Y_1), \dots, (X_{2n}, Y_{2n})$ de taille $2n$ du couple (X, Y) de loi $\mathbb{P} \in \mathcal{P}(m)$. Soit $\hat{g}_k$ la contrepartie empirique construite à partir de la deuxième moitié de cet échantillon de $\varphi_\theta f_k$ et défini pour tout $k \in 1, \dots, \delta$ par

$$\hat{g}_k(x) = \frac{1}{n} \sum_{i=n+1}^{2n} Y_i \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y) (X_i - \varphi_{\hat{\theta}} \cdot \lambda)^{\otimes k}(\mathrm{d}y),$$

où $\hat{\theta} = \tilde{\theta}_n(X_{n+1}, \dots, X_{2n})$ a été défini dans la sous-Section 1.3.4, et $b_k = b_k(n)$ est une fenêtre à calibrer. Un estimateur naturel $\hat{\delta}$ de δ est alors défini par

$$\hat{\delta} = \min(k : \|\hat{g}_k\| \leq \rho_k) - 1,$$

où $\rho_k = \rho_k(n)$ est une suite de nombre positifs qui tend vers 0 que nous fixons dans la suite. En utilisant cet estimateur et l'estimateur défini à l'équation 1.12, on définit l'estimateur plug-in $\hat{r}$ de r suivant

$$\hat{r} = \hat{r}_{\hat{\delta}}.$$

Le résultat suivant donne une borne supérieure du risque de l'estimateur $\hat{r}$ pour un nombre de chaos δ inconnu.

Théorème 1.11. *Soient $\varepsilon > d\delta \geq 2$, $\alpha, \beta > 0$ tels que $\alpha + \beta < 1$ et $2\alpha + \beta > 1/(2 + d\delta)$. On fixe $T_n = (\ln n)^{1+\varepsilon}$. Alors, si pour tout entier k , on prend :*

$$h_k = (T_n^2 n^{-1})^{1/(2+dk)}, \quad \rho_k = ((2k)!)^2 n^{(\alpha+\beta-1)/2} \text{ et } b_k = n^{-\beta/(2dk)},$$

on a, pour tout $m > 0$,

$$\limsup_{n \rightarrow +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}} \right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}(m)} R_n(\hat{r}, r) < +\infty.$$

Remarque. On déduit du Théorème 1.11 que la vitesse optimale de convergence sur $\mathcal{P}(m)$ est bornée supérieurement par $((\ln n)^{2+2\varepsilon} n^{-1})^{2/(2+d\delta)}$. Il s'agit de la même borne que celle obtenue dans le Théorème 1.9 pour δ connu.

1.4 Estimation non paramétrique à noyaux de l'intensité d'un processus de Cox

1.4.1 Estimation de l'intensité du processus de Cox

Depuis l'article fondateur de Cox (1955), de nombreuses procédures permettant d'estimer l'intensité du processus de Cox ont été construites et étudiées. Une grande partie d'entre elles se concentre sur le modèle à risque proportionnel (1.5) et la construction de l'estimateur est basée sur la maximisation de la vraisemblance — partielle ou totale, et éventuellement pénalisée — du modèle de Cox (e.g. voir le chapitre 2 du livre de Jacod et Shiryaev, 2013). Dans le cas du modèle (1.5), pour un échantillon de n copies indépendantes $(N^1, Z^1), \dots, (N^n, Z^n)$ de (N, Z) observé sur $[0, T]$, où on rappelle que Z désigne la covariable, la log-vraisemblance partielle s'écrit (e.g. voir le livre de Andersen et al., 1993) :

$$\ell(\beta) = \int_0^T \ln \left(\frac{1}{n} \sum_{i=1}^n e^{\beta^\top Z_t^i} \right) d\bar{N}_t - \frac{1}{n} \sum_{i=1}^n \int_0^T \beta^\top Z_t^i dN_t^i, \text{ si } \bar{N}_t = \frac{1}{n} \sum_{i=1}^n N_t^i.$$

Noter que seul le paramètre d'effet de la covariable est estimé. Par ailleurs, il s'agit ici d'une version non pénalisée. Des versions pénalisées ont été étudiées par Cox (1984), Cox et O'Sullivan (1990), O'Sullivan (1993), etc.

On trouve aussi dans la littérature quelques études non paramétriques. Par exemple, l'article de O'Sullivan (1993) étudie une version non paramétrique du modèle (1.5), dans laquelle le terme lié à la covariable $\beta^\top Z_t$ est remplacé par une fonction de Z_t . L'estimateur est basé sur une méthode de maximisation de la vraisemblance comme dans la formule ci-dessus, mais aussi sur le calcul de dérivées au sens de Fréchet de certaines fonctionnelles, ce qui rend l'approche difficile à utiliser en pratique.

L'article plus récent de Zhang et Kou (2010) considère le modèle général, i.e. sans spécifier la forme de l'intensité λ . La méthode d'estimation, sur la base

ici d'une observation d'une trajectoire de $(N_t)_{t \leq T}$, est une méthode à noyau et l'intensité λ est supposée stationnaire et ergodique, i.e. essentiellement la loi de λ est invariante par translation et p.s.

$$\frac{1}{T} \int_0^T \lambda(s) ds \rightarrow \mathbb{E}\lambda(0),$$

lorsque $T \rightarrow \infty$. On notera que, si ces hypothèses de stationarité et d'ergodicité se justifient dans le cadre de la chimie moléculaire dans lequel se placent les auteurs, elles apparaissent contraignantes en général, notamment dans le domaine de l'actuariat.

On remarquera qu'en règle générale on n'observe pas l'intensité du processus de Cox directement mais plutôt un co-processus. Par exemple, dans l'article de Kou et al. (2005), les auteurs étudient des expériences de biophysique à molécule individuelle pour lesquelles seuls les piques qui correspondent au temps de saut du processus de comptage et un processus sous-jacent sont observés. Dans le domaine de l'assurance automobile (voir Asmussen et Albrecher, 2010), le processus de comptage permet de modéliser les occurrences d'accidents de voitures, eux-mêmes sujets aux intempéries. Dans tous ces exemples, le processus de comptage $N = (N_t)_{t \in [0,1]}$ est accompagné d'un co-processus $Z = (Z_t)_{t \in [0,1]}$ qui est tel que la loi conditionnelle de N sachant Z est celle d'un processus de Poisson d'intensité $\lambda(t, Z)$, où λ est une fonction déterministe. Par abus, nous appellerons ici tout processus de comptage qui vérifie ces caractéristiques un processus de Cox et nous nous intéressons à l'estimation de cette fonction λ à partir d'un échantillon aléatoire $(N^1, Z^1), \dots, (N^n, Z^n)$ de taille n de (N, Z) . On remarquera que la définition d'un estimateur non-paramétrique dans cette représentation du processus de Cox est naturellement sujette au fléau de la dimension puisque la covariable Z est à valeurs dans un espace de dimension infini (voir e.g. O'Sullivan, 1993).

Une façon qui a été très largement étudiée dans la littérature de contourner ce problème est de supposer que la covariable ne dépend pas directement du temps mais que la dépendance du processus de comptage N au temps est entièrement contenue dans une fonction auxiliaire inconnue qu'il faut alors estimer. C'est pas exemple le point de vue étudié dans les articles Hansen, et al. (2015) et Gaïffas et Guillaou (2012) où les auteurs se placent dans le modèle additif d'Aalen et utilisent une procédure Lasso pour obtenir des inégalités oracles sur le risque de leur estimateur.

On peut aussi remarquer qu'en pratique, il est souvent superflu, ou à tout le moins pas absolument nécessaire pour la modélisation, d'observer la trajectoire complète du co-processus. Nous considérons ici le cas où le co-processus est observé à des temps aléatoires qui, en un certain sens, permettent de recouvrir la majorité de l'information portée par celui-ci. Dans ce modèle, le co-processus est observé en un nombre fini de temps aléatoire, circonvenant ainsi au fléau de la dimension. C'est l'approche que nous avons choisi d'étudier.

1.4.2 Présentation du modèle

Nous considérons donc le modèle suivant : soit $N = (N_t)_{t \in [0,1]}$ un processus de comptage et $Z = (Z_t)_{t \in [0,1]}$ un co-processus à valeurs dans $\mathbb{R}^d$ tels que N admet une intensité aléatoire qui dépend de t et des observations de Z aux temps aléatoires $S_1 < S_2 < \dots$.

De manière plus précise, conditionnellement à la tribu $\mathcal{S}$ engendrée par ces temps aléatoires, N est un processus de Cox d'intensité

$$\lambda_S(t, Z) = \theta_S(t, \vec{Z}_S(t)), \quad (1.13)$$

où $\theta_S(t, \cdot)$ est une fonction de $\mathbb{R}^{dM_t}$ dans $\mathbb{R}_+$, pour toute fonction $z : [0, 1] \rightarrow \mathbb{R}$, la notation $\vec{z}_S(t)$ correspond au vecteur $(z_{S_1}, \dots, z_{S_{M_t}}) \in \mathbb{R}^{dM_t}$, et M est le processus de comptage associé aux temps $S = (S_1, S_2, \dots)$.

Dans la suite, $(N^1, Z^1), \dots, (N^n, Z^n)$ est un échantillon indépendant et identiquement distribué (i.i.d.) de (N, Z) conditionnellement à $\mathcal{S}$. Notre but est de construire et d'étudier les propriétés statistiques d'un estimateur à noyaux de λ exploitant ces données. Une difficulté réside en particulier dans le fait que la dimension de ce problème d'estimation dépend du processus de comptage $(M_t)_{t \in [0,1]}$ de sorte qu'elle augmente avec le temps t . On s'attend donc à une détérioration de la précision des procédures d'estimation lorsque la variable de temps augmente.

Nous proposons, dans ce nouveau modèle, une procédure d'estimation de l'intensité (1.13) basée sur n copies indépendantes de (N, Z) . On montre les bonnes propriétés asymptotiques de l'estimateur proposé : convergence en probabilité, normalité asymptotique et une borne supérieure pour l'erreur quadratique moyenne ponctuelle. Les vitesses de convergence obtenues sont

comprises entre les vitesses obtenues classiquement dans un cadre de régression en dimension finie (voir e.g. Györfi et al. (Györfi et al., 2006)) et les vitesses minimax obtenues en régression fonctionnelle (voir les articles de Mas, 2012 et Biau et al., 2010).

1.4.3 Stratégie d'estimation

Nous proposons ici un estimateur non-paramétrique de l'intensité $\lambda(t, z)$ pour tous $t \in [0, 1]$ et $z : [0, 1] \rightarrow \mathbb{R}$, en se basant sur l'observation d'une trajectoire du processus $(M_t)_{t \in [0, 1]}$ et d'un échantillon i.i.d. de taille n du couple (N, Z) . On fixe pour cela $Z = (Z_t)_{t \in [0, 1]}$ un co-processus à valeurs dans $\mathbb{R}^d$ observé aux temps aléatoires $S_1 < S_2 < \dots$ et $N = (N_t)_{t \in [0, 1]}$ le processus de Cox d'intensité $\lambda_S(t, Z) = \theta_S(t, \vec{Z}_S(t))$.

Considérons en guise d'heuristique le cas simple où nous observons directement un échantillon $\left(\theta_S(t, \vec{Z}_S^k(t))\right)_{k=1, \dots, n}$ de l'intensité aléatoire $\lambda_S(t, Z) = \theta_S(t, \vec{Z}_S(t))$. Le problème d'estimation de la fonction d'intensité λ_S peut alors être simplement vu comme le problème d'estimation de la régression où $\lambda_S(t, \cdot)$ serait la fonction de régression. L'estimateur de Nadaraya-Watson s'écrirait alors

$$\hat{\lambda}_{NW}(t, z) = \frac{\sum_{k=1}^n \theta_S(t, \vec{Z}_S^k(t)) H_\eta(\vec{z}_S(t) - \vec{Z}_S^k(t))}{\sum_{l=1}^n H_\eta(\vec{z}_S(t) - \vec{Z}_S^l(t))}.$$

où η est une variable aléatoire positive et $\mathcal{S}$ -mesurable (appelée fenêtre dans la suite), et H_η désigne le noyau produit multivarié $\mathcal{H}_\eta^{\otimes dM_t}$ avec $\mathcal{H}$ un noyau, i.e. une fonction de $\mathbb{L}^1(\mathbb{R})$ telle que $\int_{\mathbb{R}} \mathcal{H}(u) du = 1$, et $\mathcal{H}_\eta(\cdot) = \eta^{-1} \mathcal{H}(\eta^{-1} \cdot)$.

Rappelons d'autre part que, conditionnellement à $\mathcal{S}$ et à la tribu $\mathcal{Z}$ engendrée par les trajectoires $(Z^1, \dots, Z_n)$ du co-processus, N est un processus de Poisson de fonction d'intensité $t \mapsto \theta_S(t, \vec{Z}_S^k(t))$. On en déduit l'estimateur naturel suivant

$$\int_0^t K_h(t - u) dN_u^k = \sum_{i=1}^{N_t^k} K_h(t - T_i^k),$$

où les $T_1^k, T_2^k, \dots$ sont les temps de saut de la k -ième trajectoire N^k de N , la fonction $K : \mathbb{R}_+ \rightarrow \mathbb{R}$ est un noyau, h est une fenêtre et $K_h(\cdot) = h^{-1} K(h^{-1} \cdot)$.

En exploitant les deux dernières remarques, on définit l'estimateur *plug-in* suivant

$$\hat{\lambda}_{S,h,\eta}(t, z) = \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{\hat{f}_{S,h,\eta}(\vec{z}_S(t))}.$$

où on a noté

$$\begin{aligned}\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) &= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^{N_t^k} K_h(t - T_i^k) H_\eta(\vec{z}_S(t) - \vec{Z}_S^k(t)), \\ \hat{f}_{S,h,\eta}(\vec{z}_S(t)) &= \frac{1}{n} \sum_{\ell=1}^n H_\eta(\vec{z}_S(t) - \vec{Z}_S^\ell(t)),\end{aligned}$$

L'estimateur étudié dans la suite est une version *trimmée* du précédent

$$\tilde{\lambda}_{S,h,\eta}(t, z) = \tilde{\theta}_{S,h,\eta}(t, \vec{z}_S(t)) = \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{\tilde{f}_{S,h,\eta}(\vec{z}_S(t))}, \quad (1.14)$$

où $\tilde{f}_{S,h,\eta}(\vec{z}_S(t)) = \max(\hat{f}_{S,h,\eta}(\vec{z}_S(t)), a_n)$ et $(a_n)_{n \in \mathbb{N}}$ est une suite décroissante de variables aléatoires réelles positives et $\mathcal{S}$ -mesurables définie par $a_n = (n\eta^{dM_t})^{-1/2}$.

1.4.4 Résultats

On s'intéresse aux propriétés asymptotiques de l'estimateur (1.14).

Les hypothèses du modèle sont détaillées dans le Chapitre 4. Nous les présentons en quelques lignes dans cette introduction. Dans la suite, $t \in [0, 1]$ et $z : [0, 1] \rightarrow \mathbb{R}$ sont considérés comme étant fixés.

Rappelons tout d'abord que la dimension de l'espace de départ de la fonction θ_S définie en (1.13) dépend du processus de comptage $(M_t)_{t \in [0, 1]}$ de sorte qu'elle augmente avec le temps. Remarquons de plus que, presque sûrement, $t \in]S_{M_t}, S_{M_t} + 1[= I_{M_t}$, nous ramenant ainsi à un problème d'estimation en dimension finie fixée sur l'intervalle I_{M_t} .

Partant de ce constat, nous supposons tout d'abord que, conditionnellement à la tribu $\mathcal{S}$, pour tout $u \in I_{M_t}$, $\vec{Z}_S(u)$ admet une densité conditionnelle notée f .

On suppose de plus que, conditionnellement à $\mathcal{S}$, les fonctions $\theta_S(u, \cdot)$ et f sont suffisamment régulières en un sens usuel en régression finie-dimensionnelle (voir Chapitre 4 pour plus de détails).

La régularité des fonctions $\theta_S(u, \cdot)$ et f étant conditionnelle à $\mathcal{S}$, il est nécessaire de contrôler la régularité du processus $(M_t)_{t \in [0,1]}$ afin d'obtenir des résultats asymptotiques sur notre estimateur. On suppose donc que le processus $(M_t)_{t \in [0,1]}$ vérifie des conditions sur ses moments exponentiels. Ces conditions sont détaillées dans le chapitre correspondant.

On suppose finalement que les noyaux $\mathcal{H}$ et K sont d'ordre 2 et tels que $\text{supp} \mathcal{H} = [-1, 1]$, $\text{supp} K = [0, 1]$, $\|\mathcal{H}\|_\infty < \infty$ et $K \in \mathbb{L}^4(\mathbb{R})$.

Un premier résultat permet de contrôler l'erreur quadratique moyenne conditionnelle ponctuelle de notre estimateur. Elle est définie par l'équation suivante

$$\text{MSE}_{\mathcal{S}}(t, z) = \mathbb{E} \left[\left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right)^2 \mid \mathcal{S} \right].$$

Dans tous les résultats, on a noté **(H1)** à **(H5)**, **(T1)** et **(T2)** les hypothèses discutées précédemment et présentées en détails dans le Chapitre 4 de ce manuscrit.

Théorème 1.12. *Soient $h = \eta = n^{-\frac{1}{5+dM_t}}$ deux fenêtres $\mathcal{S}$ -mesurables. Alors sous les hypothèses **(H1)** à **(H5)**, **(T1)** et **(T2)**, il existe une constante positive τ qui ne dépend que de F_0 , F_∞ , Θ et Q telle que*

$$\text{MSE}_{\mathcal{S}}(t, z) \leq e^{\tau M_t} n^{-\frac{4}{5+dM_t}}.$$

On en déduit

$$\text{MSE}(t, z) = \mathbb{E}(\text{MSE}_{\mathcal{S}}(t, z)) \leq \mathbb{E} \left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}} \right).$$

La consistance de l'estimateur (1.14) peut être obtenue directement via le Théorème 1.12. La Proposition suivante montre que la consistance de l'estimateur peut être obtenue sous un jeu d'hypothèse plus faible que celui du Théorème 1.12.

Proposition 1.13. *Soient h et η deux fenêtres $\mathcal{S}$ -mesurables telles que $h \rightarrow 0$, $\eta \rightarrow 0$ et $nh\eta^{dM_t} \rightarrow +\infty$ p.s. lorsque $n \rightarrow +\infty$. Alors sous les hypothèses (H1), (H2) et (H4), on a*

$$\tilde{\lambda}_{S,h,\eta}(t, z) \xrightarrow{\mathbb{P}} \lambda_S(t, z).$$

Pour tout noyau $\mathcal{K}$, on note désormais $m_j(\mathcal{K}) = \int u^j \mathcal{K}(u) du$ so j -ème moment. Pour terminer l'étude asymptotique de l'estimateur (1.14) nous avons montré sa normalité asymptotique. C'est l'objet du Théorème suivant.

Théorème 1.14. *Soient h et η deux fenêtres $\mathcal{S}$ -mesurables telles que $h \rightarrow 0$, $\eta \rightarrow 0$, $nh^5\eta^{dM_t} = O(1)$, $nh\eta^{dM_t+4} = O(1)$ et $nh\eta^{dM_t} \rightarrow +\infty$ p.s. lorsque $n \rightarrow +\infty$. Alors sous les hypothèses (H1) à (H5), (T1) et (T2) et pour tout $z : [0, 1] \rightarrow \mathbb{R}$ tel que $\theta_S(t, \vec{z}_S(t)) \neq 0$,*

$$(nh\eta^{dM_t})^{1/2} \sigma^{-1} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) + h^2 B_1 + \eta^2 B_2 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where

$$\begin{aligned} \sigma^2 &= f(\vec{z}_S(t)) \theta_S(t, \vec{z}_S(t)) \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t}, \\ B_1 &= \frac{1}{2} m_2(K) \frac{1}{f(\vec{z}_S(t))} \partial_1^2 (f(\vec{z}_S(t)) \theta_S(t, \vec{z}_S(t))), \\ B_2 &= \frac{1}{2} m_2(\mathcal{H}) \sum_{i=1}^{dM_t} \left(\frac{1}{f(\vec{z}_S(t))} \partial_{i+1}^2 (f(\vec{z}_S(t)) \theta_S(t, \vec{z}_S(t))) \right. \\ &\quad \left. - \frac{\theta(t, \vec{z}_S(t))}{f(\vec{z}_S(t))} \partial_i^2 f(\vec{z}_S(t)) \right). \end{aligned}$$

Remarque. Dans le Théorème 1.14, en prenant $h = \eta = n^{-\frac{1}{5+dM_t}}$, on obtient

$$(nh\eta^{dM_t})^{1/2} \sigma^{-1} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right) + \sigma^{-1} (B_1 + B_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

comme $(nh\eta^{dM_t})^{1/2} h^2 = (nh\eta^{dM_t})^{1/2} \eta^2 = 1$. On en déduit que la loi asymptotique de notre estimateur est une loi normale de biais et de variance non nuls.

De plus, le Théorème 1.12 donne les outils qui permettent de définir des fenêtres optimales en terme de MSE ponctuelle asymptotique. Cependant, comme la borne trouvée dans le Théorème 1.12 dépend de la quantité $\mathbb{E} \left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}} \right)$ qui est aléatoire, on ne peut pas conclure sur la vitesse de convergence de l'estimateur sans faire d'autres hypothèses sur le processus M . Il est tout de même intéressant de remarquer que sur quelques exemples détaillés dans le Chapitre 4, on obtient une vitesse de convergence comprise entre les vitesses usuelles obtenues dans un cadre fini-dimensionnel (voir par exemple Györfi et al., 2006) et les vitesses obtenues dans l'article de Biau et al. (2010) dans un cadre infini-dimensionnel. On peut expliquer cela par la particularité du modèle 1.13.

1.4.5 Validation Croisée

L'utilisation pratique de l'estimateur proposé repose sur le choix de deux fenêtres. On remarque que dans le cadre de la régression à noyaux, un critère classique de sélection est de choisir des fenêtres minimisant l'erreur de prédiction. Dans ce contexte, l'erreur de prédiction est classiquement estimée par validation croisée. On propose ici d'adopter une approche similaire. Ainsi, on se propose donc d'estimer l'erreur d'estimation intégrée suivante par validation croisée K -fold

$$\mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{dM_s}} \left(\tilde{\theta}_{S,h,\eta}^{-\kappa(1)}(s, z) - \theta_S(s, z) \right)^2 f_{M_s}(z) dz ds \right]. \quad (1.15)$$

On choisit de plus de se placer dans le cas plus simple où $h = \eta$.

En pratique, la validation croisée K -fold consiste à couper les données en K sous-échantillons de même taille. Pour chaque sous-échantillon, on ajuste le modèle sur les $K - 1$ autres parties des données et on calcule l'erreur d'estimation en utilisant la partie non-utilisée. On obtient finalement un estimateur de l'erreur d'estimation en agrégeant les K quantités ainsi construites.

Soit $\kappa : \{1, \dots, N\} \mapsto 1, \dots, K$ une fonction d'indexation qui associe à tout numéro d'observation i le numéro de la partie des données à laquelle l'observation a été allouée par randomisation. On note $\tilde{\theta}_{S,h}^{-k}$ l'estimateur construit sans la k -ième partie des données. Un estimateur de l'erreur

d'estimation (1.15) par validation croisée *K-fold* s'écrit alors

$$CV(\tilde{\theta}, h) = \frac{1}{n} \sum_{i=1}^n \left(\int_0^t \tilde{\theta}_{S,h}^{-\kappa(i)}(s, \vec{Z}_S^i(s))^2 ds - \int_0^t \tilde{\theta}_{S,h}^{-\kappa(i)}(s, \vec{Z}_S^i(s)) dN_s^i \right).$$

Remarque. On montre aisément que l'estimateur ainsi défini estime sans biais le terme à minimiser dans l'erreur d'estimation (1.15).

On illustre sur des simulations le bon comportement pratique de cette méthode.

A nonparametric test for Cox processes¹

Sommaire

2.1	Introduction	46
2.2	Tests for Cox processes	47
2.2.1	Principle of the test	47
2.2.2	Testing $\mathbf{H}_0$ vs $\mathbf{H}_1$	49
2.2.3	Local alternative	50
2.2.4	Non-asymptotic test	53
2.3	Simulation study	56
2.3.1	Asymptotic level study	56
2.3.2	Rejection power study	58
2.4	Application to real data	60
2.4.1	Analysis of some arrival times in a call center	60
2.4.2	Analysis of the scoring times of a soccer team	61
2.5	Proof of Theorem 2.1	62
2.5.1	Auxiliary results	62
2.5.2	Proof of Theorem 2.1	65
2.6	Proof of corollary 2.2	66
2.7	Proof of Theorem 2.3	68
2.7.1	Auxiliary results	68
2.7.2	Proof of Theorem 2.3	72
2.8	Proofs of Theorems 2.4 and 2.5	73
2.8.1	Auxiliary results	73
2.8.2	Proof of Theorem 2.4	76
2.8.3	Proof of Theorem 2.5	77

¹For the most part, this chapter has been accepted for publication in the *Journal of Statistical Planning and Inference*.

2.1 Introduction

Count process formulation is commonly used to describe and analyze many kind of data in sciences and engineering. A widely used class of such processes is the family of the so-called Cox processes or doubly stochastic Poisson processes. Compared to the standard Poisson process, the key feature of a Cox process is that its arrival rate is stochastic, depending on some covariate. In other words, if $T > 0$ denotes the observation period, $N = (N_t)_{t \in [0, T]}$ the Cox process and $\Lambda = (\Lambda(t))_{t \in [0, T]}$ the (stochastic) cumulative arrival rate then, conditioning on Λ , the distribution of N is that of a Poisson process with cumulative intensity Λ . Using randomness in the cumulative intensity, the statistician can take into account some auxiliary information, thus leading to a more realistic model. For general references, we refer the reader to the monographies by Cox and Isham (1980), Karr (1991), Kingman (1993) or Snyder and Miller (2012).

In actuarial sciences and risk theory for instance, the number of claims in the risk model may be represented by a Cox process. In this area, the central quantity is the ruin probability, i.e the probability that the surplus of the insurer is negative at some time (see e.g., Björk and Grandell, 1988; Grandell, 2012; Schmidli, 1996). Cox process also appears in biophysics and physical chemistry (see e.g., Kou et al., 2005; Kou, 2008; Zhang and Kou, 2010). In these fields, experimental data consist of photon arrival times with the arrival rate depending on the stochastic dynamics of the system under study (for example, the active and inactive states of an enzyme can have different photon emission intensities); by analyzing the photon arrival data, one aims to learn the system's biological properties. Cox process data arise in neuroscience, to analyse the form of neural spike trains, defined as a chain of action potentials emitted by a single neuron over a period of time (see e.g., Gerstner and Kistler, 2002; Reynaud-Bourret et al., 2014). Finally let us mention astrophysics as another area where Cox process data often occur (see e.g., Scargle, 1998; Carroll and Ostlie, 2007).

In general, it is tempting to incorporate abusively numerous covariates in the statistical model, though a Poisson process model is satisfactory. In this chapter, we elaborate a nonparametric test statistic to highlight the Poisson nature of a Cox process. More precisely, based on i.i.d. copies of N , we construct a nonparametric test statistic for $\mathbf{H}_0$: N is a Poisson process vs $\mathbf{H}_1$: N is not a Poisson process. This setting of i.i.d. copies of N is justified

by the fact that in many situations, the duration of observation is limited but the number of observed paths is large.

Among the various possibilities for constructing a test statistic devoted to this problem, a naive approach consists in first estimating both functions $t \mapsto \mathbb{E}[N_t|\Lambda]$ and $t \mapsto \mathbb{E}N_t$ and then testing whether these functions are equal. However, this approach suffers from two main drawbacks: the curse of dimensionality (whenever Λ takes values in a high-dimensional space) and the knowledge a priori on Λ . Another approach is to test whether time-jumps of N are Poisson time-jumps; in this direction, we refer the reader to the paper by Reynaud-Bourret et al. (2014), in which a modified Kolmogorov-Smirnov statistic is used.

In this chapter, we elaborate and study two test statistics, both based on the observation that a Cox process is a Poisson process if, and only if its mean and variance function are equal. As we shall see, this approach leads to very simple and easily implementable tests.

The chapter is organized as follows. In Section 2.2, we first present the test statistic, then we establish asymptotic performances dedicated to the problem of $\mathbf{H}_0$ vs $\mathbf{H}_1$. The case of a local alternative is also considered to study asymptotic power of the tests. The last part of Section 2.2 is devoted to the non-asymptotic study of one of the tests. Section 2.3 is devoted to a simulation study. An application to real data is presented in Section 2.4. The proofs of our results are postponed to the last sections of the chapter.

2.2 Tests for Cox processes

2.2.1 Principle of the test

Throughout the chapter, $T > 0$ is the (deterministic) duration of observation, and $N = (N_t)_{t \in [0, T]}$ is a Cox process with cumulative intensity process $\Lambda = (\Lambda(t))_{t \in [0, T]}$, such that the fourth moment of N_T is finite, i.e. $\mathbb{E}N_T^4 < \infty$, and $\mathbb{E}N_t \neq 0$ for some $t \in]0, T[$. Note that the mean function $t \mapsto \mathbb{E}N_t = \mathbb{E}\Lambda(t)$ might not be absolutely continuous, so it is not necessarily the integral of an intensity.

We let m and σ^2 the mean and variance functions of N , i.e. for all $t \in [0, T]$:

$$m(t) = \mathbb{E}N_t \text{ and } \sigma^2(t) = \text{var}(N_t).$$

Recall that for all $t \in [0, T]$ (see p. 66 in the book by Kingman, 1993):

$$\sigma^2(t) = m(t) + \text{var}(\mathbb{E}[N_t|\Lambda]) = m(t) + \text{var}(\Lambda(t)). \quad (2.1)$$

Hence, $\sigma^2(t) \geq m(t)$ that is, each N_t is overdispersed. Moreover, if $m = \sigma^2$, then $\mathbb{E}[N_t|\Lambda] = \mathbb{E}N_t$ for all $t \in [0, T]$, thus N is a Poisson process. As a consequence, N is a Poisson process if, and only if $m = \sigma^2$. This observation is the key feature for the construction of our test statistic, with this respect the problem can be written as follows:

$$\mathbf{H}_0 : \sigma^2 = m \text{ vs } \mathbf{H}_1 : \exists t \leq T \text{ with } \sigma^2(t) > m(t).$$

From now on, we let the data $N^{(1)}, \dots, N^{(n)}$ to be independent copies of N . By above, natural test statistics are based on the process $\hat{\sigma}^2 - \hat{m} = (\hat{\sigma}^2(t) - \hat{m}(t))_{t \in [0, T]}$, where $\hat{m}$ and $\hat{\sigma}^2$ are the empirical counterparts of m and σ^2 :

$$\hat{m}(t) = \frac{1}{n} \sum_{i=1}^n N_t^{(i)} \text{ and } \hat{\sigma}^2(t) = \frac{1}{n-1} \sum_{i=1}^n (N_t^{(i)} - \hat{m}(t))^2.$$

In this chapter, convergence in distribution of stochastic processes is intended with respect to the Skorokhod topology (see Chapter VI in the book by Jacod and Shiryaev, 2013).

Our first main result gives the asymptotic distribution of the process $\hat{\sigma}^2 - \hat{m}$ under $\mathbf{H}_0$ (see Section 2.5 for the proof).

Theorem 2.1. *Let $B = (B_t)_{t \in \mathbb{R}_+}$ be a standard Brownian Motion on the real line. Under $\mathbf{H}_0$, $\hat{\sigma}^2 - \hat{m}$ is a martingale and*

$$\sqrt{n} (\hat{\sigma}^2 - \hat{m}) \xrightarrow{(\text{law})} (B_{2m(t)^2})_{t \leq T}.$$

As far as we know, the martingale property for $\hat{\sigma}^2 - \hat{m}$ has not been observed yet. This property, which is interesting by itself, plays a crucial role in the derivation of the asymptotic result.

2.2.2 Testing $\mathbf{H}_0$ vs $\mathbf{H}_1$

Various test statistics based on the process $\hat{\sigma}^2 - \hat{m}$ can be derived for our purpose. In this chapter, we shall concentrate on two test statistics for testing whether a Cox process is Poisson or not. These statistics, denoted by $\hat{S}_1$ and $\hat{S}_2$, involve either the supremum or the integral of the process $\hat{\sigma}^2 - \hat{m}$. They are defined as follows:

$$\hat{S}_1 = \sup_{t \leq T} (\hat{\sigma}^2(t) - \hat{m}(t)), \text{ and } \hat{S}_2 = \int_0^T (\hat{\sigma}^2(t) - \hat{m}(t)) dt.$$

These test statistics are natural for testing if a nonnegative function is equal to 0. Moreover they are compatible with the unilateral nature of the problem, since the alternative hypothesis may be written $\mathbf{H}_1$: $\sigma^2(t) > m(t)$ for some $t \leq T$.

We now present the asymptotic properties of $\hat{S}_1$ and $\hat{S}_2$.

Corollary 2.2. *Let $\hat{I}^2 = \int_0^T (T - t) \hat{m}(t)^2 dt$.*

(i) *Under $\mathbf{H}_0$,*

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{(\text{law})} |\mathcal{N}(0, 2)|, \text{ and } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{(\text{law})} \mathcal{N}(0, 4).$$

(ii) *Under $\mathbf{H}_1$,*

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{\text{prob.}} +\infty, \text{ and } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{\text{prob.}} +\infty.$$

This result is proved in Section 2.6. By above, the test statistics $\hat{S}_1/\hat{m}(T)$ and $\hat{S}_2/\hat{I}$ define tests with maximal asymptotic power. The rejection regions for tests of level $\alpha \in]0, 1[$ are:

$$R_1(\alpha) = \left\{ \frac{\hat{S}_1}{\hat{m}(T)} \geq \sqrt{\frac{2}{n}} q_{1-\alpha/2} \right\} \text{ and } R_2(\alpha) = \left\{ \frac{\hat{S}_2}{\hat{I}} \geq \frac{2}{\sqrt{n}} q_{1-\alpha} \right\}, \quad (2.2)$$

where for each $\beta \in]0, 1[$, q_β is the $\mathcal{N}(0, 1)$ -quantile of order β .

Remark. A close inspection of the proof of Corollary 2.2 reveals that a more general setting may be reached. Indeed, for the test of $\mathbf{H}_0$ vs $\mathbf{H}_1$,

we only need to assume that N is in some class of overdispersed counting processes (i.e. $\text{var}(N_t) \geq \mathbb{E}N_t$ for all $t \in [0, T]$) which satisfies the property: $\text{var}(N_t) = \mathbb{E}N_t$ for all $t \in [0, T]$ if, and only if N is a Poisson process. The archetype of such a class of counting processes is given by the Cox process. It is also satisfied by other classes, such as some subclasses of Hawkes process for instance. In this direction, our test is more or less a functional version of the classical overdispersion test, that is used for testing the Poisson distribution of a sequence of count data (see for instance Rao and Chakravarti, 1956 or Bohning, 1994). Recall that overdispersion tests are widely used in actuarial science for the study of claims counts (e.g. Denuit et al., 2007).

2.2.3 Local alternative

For testing $\mathbf{H}_0$ vs $\mathbf{H}_1$, we have seen that the statistics $\hat{S}_1$ and $\hat{S}_2$ provide tests with maximal asymptotic power. The aim of this section is to proceed to a theoretical comparison between the two tests.

One of the most common way to analyze the asymptotic properties of test statistics when the null hypothesis $\mathbf{H}_0$ is false is to consider an artificial situation, represented by a so-called *local alternative* $\mathbf{H}_1^n$, under which the process which actually generates the data changes systematically as the sample size is increased. Specifically, it is necessary that as the sample size tends to infinity, the distance between the null hypothesis $\mathbf{H}_0$ and the alternative $\mathbf{H}_1^n$ should diminish. With this respect, the asymptotic power under the local alternative turns to be an interesting quantity for comparing several tests. We refer the reader to Engle (1984) and the book by van der Vaart (2000) for a general overview on local alternatives.

In this section, we assume in addition that the Cox process N has an intensity $\lambda = (\lambda(t))_{t \in [0, T]}$, i.e. with probability 1, Λ is absolutely continuous, and

$$\Lambda(t) = \int_0^t \lambda(s) ds, \quad \forall t \in [0, T].$$

We introduce a vanishing sequence of positive numbers $(d_n)_n$. A classical way to define the local alternative is to consider a *Pitman drift* (see Davidson and MacKinnon, 2006, and the references therein), which specifies the

intensity as a local deviation from the null hypothesis.

$\mathbf{H}_1^n$: There exists a bounded non-null function $\lambda_0 : [0, T] \rightarrow \mathbb{R}_+$ and a

process $\Delta = (\Delta_t)_{t \in [0, T]}$ with $\sup_{t \in [0, T]} \mathbb{E} \Delta_t^6 < \infty$ and $\text{var}(\int_0^{t_0} \Delta_s ds) > 0$

for some $t_0 \in [0, T]$, and such that $\lambda = \lambda_0 + d_n \Delta$ w.p. 1.

Here, the assumption on the sixth moment $\sup_{t \in [0, T]} \mathbb{E} \Delta_t^6 < \infty$ ensures that the Cox process N satisfies $\mathbb{E} N_T^6 < \infty$ as well. Moreover, the condition $\text{var}(\int_0^{t_0} \Delta_s ds) > 0$ for some $t_0 \in [0, T]$ implies $\sigma^2(t_0) > m(t_0)$ (see subsection 2.7.2 for the proof), thus hypothesis $\mathbf{H}_1^n$ is contained in $\mathbf{H}_1$.

Also observe that when n gets larger and $\mathbf{H}_1^n$ holds, N becomes closer to a Poisson process. Thus, $(d_n)_n$ has to be understood as a separation rate from $\mathbf{H}_1^n$ to the null hypothesis $\mathbf{H}_0$. In particular, next result -whose proof has to be found in Section 2.3- states that in view of a consistent test for $\mathbf{H}_0$ vs $\mathbf{H}_1^n$, it is necessary and sufficient that d_n^2 tends to 0 slower than $1/\sqrt{n}$.

Theorem 2.3. *Let $B = (B_t)_{t \in \mathbb{R}_+}$ be a standard Brownian Motion on the real line. Assume that $\mathbf{H}_1^n$ holds, and denote by m_0 and v the functions defined for all $t \in [0, T]$ by*

$$m_0(t) = \int_0^t \lambda_0(s) ds, \text{ and } v(t) = \text{var}(\int_0^t \Delta_s ds).$$

Moreover, we let $I_0^2 = \int_0^T (T-t)m_0(t)^2 dt$.

(i) If $\sqrt{n} d_n^2 \rightarrow \infty$, then

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{\text{prob.}} +\infty, \text{ and } \sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{\text{prob.}} +\infty.$$

(ii) If $\sqrt{n} d_n^2 \rightarrow d < \infty$, then

$$\begin{aligned} \sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} &\xrightarrow{(\text{law})} \frac{1}{m_0(T)} \sup_{t \leq T} (B_{2m_0(t)^2} + dv(t)), \text{ and} \\ \sqrt{n} \frac{\hat{S}_2}{\hat{I}} &\xrightarrow{(\text{law})} 2\mathcal{N}(0, 1) + \frac{d}{I_0} \int_0^T v(t) dt. \end{aligned}$$

In the problem $\mathbf{H}_0$ vs $\mathbf{H}_1^n$, we consider the tests defined by the rejection regions in (2.2), with $\alpha \in]0, 1[$. For a power study, we assume from now on that $\mathbf{H}_1^n$ holds. By above, if $\sqrt{n}d_n^2 \rightarrow d < \infty$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(R_2(\alpha)) = 1 - \Phi\left(q_{1-\alpha} - \frac{d}{2I_0} \int_0^T v(t)dt\right) < 1, \quad (2.3)$$

where Φ stands for the cumulative distribution function of the $\mathcal{N}(0, 1)$ distribution. However, we only have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(R_1(\alpha)) \leq \mathbb{P}\left(\frac{1}{m_0(T)} \sup_{t \leq T} (B_{2m_0(t)^2} + dv(t)) \geq \sqrt{2}q_{1-\alpha/2}\right) < 1,$$

according to the Portmanteau Theorem, as the limit distribution may have a mass at point $\sqrt{2}q_{1-\alpha/2}$. At least, we deduce from above and part (i) of Theorem 2.3 that both tests defined by $R_1(\alpha)$ and $R_2(\alpha)$ are consistent if, and only if $\sqrt{n}d_n^2 \rightarrow \infty$.

In the rest of the section, we assume that $\sqrt{n}d_n^2 \rightarrow d < \infty$. For a comparison of the tests, we need an additional assumption ensuring that the limit distribution associated with statistic $\hat{S}_1$ is continuous. To this aim, we suppose now that $\lambda_0(t) > 0$ for all $t \in [0, T]$, and we let ℓ_0 the function such that $\ell_0(t) = 2m_0(t)^2$ for all $t \in [0, T]$. Then, ℓ_0 is a continuous and increasing function with $\ell_0(0) = 0$, and

$$\sup_{t \leq T} (B_{\ell_0(t)} + dv(t)) = \sup_{s \leq \ell_0(T)} (B_s + dv \circ \ell_0^{-1}(s)). \quad (2.4)$$

Now observe that function $v \circ \ell_0^{-1}$ is absolutely continuous. Thus, by the Girsanov Theorem (Revuz and Yor, 2013), there exists a probability measure Q such that the distribution of the random variable in (2.4) equals the distribution under Q of the supremum of a standard Brownian Motion over $[0, \ell_0(T)]$. According to Proposition III.3.7 in the book by Revuz and Yor (2013), this distribution is continuous, which proves that the distribution of the random variable in (2.4) is also continuous. As consequence, the Portmanteau Theorem and Theorem 2.3 (ii) give

$$\lim_{n \rightarrow \infty} \mathbb{P}(R_1(\alpha)) = \mathbb{P}\left(\frac{1}{m_0(T)} \sup_{s \leq \ell_0(T)} (B_s + dv \circ \ell_0^{-1}(s)) \geq \sqrt{2}q_{1-\alpha/2}\right).$$

Unfortunately, the latter probability is not known, except for some special cases that we now study.

In addition to $\mathbf{H}_1^n$, we suppose that $\lambda_0(t) = \lambda_0 > 0$ and $\Delta_t = Z$ for all $t \in [0, T]$, where Z is a random variable with variance w^2 . In particular, we study the case of a small deviation from an homogeneous Poisson process. Then, formula (2.4) writes

$$\sup_{t \leq T} (B_{\ell_0(t)} + dv(t)) = \sup_{s \leq 2\lambda_0^2 T^2} (B_s + \frac{dw^2}{2\lambda_0^2} s).$$

Obviously, we have in this case $v(t) = w^2 t^2$, $m_0(t) = \lambda_0 t$ and $I_0^2 = \lambda_0^2 T^4 / 12$ for all $t \in [0, T]$. Setting $x = dw^2 T$, we obtain with Theorem 2.3 (ii) and the distribution of the supremum of a drifted Brownian Motion (see p. 250 in the book by Borodin and Salminen, 2012):

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(R_1(\alpha)) &= \exp\left(\frac{\sqrt{2}}{\lambda_0} x q_{1-\alpha/2}\right) \left(1 - \Phi\left(q_{1-\alpha/2} + \frac{x}{\sqrt{2}\lambda_0}\right)\right) \\ &\quad + 1 - \Phi\left(q_{1-\alpha/2} - \frac{x}{\sqrt{2}\lambda_0}\right). \end{aligned}$$

Moreover, by (2.3):

$$\lim_{n \rightarrow \infty} \mathbb{P}(R_2(\alpha)) = 1 - \Phi\left(q_{1-\alpha} - \frac{x}{\sqrt{3}\lambda_0}\right).$$

Assume $\lambda_0 = 1$ and denote by g_1 and g_2 respectively the above functions of x . Figure 2.1 shows a comparison of these two quantities as functions of x and for $\alpha = 0.05$ or $\alpha = 0.1$. This example suggests a better power for the test induced by $R_1(\alpha)$. The numerical experiments given in the next section also suggest a better power for the first test.

2.2.4 Non-asymptotic test

For simplicity, now let $T = 1$. Recall that, according to Corollary 2.2, the test statistic $\hat{S}_1 / \hat{m}(1)$ defines an asymptotic test with maximal power and a rejection region for a test of level $\alpha \in]0, 1[$ is

$$R'_1(\alpha) = \left\{ \hat{S}_1 \geq \frac{\sqrt{2} \hat{m}(1)}{\sqrt{n}} q_{1-\alpha/2} \right\}. \quad (2.5)$$

We now turn to the non-asymptotic study of the test statistic. Indeed, we may now specify to what extent a non asymptotic rejection region is close or not to the ideal situation described by $R'_1(\alpha)$.

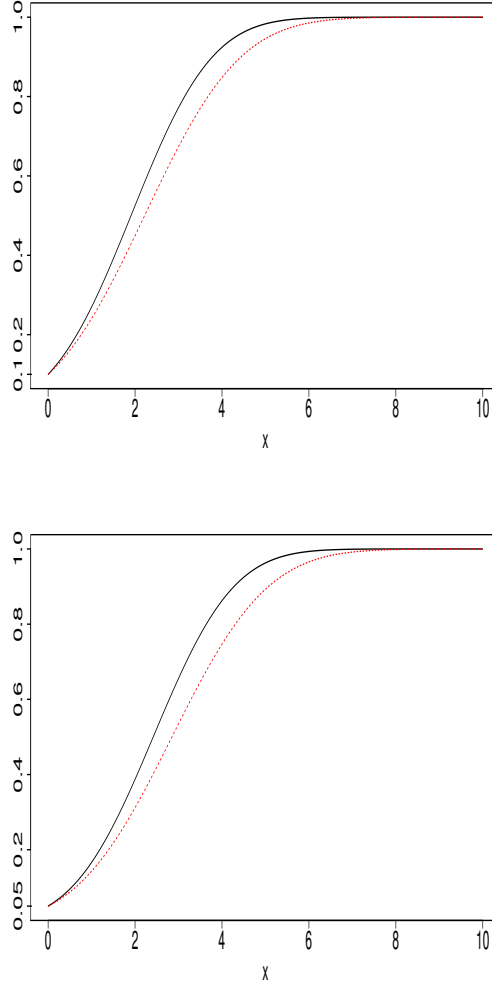


Figure 2.1: Plots of g_1 (full lines) and g_2 (dashed lines), with $\alpha = 0.1$ for the curves on the left, $\alpha = 0.05$ for the curves on the right.

In the next result, we assume that we have at hand an upper bound for $m(1)$. Such a bound can be easily found, using for instance the empirical mean.

Theorem 2.4. *Let $\bar{m} \geq m(1)$, $\gamma > 1$, $\alpha \in]0, 1[$, and denote a, b the positive*

numbers defined by

$$(n-1)a = n \frac{\bar{m}^2}{2} + \max(9\bar{m}^3, 4\bar{m}\gamma^2(\ln n)^2) \sqrt{\frac{\gamma}{2} n \ln n}$$

$$(n-1)b = \max(4\gamma \ln n, 6\bar{m}).$$

Then, under $\mathbf{H}_0$, we have $\mathbb{P}(\hat{S}_1 \geq \underline{q}) \leq \alpha$, provided $15n^{1-\gamma} \leq \alpha$ and

$$\underline{q} = \sqrt{\frac{8a}{n-1} \ln\left(\frac{1}{\alpha}\right)} + \frac{2b}{3} \ln\left(\frac{1}{\alpha}\right) + \frac{5b}{\alpha} n^{1-\gamma} + 8\sqrt{\frac{a}{\alpha} n^{-\gamma}}.$$

For a comparison with the asymptotic rejection region defined by (2.5), we observe that when $\gamma > 2$, we have the following asymptotic expansion:

$$\underline{q} = \frac{2\bar{m}}{\sqrt{n}} \sqrt{\ln\left(\frac{1}{\alpha}\right)} + O\left(\frac{(\ln n)^{5/4}}{n}\right).$$

Recall that, as $\alpha \rightarrow 0$,

$$\frac{q_{1-\alpha/2}}{\sqrt{\ln\left(\frac{1}{\alpha}\right)}} = \sqrt{2} + o(1).$$

Here, we can note the closeness of asymptotic rejection region (2.5) and non-asymptotic one, as given in previous theorem.

In view of a power study, we restrict the alternative hypothesis $\mathbf{H}_1$ to the local alternative hypothesis $\mathbf{H}'_1 : \sigma^2(t) \geq m(t) + s$ for some $t \in [0, 1]$. Here, s may depend on n . Our goal here is to give a lower bound for the power of the test, whenever s is in a range of values ensuring that the test is asymptotically of maximal power. With this respect, we observe that whenever s tends to 0, it must not converge too fastly, otherwise our test may have an asymptotic power strictly less than 1.

Theorem 2.5. Let $\bar{\ell} \geq \ln \mathbb{E}e^{N_1}$, $\lambda > 0$ and s such that

$$s\sqrt{n-1} \geq 4(\bar{\ell} + (\lambda + 1) \ln n)^2 \sqrt{\lambda \ln n}.$$

Then, under the alternative hypothesis $\mathbf{H}'_1$, we have

$$\mathbb{P}(\hat{S}_1 \geq \underline{q}) \geq 1 - \frac{1}{n^\lambda} \left[1 + \exp\left(\frac{1}{\lambda^2} (\sqrt{n-1} \underline{q} + e^{\bar{\ell}} n^{-\lambda/2})\right) \right],$$

where $\underline{q}$ is defined in Theorem 2.4.

For a more readable version of the power of the test statistic, note that, whenever $\gamma > 2$ and $\lambda > 1$, we have for large n :

$$\mathbb{P}(\hat{S}_1 \geq \underline{q}) \geq 1 - \frac{1}{n^\lambda} \left[1 + \exp \left(\frac{2\bar{m}}{\lambda^2} \sqrt{\ln \left(\frac{1}{\alpha} \right)} \right) \right] + O \left(\frac{(\ln n)^{5/4}}{n^{\lambda+1/2}} \right).$$

Roughly, the power of the test tends to 1 at a polynomial rate.

2.3 Simulation study

In this section we illustrate the good properties of our tests with a simulation study. We consider n_{MC} replications of Monte Carlo simulations and we study the performances of our tests in terms of asymptotic level and rejection power.

Throughout this section, we fix $T = 1$. In the following, function λ denotes the intensity function of N , *i.e.* the first derivative of its cumulative intensity function Λ .

2.3.1 Asymptotic level study

Model 1. We first consider the following model for the asymptotic level study,

$$\lambda(t) = \beta t^{\beta-1}, \text{ with } \beta > 0. \quad (2.6)$$

This is the intensity function of a so-called Weibull process, which is frequently used in Reliability Theory for instance. Remark that function λ is decreasing for $\beta < 1$, constant for $\beta = 1$ and increasing for $\beta > 1$.

In Table 2.1 we evaluate the empirical rejection frequency for both tests using the rejection regions defined in (2.2) for levels $\alpha = 5\%$ and $\alpha = 10\%$. This evaluation is based on $n_{MC} = 10,000$ Monte Carlo simulations (for each value of $\beta \in \{1/2, 1, 2\}$), with sample sizes $n = 100$ or $n = 500$. We note that test statistic $\hat{S}_1$ has a similar behavior than test statistic $\hat{S}_2$ at this range. For both statistics, the empirical rejection frequency is close to the nominal value even with the smallest sample size $n = 100$.

Model 2. We define the second model as follows,

$$\lambda(t) = \exp(\theta t), \text{ with } \theta > 0. \quad (2.7)$$

n	$\beta = 1/2$		$\beta = 1$		$\beta = 2$	
	100	500	100	500	100	500
	$\alpha = 5\%$					
$\hat{S}_1$	6.55	5.84	5.99	5.52	5.69	5.56
$\hat{S}_2$	6.27	5.38	5.74	5.44	6.00	5.71
	$\alpha = 10\%$					
$\hat{S}_1$	10.96	10.29	10.25	9.95	10.45	10.25
$\hat{S}_2$	10.55	10.05	10.38	10.02	10.29	10.06

Table 2.1: Empirical rejection frequency (%) for $n_{MC} = 10,000$ under the model (2.6).

Again, we evaluate in Table 2.2 the empirical rejection frequency for both tests using the rejection regions defined in (2.2) for levels $\alpha = 5\%$ and $\alpha = 10\%$, and this evaluation is based on $n_{MC} = 10,000$ Monte Carlo simulations, for each value of $\theta \in \{1/2, 1, 2\}$ and sample sizes $n = 100$ or $n = 500$. The same remarks as for model (2.6) apply here.

n	$\theta = 1/2$		$\theta = 1$		$\theta = 2$	
	100	500	100	500	100	500
	$\alpha = 5\%$					
$\hat{S}_1$	5.95	5.55	5.81	5.74	6.15	5.41
$\hat{S}_2$	5.57	5.26	5.67	5.54	5.68	5.43
	$\alpha = 10\%$					
$\hat{S}_1$	10.74	10.23	10.42	10.61	11.20	10.00
$\hat{S}_2$	10.19	10.40	10.14	10.56	10.60	9.77

Table 2.2: Empirical rejection frequency (%) for $n_{MC} = 10,000$ under the model (2.7).

2.3.2 Rejection power study

Model 1. For the asymptotic power study, we first consider the model defined by

$$\lambda(t) = \exp(\theta Zt), \quad (2.8)$$

with $Z \sim 2 + \text{Beta}(\frac{1}{2}, \frac{1}{2})$ and $\theta \in [0, 1]$. Figure 2.2 represents the empirical rejection frequency for different values of θ between 0 and 1 in model (2.8). For $\theta = 0$, the simulated process is an homogeneous Poisson process with intensity 1. The process deviates from the homogeneous Poisson process as θ increases. We observe on Figure 2.2 that both tests catch this behavior for small θ 's. In both cases the power goes to 1 for higher values of θ . The test statistic $\hat{S}_1$ has better performances, its power curve being always above that of test statistic $\hat{S}_2$.

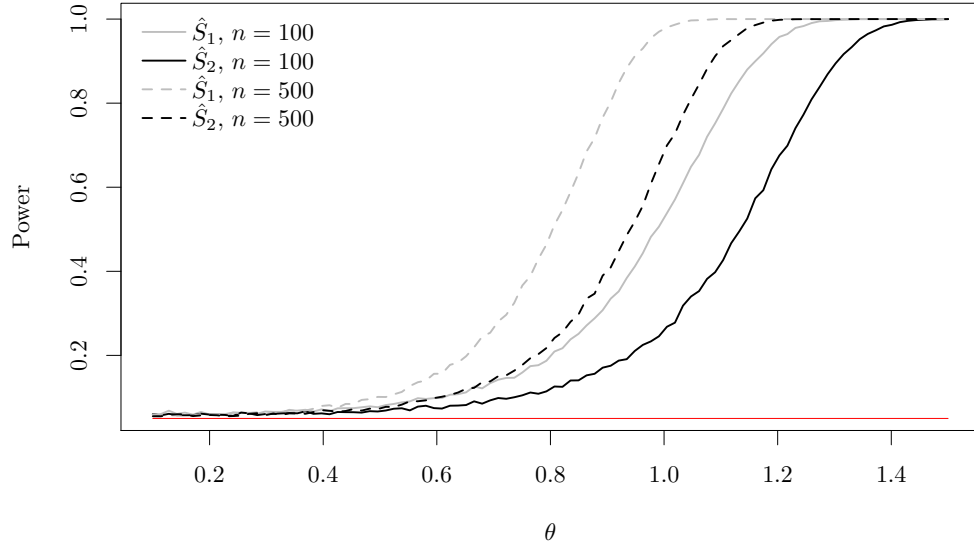


Figure 2.2: Empirical rejection frequency under (2.8) for $n_{MC} = 10,000$, 100 and 500 trajectories and $\alpha = 5\%$. The red horizontal line represents the value of α .

Model 2. We define the second model as follows,

$$\lambda(t) = \exp(\theta \sin(Z_t)), \quad (2.9)$$

with $(Z_t)_{t \in [0,1]}$ a standard Brownian Motion and $\theta \in [0, 1]$. This model differs from the previous one as the covariate depends on the time variable. Figure 2.3 represents the empirical rejection frequency for θ varying in $[0, 3]$ in model (2.9). The same remarks as for model (2.8) apply here. Note that the two test statistics look closer on this example.

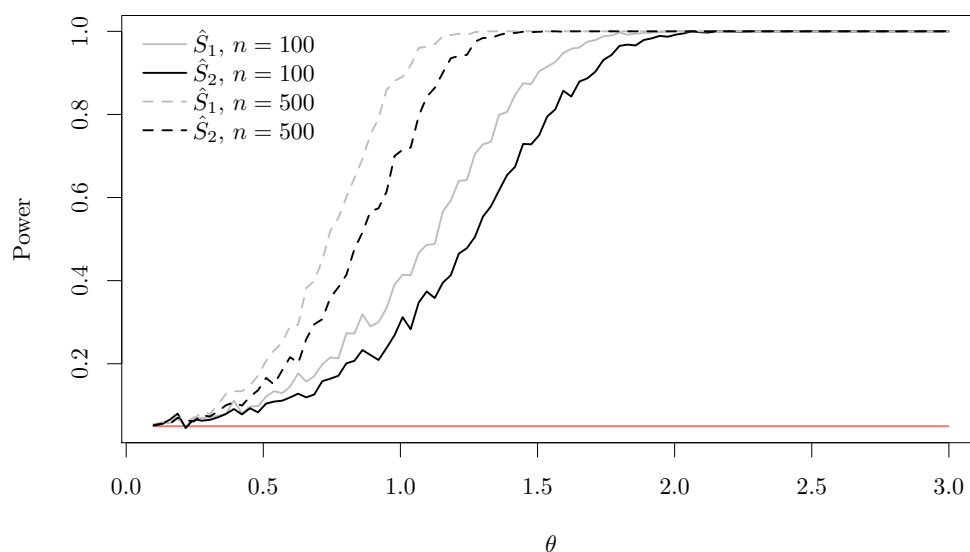


Figure 2.3: Empirical rejection frequency under (2.9) for $n_{MC} = 10,000$, 100 and 500 trajectories and $\alpha = 5\%$. The red horizontal line represents the value of α .

2.4 Application to real data

2.4.1 Analysis of some arrival times in a call center

The use of Poisson processes has been often considered as a first approach for modeling the arrival times in call centers and more generally queue systems. See Asmussen (2008) for the nice theory developed around this assumption.

As in the papers by Brown et al. (2005) and Mandelbaum et al. (2000), we consider a call center for an anonymous Israel's banks. A description of the calls received over the year 1999 is available online¹. The call center is open on weekdays (Sunday to Thursday in Israel), from 7 A.M. to midnight, and the calls are separated in different classes, depending on the needs of the customers. Each call can be described as follows. A customer calls one of the phone numbers of the call center. Except for rare busy signals, the customer is then connected to an interactive voice respond unit (IVR or VRU) and identifies himself/herself. While using the VRU, the customer receives recorded information. He/She can choose to perform some self-service transactions ($\sim 65\%$ of the calls) or indicate the need to speak with an agent ($\sim 35\%$). Here, we are interested in the latter, which represents roughly 30,000 to 40,000 calls per month. Each call record in the database also includes a categorical description of the type of service requested. The main call types are regular (PS in the database), stock transaction (NE), new or potential customer (NW), and Internet assistance (IN). Mandelbaum et al. (2000) and Brown et al. (2005) described the process of collecting and cleaning the data and provided complete descriptive analysis of the data.

In this study, we concentrate on IN calls recording between 3:25 P.M. and 3:35 P.M. all weekdays of year 1999. Times at which calls enter the VRU represent the arrival times of a counting process. The dataset then consists in 258 trajectories that we can assume to be realizations of *i.i.d.* Cox processes.

The results of the statistical study are presented in Table 2.3. One can see that the null hypothesis $\mathbf{H}_0$ is highly rejected using both statistics. This result suggests that even on a short period of time, these arrival times, which depend on complex human behavior, are strongly influenced by some covari-

¹<http://ie.technion.ac.il/serveng/callcenterdata/>

ates. One might easily imagine that weather conditions or other company intrinsic variables (*e.g.* number of recent opened accounts) could reduce this overdispersion and help to explain the number of IN phone calls.

	$\hat{S}_1$	$\hat{S}_2$
p -values	1.95×10^{-6}	1.05×10^{-6}

Table 2.3: p -values of both tests for the call center dataset.

Finally mention that Brown et al. (2005) also studied IN calls but did not reject the Poisson assumption. However, their study consists in testing the exponential distribution for the interarrival times of IN calls occurring in a single day, which is not compatible with our asymptotic and cannot help to determine if some covariates influence the daily calls process.

2.4.2 Analysis of the scoring times of a soccer team

As seen in Heuer et al. (2010), Poisson processes may also be used to model scoring goals during a soccer match. Nevertheless, one could suspect the influence of some covariates such as the behavior of the spectators or fitness fluctuations of the team under study. Thus, we propose to test the Poisson process assumption $\mathbf{H}_0$ for the scoring times of Arsenal soccer club first team.

To this end, we collected on the SoccerSTATS.com website their scoring times (in minutes) for each match in "Premier League" over three seasons (from 2012 to 2015). For each match, the scoring times of the team define the jump times of a counting process, giving a total of 229 counting paths. We can assume that these data are *i.i.d.* realizations of Cox processes.

The results of the statistical study are presented in Table 2.4. For both statistics $\hat{S}_1$ and $\hat{S}_2$, we cannot reject the null hypothesis $\mathbf{H}_0$ and the Poisson process seems to be a reasonable approximation for these counting processes.

	$\hat{S}_1$	$\hat{S}_2$
p -values	0.419	0.298

Table 2.4: p -values of both tests for the soccer goals dataset.

Recall that the analysis in Heuer et al. (2010) also suggests that the Poisson process is relevant for modeling the scoring goals of a given team in the German Bundesliga.

2.5 Proof of Theorem 2.1

In the rest of the chapter, we assume for notational simplicity that $T = 1$. Let $\hat{M} = (\hat{M}_t)_{t \in [0,1]}$ be the process defined by $\hat{M}_t = \hat{\sigma}^2(t) - \hat{m}(t)$ and $\hat{\tau} = (\hat{\tau}_t)_{t \in [0,1]}$ the process such that for all $t \in [0, 1]$,

$$\hat{\tau}_t = \frac{4}{n-1} \int_0^t \hat{\sigma}^2(s) dm(s).$$

Martingale properties and predictable σ -field are implicitly with respect to the natural filtration generated by the sample $N^{(1)}, \dots, N^{(n)}$. As usual, $\langle X \rangle$ stands for the predictable quadratic variation of a martingale X .

2.5.1 Auxiliary results

Lemma 2.6. *Under $\mathbf{H}_0$, the process $\hat{M}$ is a martingale, and $\langle \hat{M} \rangle = \hat{\tau}$.*

Proof. First we prove that $\hat{M}$ is a martingale. Observe that for all $t \in [0, 1]$,

$$\hat{\sigma}^2(t) = \frac{1}{n-1} \sum_{k=1}^n (N_t^{(k)})^2 - \frac{n}{n-1} \hat{m}(t)^2. \quad (2.10)$$

We now decompose $\hat{M}_t$ as a sum of stochastic integrals. In the sequel, we let $\bar{N}^{(k)} = N^{(k)} - m$. Note that $\bar{N}^{(k)}$ is a martingale. According to the integration by parts formula (Proposition 0.4.5 in the book by Revuz and Yor, 2013), we have

$$\begin{aligned} \sum_{k=1}^n (N_t^{(k)})^2 &= \sum_{k=1}^n \left[2 \int_0^t N_{s^-}^{(k)} dN_s^{(k)} + N_t^{(k)} \right] \\ &= 2 \sum_{k=1}^n \int_0^t N_{s^-}^{(k)} dN_s^{(k)} + n \hat{m}(t) \\ &= 2 \sum_{k=1}^n \int_0^t N_{s^-}^{(k)} d\bar{N}_s^{(k)} + 2n \int_0^t \hat{m}(s^-) dm(s) + n \hat{m}(t). \end{aligned} \quad (2.11)$$

Moreover, by the integration by parts formula,

$$\hat{m}(t)^2 = 2 \int_0^t \hat{m}(s^-) d\hat{m}(s) + \sum_{s \leq t} (\Delta \hat{m}(s))^2.$$

Using the fact that two independent Poisson processes do not jump at the same time (Proposition XII.1.5 in the book by Revuz and Yor, 2013), we deduce that

$$\begin{aligned} \sum_{s \leq t} (\Delta \hat{m}(s))^2 &= \frac{1}{n^2} \sum_{s \leq t} \left(\sum_{k=1}^n \Delta N_s^{(k)} \right)^2 = \frac{1}{n^2} \sum_{s \leq t} \sum_{k=1}^n \Delta N_s^{(k)} \\ &= \frac{1}{n} \hat{m}(t). \end{aligned}$$

Hence,

$$\hat{m}(t)^2 = 2 \int_0^t \hat{m}(s^-) d\hat{m}(s) + \frac{1}{n} \hat{m}(t).$$

Then, combining (2.10) and (2.11), we obtain

$$\begin{aligned} \hat{M}_t &= -\hat{m}(t) + \frac{1}{n-1} \sum_{k=1}^n (N_t^{(k)})^2 - \frac{n}{n-1} \hat{m}(t)^2 \\ &= \frac{2}{n-1} \sum_{k=1}^n \int_0^t N_{s^-}^{(k)} d\bar{N}_s^{(k)} - \frac{2n}{n-1} \int_0^t \hat{m}(s^-) d(\hat{m}(s) - m(s)). \quad (2.12) \end{aligned}$$

Since $\hat{m} - m$ and each of the $\bar{N}^{(k)}$'s are martingales and the integrands are predictable, we deduce that all the integrals in this formula are local martingales. It is a classical exercise to prove that they are of class DL (see Definition IV.1.6 in the book by Revuz and Yor, 2013), so that they are martingales, as well as $\hat{M}$.

In view of computing the predictable quadratic variation of $\hat{M}$, we first observe that by the integration by parts formula,

$$\hat{M}_t^2 = 2 \int_0^t \hat{M}_s^- d\hat{M}_s + \sum_{s \leq t} (\Delta \hat{M}_s)^2. \quad (2.13)$$

But, by (2.12),

$$\Delta \hat{M}_s = \frac{2}{n-1} \sum_{k=1}^n N_{s^-}^{(k)} \Delta N_s^{(k)} - \frac{2n}{n-1} \hat{m}(s^-) \Delta \hat{m}(s).$$

Again, we shall make use of the fact that two Poisson processes do not jump at the same time. Hence, if s is a time-jump for $N^{(k)}$,

$$\Delta \hat{M}_s = \frac{2}{n-1} (N_{s^-}^{(k)} - \hat{m}(s^-)) = \frac{2}{n-1} (N_{s^-}^{(k)} - \hat{m}(s^-)) \Delta N_s^{(k)}. \quad (2.14)$$

Thus,

$$\begin{aligned} \sum_{s \leq t} (\Delta \hat{M}_s)^2 &= \sum_{s \leq t} \sum_{k=1}^n (\Delta \hat{M}_s)^2 \mathbf{1}_{\{\Delta N_s^{(k)}=1\}} \\ &= \frac{4}{(n-1)^2} \sum_{k=1}^n \int_0^t (N_{s^-}^{(k)} - \hat{m}(s^-))^2 dN_s^{(k)}. \end{aligned}$$

By (2.13), we have

$$\begin{aligned} \hat{M}_t^2 &= 2 \int_0^t \hat{M}_{s^-} d\hat{M}_s + \frac{4}{(n-1)^2} \sum_{k=1}^n \int_0^t (N_{s^-}^{(k)} - \hat{m}(s^-))^2 d\bar{N}_s^{(k)} \\ &\quad + \frac{4}{(n-1)^2} \sum_{k=1}^n \int_0^t (N_{s^-}^{(k)} - \hat{m}(s^-))^2 dm(s). \end{aligned}$$

As above, we can conclude from the fact that both $\hat{M}$ and $\bar{N}^{(k)}$ are martingales that the first two terms on the right-hand side are martingales. Last term, namely

$$\frac{4}{n-1} \int_0^t \hat{\sigma}^2(s^-) dm(s) = \frac{4}{n-1} \int_0^t \hat{\sigma}^2(s) dm(s),$$

where equality holds by continuity of m , is predictable. Hence, it is the predictable quadratic variation of $\hat{M}$. $\square$

Lemma 2.7. *Under $\mathbf{H}_0$, we have:*

$$n \mathbb{E} \sup_{t \leq 1} |\Delta \hat{M}_t|^2 \rightarrow 0.$$

Proof. First observe that for all $u \in]0, 1[$:

$$\begin{aligned} n \mathbb{E} \sup_{t \leq 1} |\Delta \hat{M}_t|^2 &\leq u + n \mathbb{E} \sup_{t \leq 1} |\Delta \hat{M}_t|^2 \mathbf{1}_{\{\sqrt{n} \sup_{t \leq 1} |\Delta \hat{M}_t| > u\}} \\ &\leq u + 2 \int_0^\infty x \mathbb{P}(\sqrt{n} \sup_{t \leq 1} |\Delta \hat{M}_t| \mathbf{1}_{\{\sqrt{n} \sup_{t \leq 1} |\Delta \hat{M}_t| > u\}} > x) dx \\ &\leq 2u + 2 \int_u^\infty x \mathbb{P}(\sqrt{n} \sup_{t \leq 1} |\Delta \hat{M}_t| > x) dx. \end{aligned} \quad (2.15)$$

But, according to (2.14), if t is a time-jump for $N^{(k)}$, we have

$$\Delta \hat{M}_t = \frac{2}{n-1} (N_{t^-}^{(k)} - \hat{m}(t^-)),$$

and hence,

$$\sup_{t \leq 1} |\Delta \hat{M}_t| \leq \frac{2}{n-1} \sup_{k \leq n} \sup_{t \leq 1} |N_t^{(k)} - \hat{m}(t)|.$$

Thus, for all $x > 0$:

$$\begin{aligned} \mathbb{P}(\sqrt{n} \sup_{t \leq 1} |\Delta \hat{M}_t| \geq x) &\leq n \mathbb{P}\left(\sup_{t \leq 1} |N_t^{(1)} - \hat{m}(t)| \geq \frac{(n-1)x}{2\sqrt{n}}\right) \\ &\leq 8 \sup_{t \leq 1} \mathbb{E}|N_t^{(1)} - \hat{m}(t)|^3 \frac{n^{5/2}}{(n-1)^3 x^2}, \end{aligned}$$

where the last inequality is due to Doob's Inequality (see Revuz and Yor, 2013) applied to the martingale $N^{(1)} - \hat{m}$. A direct calculation shows that exists a constant $C > 0$ (independent of n) such that

$$\sup_{t \leq 1} \mathbb{E}|N_t^{(1)} - \hat{m}(t)|^3 \leq C.$$

Consequently, by (2.15):

$$\begin{aligned} n \mathbb{E} \sup_{t \leq 1} |\Delta \hat{M}_t|^2 &\leq 2u + \frac{16Cn^{5/2}}{(n-1)^3} \int_u^\infty \frac{dx}{x^2} \\ &\leq 2u + \frac{16Cn^{5/2}}{u(n-1)^3}. \end{aligned}$$

Taking for instance $u = n^{-1/4}$ gives the result. $\square$

2.5.2 Proof of Theorem 2.1

According to Theorem VIII.3.22 in the book by Jacod and Shiryaev (2013), the sequence of square integrable martingales $(\sqrt{n}\hat{M})_n$ converges in distribution to a continuous Gaussian martingale M such that $\langle M \rangle = 2m^2$ if, for all $t \in [0, 1]$ and $\varepsilon > 0$,

$$\langle \sqrt{n}\hat{M} \rangle_t \rightarrow 2m(t)^2 \text{ and } \int_{\mathbb{R} \times [0, t]} |x|^2 \mathbf{1}_{\{|x| > \varepsilon\}} \nu_n(dx, ds) \rightarrow 0, \quad (2.16)$$

both in probability, where ν_n stands for the predictable compensator of the random jump measure associated to the martingale $\sqrt{n}\hat{M}$. Regarding the first property, we know from Lemma 2.6 that

$$\lim_{n \rightarrow \infty} \langle \sqrt{n}\hat{M} \rangle_t = \lim_{n \rightarrow \infty} n\hat{\tau}_t = 4 \int_0^t \sigma^2(s) dm(s),$$

in probability. Since $\sigma^2 = m$ under $\mathbf{H}_0$, we deduce that $\lim_{n \rightarrow \infty} \langle \sqrt{n}\hat{M} \rangle_t = 2m(t)^2$ in probability. In order to prove the second property in (2.16), we fix $\varepsilon > 0$ and we let U and V be the processes defined for all $t \in [0, 1]$ by

$$U_t = \int_{\mathbb{R} \times [0, t]} |x|^2 \mathbf{1}_{\{|x| > \varepsilon\}} \nu_n(dx, ds) \text{ and } V_t = n \sum_{s \leq t} |\Delta \hat{M}_s|^2 \mathbf{1}_{\{\sqrt{n}|\Delta \hat{M}_s| > \varepsilon\}}.$$

Observing that U is L-dominated by the increasing adapted process V , we deduce from the Lenglart Inequality (see p.35 in the book by Jacod and Shiryaev, 2013) that for all $t \in [0, 1]$ and $\alpha, \eta > 0$:

$$\mathbb{P}(U_t \geq \eta) \leq \frac{1}{\eta} (\alpha + \mathbb{E} \sup_{s \leq t} \Delta V_s) + \mathbb{P}(V_t \geq \alpha).$$

But, $\{V_t > 0\} = \{\sqrt{n} \sup_{s \leq t} |\Delta \hat{M}_s| > \varepsilon\}$ and $\sup_{s \leq t} \Delta V_s \leq n \sup_{s \leq t} |\Delta \hat{M}_s|^2$. Thus, letting $\alpha \searrow 0$, we obtain with the help of Markov's Inequality:

$$\begin{aligned} \mathbb{P}(U_t \geq \eta) &\leq \frac{n}{\eta} \mathbb{E} \sup_{s \leq t} |\Delta \hat{M}_s|^2 + \mathbb{P}(\sqrt{n} \sup_{s \leq t} |\Delta \hat{M}_s| > \varepsilon) \\ &\leq \left(\frac{1}{\eta} + \frac{1}{\varepsilon^2} \right) n \mathbb{E} \sup_{s \leq t} |\Delta \hat{M}_s|^2. \end{aligned}$$

We conclude from Lemma 2.7 that U_t converges to 0 in probability. Hence, both properties in (2.16) are satisfied so that the sequence of square integrable martingales $(\sqrt{n}\hat{M})_n$ converges in distribution to a continuous Gaussian martingale M , with $\langle M \rangle = 2m^2$. The Dambis-Dubins-Schwarz Theorem (see Theorem V.1.6 in the book by Revuz and Yor, 2013) then gives $M = B_{2m^2}$, where B is a standard real Brownian Motion. $\square$

2.6 Proof of corollary 2.2

(i) Let $\mathbb{D}$ be the space of càdlàg functions from $[0, 1]$ to $\mathbb{R}$, equipped with the Skorokhod topology. By continuity of the application $\mathbb{D} \ni x \mapsto \sup_{t \leq T} x(t)$,

we deduce from Theorem 2.1 that

$$\sqrt{n} \hat{S}_1 = \sqrt{n} \sup_{t \leq 1} (\hat{\sigma}^2(t) - \hat{m}(t)) \xrightarrow{(law)} \sup_{t \leq 1} B_{2m(t)^2} = \sup_{t \leq 2m(1)^2} B_t.$$

According to the reflection principle (Proposition III.3.7 in the book by Revuz and Yor, 2013), the distribution of the latter term is $\sqrt{2} m(1) |\mathcal{N}(0, 1)|$. Hence the result follows for the statistics $\hat{S}_1$. Similarly, by continuity of $\mathbb{D} \ni x \mapsto \int_0^1 x(t) dt$, we have

$$\sqrt{n} \hat{S}_2 = \sqrt{n} \int_0^1 (\hat{\sigma}^2(t) - \hat{m}(t)) dt \xrightarrow{(law)} \int_0^1 B_{2m(t)^2} dt,$$

and the distribution of the limit is $\mathcal{N}(0, 4 \int_0^1 (1-t)m(t)^2 dt)$. Moreover, using the fact that $\hat{m} - m$ is a martingale, we easily prove with Doob's Inequality that $\sup_{t \leq 1} |\hat{m}(t) - m(t)|$ converges in probability to 0. Putting all pieces together and applying Slutsky's Lemma gives the result.

(ii) Under $\mathbf{H}_1$, there exists $t_0 \in [0, 1]$ such that $\sigma^2(t_0) > m(t_0)$. Then,

$$\begin{aligned} \sqrt{n} \hat{S}_1 &\geq \sqrt{n} (\hat{\sigma}^2(t_0) - \hat{m}(t_0)) \\ &\geq \sqrt{n} (\hat{\sigma}^2(t_0) - \sigma^2(t_0)) + \sqrt{n} (m(t_0) - \hat{m}(t_0)) + \sqrt{n} (\sigma^2(t_0) - m(t_0)). \end{aligned}$$

The latter term tends to $+\infty$, while the Central Limit Theorem (that can be used because $\mathbb{E}N_1^4 < \infty$) shows that the sequences induced by the first two terms on the right-hand side are stochastically bounded, hence the result with $\hat{S}_1$. Regarding $\hat{S}_2$, we first observe that under $\mathbf{H}_1$, $\int_0^1 (\sigma^2(t) - m(t)) dt > 0$, because σ and m are right-continuous functions and $\sigma^2 \geq m$. Thus, we only need to prove that the sequences $(\sqrt{n} \int_0^1 (m(t) - \hat{m}(t)) dt)_n$ and $(\sqrt{n} \int_0^1 (\hat{\sigma}^2(t) - \sigma^2(t)) dt)_n$ are stochastically bounded. Let us focus on the second sequence. We have

$$\begin{aligned} \sqrt{n} \int_0^1 (\hat{\sigma}^2(t) - \sigma^2(t)) dt &= \sqrt{n} \left(\frac{1}{n-1} \sum_{i=1}^n \int_0^1 (N_t^{(i)})^2 dt - \mathbb{E} \int_0^1 N_t^2 dt \right) \\ &\quad - \sqrt{n} \int_0^1 \left(\frac{n}{n-1} \hat{m}(t)^2 - m(t)^2 \right) dt. \end{aligned} \quad (2.17)$$

Since $\mathbb{E}N_1^4 < \infty$, the sequence induced by the first term on the right-hand side is stochastically bounded according to the Central Limit Theorem.

Regarding the latter term in (2.17), we observe that

$$\begin{aligned} \left| \int_0^1 \left(\frac{n}{n-1} \hat{m}(t)^2 - m(t)^2 \right) dt \right| &\leq 2(\hat{m}(1) + m(1)) \int_0^1 |\hat{m}(t) - m(t)| dt \\ &\quad + \frac{1}{\sqrt{n}} m(1)^2. \end{aligned}$$

By the Cauchy-Schwarz Inequality, there exists a constant $C > 0$ such that the L^1 -norm of the leftmost term is bounded by

$$\begin{aligned} C \left[1 + \sqrt{n} \mathbb{E}^{1/2} \left(\int_0^1 |\hat{m}(t) - m(t)| dt \right)^2 \right] &\leq C \left[1 + \left(\int_0^1 \text{var}(N_t) dt \right)^{1/2} \right] \\ &\leq C(1 + \mathbb{E} N_1^2). \end{aligned}$$

Thus, $(\sqrt{n} \int_0^1 (\hat{\sigma}^2(t) - \sigma^2(t)) dt)_n$ is stochastically bounded. $\square$

2.7 Proof of Theorem 2.3

In this section, we assume that $\mathbf{H}_1^n$ holds, hence in particular N is a Cox process with intensity $\lambda = \lambda_0 + d_n \Delta$ that depends on n .

For simplicity, we let $Z^{(n)} = (Z_t^{(n)})_{t \in [0,1]}$ the centered process defined for all $t \in [0, 1]$ by

$$Z_t^{(n)} = \sqrt{n}(\hat{\sigma}^2(t) - \sigma^2(t) + m(t) - \hat{m}(t)).$$

2.7.1 Auxiliary results

Lemma 2.8. *There exists $C > 0$ independent of n such that for all $s, t \in [0, 1]$,*

$$\mathbb{E}|N_t - N_s|^6 \leq C|t - s|^6.$$

Proof. Without loss of generality, we assume that $s \leq t$. Recall that the distribution of $N_t - N_s$ given Λ follows a Poisson distribution with parameter $\Lambda(t) - \Lambda(s) = \int_s^t \lambda(u) du$, and that the sixth moment of a Poisson distribution with parameter $\mu > 0$ is bounded by some constant $C > 0$

multiplied by μ^6 . Thus, using Jensen's Inequality, we get

$$\begin{aligned}\mathbb{E}(N_t - N_s)^6 &= \mathbb{E} \mathbb{E}[(N_t - N_s)^6 | \Lambda] \\ &\leq C \mathbb{E} \left(\int_s^t \lambda(u) du \right)^6 \\ &\leq C(t-s)^5 \int_s^t \mathbb{E} \lambda^6(u) du \\ &\leq 2^6 C \left(\sup_{t \in [0,1]} \lambda_0(t)^6 + \sup_{t \in [0,1]} \mathbb{E} |\Delta_t^6| \right) (t-s)^6,\end{aligned}$$

hence the lemma. $\square$

Lemma 2.9. *For all $t \in [0, 1]$, $((Z_t^{(n)})^2)_n$ is a uniformly integrable sequence.*

Proof. According to the Rosenthal Inequality and Lemma 2.8, there exists a constant $C > 0$ that does not depend on n such that

$$n^{3/2} \mathbb{E} |\hat{m}(t) - m(t)|^3 \leq C \text{ and } n^{3/2} \mathbb{E} |\hat{\sigma}^2(t) - \sigma^2(t)|^3 \leq C.$$

Thus, we deduce that $\sup_n \mathbb{E} |Z_t^{(n)}|^3 < \infty$, which implies that $((Z_t^{(n)})^2)_n$ is a uniformly integrable sequence. $\square$

Lemma 2.10. *The sequence of processes $(Z^{(n)})_n$ is tight.*

Proof. For an integrable real random variable Z , we let $\{Z\} = Z - \mathbb{E}Z$. First observe that we have, for all $t \in [0, 1]$:

$$\begin{aligned}Z_t^{(n)} &= \sqrt{n} \left(\frac{1}{n-1} \mathbb{E} N_t^2 + \frac{1}{n-1} \sum_{i=1}^n \{(N_t^{(i)})^2\} \right. \\ &\quad \left. - \{\hat{m}(t)\}^2 + \{\hat{m}(t)\}(1 - 2m(t)) \right).\end{aligned}$$

We shall make use of the classical criterion of tightness (see e.g. Theorem 13.6 in the book by Billingsley, 2013). Clearly, the sequence will be proved to be tight if each of the sequences of processes defined by

$$X^{1,n} = \sqrt{n} \{\hat{m}\}, \quad X^{2,n} = \sqrt{n} \{\hat{m}\}^2 \text{ and } X^{3,n} = \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \{(N_t^{(i)})^2\}$$

satisfy the inequality

$$\mathbb{E}(X_t^{k,n} - X_s^{k,n})^2 \leq C(F(t) - F(s))^2, \quad \forall 0 \leq s \leq t \leq 1, \quad (2.18)$$

for a constant $C > 0$ and some nondecreasing and continuous function F , both independent of n . We only prove it for $X^{3,n}$. In the sequel, $C > 0$ is a constant, independent of n , and whose value may change from line to line. Observe that

$$\begin{aligned} \mathbb{E}(X_t^{3,n} - X_s^{3,n})^2 &= \frac{n}{n-1} \mathbb{E}(\{N_t^2\} - \{N_s^2\})^2 \\ &\leq C \mathbb{E}(N_t^2 - N_s^2)^2 + C(\mathbb{E}N_t^2 - \mathbb{E}N_s^2)^2 \\ &\leq C\{\mathbb{E}(N_t - N_s)^4\}^{1/2} + C\mathbb{E}(N_t - N_s)^2, \end{aligned}$$

by Cauchy-Schwarz's Inequality. Then, Lemma 2.8 gives

$$\mathbb{E}(X_t^{3,n} - X_s^{3,n})^2 \leq C(t - s)^2.$$

Consequently, (2.18) holds for $k = 3$, with the continuous and nondecreasing function $F(t) = t$. $\square$

Lemma 2.11. *Let $B = (B_t)_{t \in \mathbb{R}_+}$ be a real and standard Brownian Motion. Then, if m_0 is the function defined for all $t \in [0, 1]$ by $m_0(t) = \int_0^t \lambda_0(u) du$, we have*

$$Z^{(n)} \xrightarrow{(\text{law})} (B_{2m_0^2(t)})_{t \in [0,1]}.$$

Proof. In the sequel, for $p = 1$ or 2 , we let

$$\hat{m}_p(t) = \frac{1}{n} \sum_{i=1}^n (N_t^{(i)})^p \text{ and } m_p(t) = \mathbb{E}N_t^p.$$

Let $k \geq 1$ and $0 \leq t_1 < \dots < t_k \leq 1$. According to the Central Limit Theorem for triangular arrays (for instance the Lyapounov condition is easily seen to be true according to Lemma 2.8), we know that the $2k$ -dimensional random vector defined by

$$\sqrt{n} \begin{pmatrix} \hat{m}_1(t_j) - m_1(t_j) \\ \hat{m}_2(t_j) - m_2(t_j) \end{pmatrix}_{j=1, \dots, k}$$

converges to a normal distribution. Now apply the δ -method to deduce that the $3k$ -dimensional random vector

$$\sqrt{n} \begin{pmatrix} \hat{m}_1(t_j) - m_1(t_j) \\ \hat{m}_2(t_j) - m_2(t_j) \\ \hat{m}_1(t_j)^2 - m_1(t_j)^2 \end{pmatrix}_{j=1, \dots, k}$$

also converges to a normal distribution. Thus,

$$\sqrt{n}(\hat{m}_2(t_j) - \hat{m}_1(t_j)^2 - m_2(t_j) + m_1(t_j)^2 - \hat{m}_1(t_j) + m_1(t_j))_{j=1, \dots, k},$$

converges to a k -dimensional normal distribution with mean μ and covariance matrix Σ , as well as $(Z_{t_j}^{(n)})_{j=1, \dots, k}$. Since for all $j = 1, \dots, k$, $\hat{m}(t_j)$ and $\hat{\sigma}^2(t_j)$ are unbiased estimators of $m(t_j)$ and $\sigma^2(t_j)$, $\mathbb{E}Z_{t_j}^{(n)} = 0$. Thus, by Lemma 2.9, $\mu = 0$. We now proceed to compute the variance matrix Σ . Let $i, j = 1, \dots, k$. With the notation $\{Z\} = Z - \mathbb{E}Z$ for an integrable real random variable Z , we easily see that the difference between $\mathbb{E}Z_{t_i}^{(n)}Z_{t_j}^{(n)}$ and

$$n\mathbb{E}(\{\hat{m}_2(t_i)\} - \{\hat{m}_1^2(t_i)\} - \{\hat{m}_1(t_i)\})(\{\hat{m}_2(t_j)\} - \{\hat{m}_1^2(t_j)\} - \{\hat{m}_1(t_j)\})$$

tends to 0 as $n \rightarrow \infty$. Moreover, the difference between the latter term and

$$\frac{1}{n}\mathbb{E}\left(\sum_{\ell=1}^n\{(N_{t_i}^{(\ell)})^2 - N_{t_i}^{(\ell)}(2m(t_i) + 1)\}\right)\left(\sum_{\ell=1}^n\{(N_{t_j}^{(\ell)})^2 - N_{t_j}^{(\ell)}(2m(t_j) + 1)\}\right),$$

denoted by A , vanishes as well. But, by independence of the processes $(N^{(\ell)})_{\ell \leq n}$, we have

$$\begin{aligned} A &= \mathbb{E}(\{N_{t_i}^2\} - \{N_{t_i}\}(2m(t_i) + 1))(\{N_{t_j}^2\} - \{N_{t_j}\}(2m(t_j) + 1)) \\ &= \text{cov}(N_{t_i}^2 - (2m(t_i) + 1)N_{t_i}, N_{t_j}^2 - (2m(t_j) + 1)N_{t_j}). \end{aligned}$$

Recall that, under $\mathbf{H}_1^n$, N is a Cox process with intensity $\lambda = \lambda_0 + d_n\Delta$. Since $(d_n)_n$ tends to 0, easy calculations prove that, as $n \rightarrow \infty$, A converges to

$$\text{cov}(P_{t_i}^2 - (2m_0(t_i) + 1)P_{t_i}, P_{t_j}^2 - (2m_0(t_j) + 1)P_{t_j}), \quad (2.19)$$

where P is a Poisson process with intensity λ_0 . Now use the properties of the Poisson process and the fact that, for a random variable Q that follows a Poisson distribution with parameter $\mu > 0$, we have

$$\mathbb{E}(Q - \mu)^2 = \mathbb{E}(Q - \mu)^3 = \mu \text{ and } \mathbb{E}(Q - \mu)^4 = \mu + 3\mu^2.$$

Moreover, for all $t \in [0, 1]$, $P_t^2 - (2m_0(t) + 1)P_t + m_0(t)^2 = \{P_t\}^2 - \{P_t\}$. We easily deduce from the above properties and the independence of the increments of a Poisson process that the covariance in (2.19) equals $2m_0(t_i \wedge t_j)^2$. Thus, by Lemma 2.9, the (i, j) term of matrix Σ is given by the previous formula. The sequence of processes $(Z^{(n)})_n$ being tight according to Lemma 2.10, we have proved that $Z^{(n)}$ converges in distribution to a centered Gaussian process Z such that if $s, t \in [0, 1]$, $\mathbb{E}Z_s Z_t = 2m_0(s \wedge t)^2$. Such a Gaussian process can be written $B_{2m_0^2}$ where B is a standard Brownian Motion on the line, hence the result. $\square$

2.7.2 Proof of Theorem 2.3

Recall that, by assumption, $v(t) \neq 0$ for some $t \in [0, 1]$. Observe that according to (2.1), we have under $\mathbf{H}_1^n$, for all $t \in [0, 1]$:

$$\begin{aligned} \sigma^2(t) - m(t) &= \text{var}(\Lambda(t)) \\ &= \text{var}\left(\int_0^t (\lambda_0(s) + d_n \Delta_s) ds\right) \\ &= d_n^2 v(t). \end{aligned}$$

Thus,

$$\sqrt{n} \hat{S}_1 = \sup_{t \leq 1} (Z_t^{(n)} + \sqrt{n} d_n^2 v(t)).$$

Moreover, it is an easy exercise to prove that $\hat{m}(T) \rightarrow m_0(T)$ and $\hat{I} \rightarrow I_0$, both in probability. If $\sqrt{n} d_n^2 \rightarrow \infty$, we then have by Lemma 2.11,

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{\text{prob.}} +\infty,$$

and similarly for $\hat{S}_2$, hence (i). We now assume that $\sqrt{n} d_n^2 \rightarrow d < \infty$. By Lemma 2.11 and Slutsky's Lemma, the continuity of the underlying functional gives:

$$\sqrt{n} \frac{\hat{S}_2}{\hat{I}} \xrightarrow{(\text{law})} \frac{1}{I_0} \int_0^1 (B_{2m_0(t)^2} + dv(t)) dt.$$

Observing now that the distribution of the latter term equals

$$2\mathcal{N}(0, 1) + \frac{d}{I_0} \int_0^1 v(t) dt$$

gives the result for $\hat{S}_2$. Regarding $\hat{S}_1$,

$$\sqrt{n} \frac{\hat{S}_1}{\hat{m}(T)} \xrightarrow{(\text{law})} \frac{1}{m_0(T)} \sup_{t \leq 1} (B_{2m_0(t)^2} + dv(t)),$$

hence the theorem. $\square$

2.8 Proofs of Theorems 2.4 and 2.5

In the sequel, we let $T = 1$.

2.8.1 Auxiliary results

From now on, we let ψ be the function defined for all $x > 0$ by

$$\psi(x) = 2 \frac{(x+1) \ln(x+1) - x}{x^2}. \quad (2.20)$$

Lemma 2.12. *Under $\mathbf{H}_0$, we have for all $\varepsilon > 0$:*

$$\mathbb{P}(\sup_{t \leq 1} |\Delta \hat{M}_t| \geq \varepsilon) \leq 2n \exp \left(- \frac{(n-1)^2 \varepsilon^2}{8m(1)} \psi \left(\frac{(n-1)\varepsilon}{2m(1)} \right) \right).$$

Proof. According to (2.14), if t is a time-jump for $N^{(k)}$, we have

$$\Delta \hat{M}_t = \frac{2}{n-1} (N_{t^-}^{(k)} - \hat{m}(t^-)).$$

Hence,

$$\sup_{t \leq 1} |\Delta \hat{M}_t| \leq \frac{2}{n-1} \sup_{k \leq n} \sup_{t \leq 1} |N_t^{(k)} - \hat{m}(t)|,$$

and then

$$\mathbb{P}(\sup_{t \leq 1} |\Delta \hat{M}_t| \geq \varepsilon) \leq n \mathbb{P}(2 \sup_{t \leq 1} |N_t^{(1)} - \hat{m}(t)| \geq (n-1)\varepsilon).$$

Note that since $N^{(k)} - m$ is a martingale with jumps less than 1, $N^{(1)} - \hat{m}$ also is a martingale whose jumps are bounded by 1. Moreover, $(N^{(k)} - m)^2 - m$ is a martingale, so that the predictable quadratic variation of $N^{(1)} - m$

equals $(1 - 1/n)m$. Consequently, by above and the exponential inequality for martingales (see Shorack and Wellner, 2009):

$$\mathbb{P}(\sup_{t \leq 1} |\Delta \hat{M}_t| \geq \varepsilon) \leq 2n \exp \left(- \frac{(n-1)^2 \varepsilon^2}{8m(1)} \psi \left(\frac{(n-1)\varepsilon}{2m(1)} \right) \right),$$

which is the desired result. $\square$

Proposition 2.13. *Let $a, b, q > 0$. Under $\mathbf{H}_0$, $\mathbb{P}(\sup_{t \leq 1} \hat{M}_t > q)$ is smaller than*

$$\begin{aligned} & \exp \left(- \frac{(n-1)q^2}{8a} \psi \left(\frac{(n-1)qb}{4a} \right) \right) + \exp \left(- \frac{2((n-1)a - nm(1)^2/2)^2}{c^2 n} \right) \\ & + 2n \exp \left(- \frac{c}{2m(1)^2} \psi \left(\frac{\sqrt{c}}{m(1)^{3/2}} \right) \right) + 2n \exp \left(- \frac{(n-1)^2 b^2}{8m(1)} \psi \left(\frac{(n-1)b}{2m(1)} \right) \right). \end{aligned}$$

Proof. Let T be the stopping time defined by

$$T = \inf \left\{ t \in [0, 1] : \int_0^t \hat{\sigma}^2(s) dm(s) > a \text{ or } \sup_{s \leq t} |\Delta \hat{M}_s| > b \right\},$$

with the usual convention that $\inf \emptyset = +\infty$. Thus, for all $q > 0$:

$$\begin{aligned} \mathbb{P} \left(\sup_{t \leq 1} \hat{M}_t > q \right) & \leq \mathbb{P} \left(\sup_{t \leq 1} \hat{M}_t > q, T > 1 \right) + \mathbb{P}(T \leq 1) \\ & \leq \mathbb{P} \left(\sup_{t \leq T \wedge 1} \hat{M}_t > q \right) + \mathbb{P} \left(\int_0^1 \hat{\sigma}^2(s) dm(s) > a \right) \\ & \quad + \mathbb{P} \left(\sup_{t \leq 1} |\Delta \hat{M}_t| > b \right), \end{aligned} \tag{2.21}$$

by definition of T . We now proceed to bound the first term on the right-hand side in (2.21), denoted p_1 . Observe that by Lemma 2.6,

$$p_1 = \mathbb{P} \left(\sup_{t \leq T \wedge 1} \hat{M}_t > q, \hat{\tau}_{1 \wedge T} \leq \frac{4a}{n-1} \right).$$

Since $\hat{\tau}$ is the predictable compensator of martingale $\hat{M}$, we thus deduce from the exponential inequality for martingales (see Shorack and Wellner, 2009) that

$$p_1 \leq \exp \left(- \frac{(n-1)q^2}{8a} \psi \left(\frac{(n-1)qb}{4a} \right) \right). \tag{2.22}$$

We now bound the second term in the right-hand side of (2.21), further denoted p_2 . First observe that for all $s \leq 1$:

$$(n-1)\hat{\sigma}^2(s) = \sum_{k=1}^n (N_s^{(k)} - m(s))^2 - n(m(s) - \hat{m}(s))^2.$$

Hence,

$$\int_0^1 \hat{\sigma}^2(s) dm(s) \leq m(1) \frac{1}{n-1} \sum_{k=1}^n X_k,$$

where, for all $k = 1, \dots, n$, $X_k = \int_0^1 (N_s^{(k)} - m(s))^2 dm(s)$. Thus, for all $c > 0$:

$$\begin{aligned} p_2 &\leq \mathbb{P}\left(\sum_{k=1}^n X_k \wedge c > (n-1)a\right) + n\mathbb{P}(X_1 > c) \\ &\leq \mathbb{P}\left(\sum_{k=1}^n (X_k \wedge c - \mathbb{E}X_1 \wedge c) > (n-1)a - n\frac{m(1)^2}{2}\right) \\ &\quad + n\mathbb{P}\left(\sup_{t \leq 1} |N_t - m(t)| \geq \sqrt{\frac{c}{m(1)}}\right), \end{aligned}$$

since

$$\mathbb{E}X_1 \wedge c \leq \int_0^1 \mathbb{E}(N_s - m(s))^2 dm(s) = \frac{m(1)^2}{2}.$$

We now make use of Hoeffding inequality (see Shorack and Wellner, 2009) for the first probability on the right-hand side, and the exponential inequality to martingale $N - m$ for the second probability, to obtain that p_2 is smaller than

$$\exp\left(-\frac{2((n-1)a - nm(1)^2/2)^2}{c^2n}\right) + 2n \exp\left(-\frac{c}{2m(1)^2} \psi\left(\frac{\sqrt{c}}{m(1)^{3/2}}\right)\right).$$

By Lemma 2.12, (2.21), (2.22) and above, we deduce the desired result. $\square$

2.8.2 Proof of Theorem 2.4

We know from Shorack and Wellner (2009) that $\psi(x) \geq 1/(1+x/3)$ for all $x > 0$. Thus, according to Proposition 2.13, for all $a, b, c > 0$, we have $\mathbb{P}(\hat{S}_1 \geq q) \leq \alpha$ provided $q \geq q_0$, where q_0 is defined by

$$\exp\left(-\frac{3(n-1)q_0^2}{24a+2(n-1)q_0b}\right) = \alpha - \xi,$$

and $\xi = \xi_1 + \xi_2 + \xi_3$, such that $\xi \leq \alpha$. In this relation, we have denoted

$$\begin{aligned}\xi_1 &= \exp\left(-\frac{2((n-1)a - nm(1)^2/2)^2}{c^2n}\right), \\ \xi_2 &= 2n \exp\left(-\frac{3}{2m(1)^{1/2}} \frac{c}{3m(1)^{3/2} + \sqrt{c}}\right) \text{ and} \\ \xi_3 &= 2n \exp\left(-\frac{3}{4} \frac{(n-1)^2b^2}{6m(1) + (n-1)b}\right).\end{aligned}$$

Letting $\phi = \ln(\alpha - \xi)$, we deduce from easy calculations that

$$q_0 = -\frac{b\phi}{3} + \sqrt{\frac{b^2\phi^2}{9} - \frac{8a\phi}{n-1}}.$$

Using the inequalities $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x, y \geq 0$ and $-\ln(1-x) \leq 3x/2$ provided $x \in [0, 1/3]$, we obtain the following upper bound when $\xi \leq \alpha/3$:

$$\begin{aligned}q_0 &\leq -\frac{2b\phi}{3} + \sqrt{-\frac{8a\phi}{n-1}} \\ &\leq \frac{2b}{3} \ln\left(\frac{1}{\alpha}\right) + \sqrt{\frac{8a}{n-1} \ln\left(\frac{1}{\alpha}\right)} + \frac{b\xi}{\alpha} + \sqrt{\frac{12a\xi}{\alpha(n-1)}}.\end{aligned}\tag{2.23}$$

For $\gamma > 1$, we now set a, b, c such that

$$(n-1)b = 2 \max(2\gamma \ln n, 3\bar{m}), \quad c = \max(9\bar{m}^3, 4\bar{m}\gamma^2(\ln n)^2) \text{ and} \tag{2.24}$$

$$(n-1)a = n\frac{\bar{m}^2}{2} + c\sqrt{\frac{\gamma}{2}n \ln n}. \tag{2.25}$$

With this choice, we get

$$\begin{aligned}\xi_3 &\leq 2n \exp\left(-\max(\gamma \ln n, 3\bar{m})\right) \leq 2n^{1-\gamma}, \\ \xi_2 &\leq 2n \exp\left(-\frac{1}{2\sqrt{\bar{m}}} \sqrt{c}\right) \leq 2n^{1-\gamma}, \text{ and} \\ \xi_1 &\leq e^{-\gamma \ln n} = n^{-\gamma},\end{aligned}$$

thus $\xi \leq 5n^{1-\gamma}$. By (2.23), we can conclude that if $5n^{1-\gamma} \leq \alpha/3$:

$$q_0 \leq \frac{2b}{3} \ln\left(\frac{1}{\alpha}\right) + \sqrt{\frac{8a}{n-1} \ln\left(\frac{1}{\alpha}\right)} + \frac{5b}{\alpha} n^{1-\gamma} + 8\sqrt{\frac{a}{\alpha} n^{-\gamma}},$$

hence the theorem. $\square$

2.8.3 Proof of Theorem 2.5

Let $t_0 \in [0, 1]$ be such that $\sigma^2(t_0) \geq m(t_0) + s$. Then,

$$\begin{aligned} \mathbb{P}(\hat{S}_1 \geq \underline{q}) &\geq \mathbb{P}(\hat{\sigma}^2(t_0) - \hat{m}(t_0) \geq \underline{q}) \\ &\geq \mathbb{P}(\hat{\sigma}^2(t_0) - \sigma^2(t_0) + m(t_0) - \hat{m}(t_0) \geq \underline{q} - s) \\ &\geq 1 - \mathbb{P}(\hat{m}(t_0) - m(t_0) + \sigma^2(t_0) - \hat{\sigma}^2(t_0) > s - \underline{q}). \end{aligned}$$

The rest of the proof is devoted to give an upper bound for the last term. We shall make use of the Mac Diarmid inequality that requires boundedness of the random variables. With this in mind, we need to truncate the random variables. Thus, for all $K > 0$, we now introduce the quantities :

$$\hat{\sigma}_K^2 = \frac{1}{n-1} \sum_{i=1}^n (N_{t_0}^{(i)} - \hat{m}_K)^2, \text{ where } \hat{m}_K = \frac{1}{n} \sum_{i=1}^n \min(N_{t_0}^{(i)}, K),$$

and their theoretical counterparts

$$\sigma_K^2 = \text{var}(\min(N_{t_0}, K)) \text{ and } m_K = \mathbb{E} \min(N_{t_0}, K).$$

We observe that $m(t_0) - m_K = \mathbb{E} N_{t_0} - \mathbb{E} \min(N_{t_0}, K) \geq 0$, and

$$\begin{aligned} \sigma_K^2 - \sigma^2(t_0) &\geq \mathbb{E}(\min(N_{t_0}, K)^2 - N_{t_0}^2) \\ &\geq \mathbb{E}(K^2 - N_{t_0}^2) \mathbf{1}_{\{N_{t_0} > K\}} \\ &\geq -\mathbb{E} N_{t_0}^2 \mathbf{1}_{\{N_{t_0} > K\}} \\ &\geq -3\mathbb{E} e^{N_{t_0}} e^{-K/2} \\ &\geq -3e^{\bar{\ell} - K/2}, \end{aligned}$$

according to the Markov inequality, since $\bar{\ell} \geq \ln \mathbb{E} e^{N_{t_0}}$ and $x^4 \leq 5e^x$ for all $x \geq 0$. Considering the event $\{N_{t_0}^{(i)} \leq K, \text{ for all } i = 1, \dots, n\}$, we thus

deduce that

$$\begin{aligned}
& \mathbb{P}(\hat{m}(t_0) - m(t_0) + \sigma^2(t_0) - \hat{\sigma}^2(t_0) > s - \underline{q}) \\
& \leq \mathbb{P}(\hat{m}_K - m_K + \sigma_K^2 - \hat{\sigma}_K^2 > s - \underline{q} + m(t_0) - m_K + \sigma_K^2 - \sigma^2(t_0)) \\
& \quad + n\mathbb{P}(N_{t_0} > K) \\
& \leq \mathbb{P}(\hat{m}_K - m_K + \sigma_K^2 - \hat{\sigma}_K^2 > s - \underline{q} - 3e^{\bar{\ell}-K/2}) + ne^{\bar{\ell}-K}.
\end{aligned}$$

We now use the Mac Diarmid inequality (see Devroye et al., 1996) for the first term on the right hand side. The above sum is then smaller than

$$\begin{aligned}
& ne^{\bar{\ell}-K} + \exp\left(-\frac{n-1}{16}\left[\frac{s-\underline{q}-3e^{\bar{\ell}-K/2}}{K^2}\right]^2\right) \\
& \leq ne^{\bar{\ell}-K} + \exp\left(-\frac{(n-1)s^2}{16K^4}\left[1-\frac{2}{s}(\underline{q}+3e^{\bar{\ell}-K/2})\right]\right) \\
& \leq ne^{\bar{\ell}-K} + \exp\left(-\frac{(n-1)s^2}{16K^4} + \frac{(n-1)s}{8K^4}(\underline{q}+3e^{\bar{\ell}-K/2})\right),
\end{aligned}$$

where we used the inequality $(1-x)^2 \geq 1-2x$ for $x \geq 0$. We now choose $K = \bar{\ell} + (\lambda+1)\ln n$. Recalling that

$$(n-1)s^2 = 16(\bar{\ell} + (\lambda+1)\ln n)^4 \lambda \ln n,$$

we deduce that

$$\frac{s^2}{K^4} = \frac{16\lambda}{n-1} \ln n \text{ and } ne^{\bar{\ell}-K} = \frac{1}{n^\lambda}.$$

Thus, by above,

$$\begin{aligned}
& ne^{\bar{\ell}-K} + \exp\left(-\frac{n-1}{16}\left[\frac{s-\underline{q}-3e^{\bar{\ell}-K/2}}{K^2}\right]^2\right) \\
& \leq \frac{1}{n^\lambda} + \frac{1}{n^\lambda} \exp\left(\frac{\sqrt{n-1}}{2(\bar{\ell} + (\lambda+1)\ln n)^2} \sqrt{\lambda \ln n} (\underline{q} + 3e^{\bar{\ell}/2} n^{-(\lambda+1)/2})\right) \\
& \leq \frac{1}{n^\lambda} + \frac{1}{n^\lambda} \exp\left(\frac{1}{2\lambda^2} (\sqrt{n-1} \underline{q} + 3e^{\bar{\ell}} n^{-\lambda/2})\right).
\end{aligned}$$

As a conclusion, we have obtained the inequality

$$\begin{aligned}
& \mathbb{P}(\hat{m}(t_0) - m(t_0) + \sigma^2(t_0) - \hat{\sigma}^2(t_0) > s - \underline{q}) \\
& \leq \frac{1}{n^\lambda} + \frac{1}{n^\lambda} \exp\left(\frac{1}{\lambda^2} (\sqrt{n-1} \underline{q} + e^{\bar{\ell}} n^{-\lambda/2})\right),
\end{aligned}$$

hence the result. $\square$

Minimax regression estimation onto a Poisson point process covariate¹

Sommaire

3.1	Introduction	80
3.1.1	Regression estimation	80
3.1.2	Minimax regression for Poisson coprocess	80
3.1.3	Chaotic decomposition in the Poisson space	81
3.1.4	Organization of the chapter	82
3.2	Model	82
3.3	Minimax properties for known δ	84
3.3.1	Chaos estimator	84
3.3.2	Results	86
3.4	Adaptive properties for unknown δ	87
3.5	Proof of Theorem 3.1	88
3.5.1	Technical results	88
3.5.2	Proof of Theorem 3.1	93
3.6	Proof of Theorem 3.2	93
3.6.1	Technical results	94
3.6.2	Proof of Theorem 3.2	95
3.7	Proof of Theorem 3.3	98
3.7.1	Technical results	99
3.7.2	Proof of Theorem 3.3	105

¹This chapter has been accepted for publication in *ESAIM: Probability and Statistics*.

3.1 Introduction

3.1.1 Regression estimation

Regression estimation is a central problem in statistics. It is widely used and studied in the literature. Among all the methods explored to deal with the regression problem, nonparametric statistics have been widely investigated (see the monographies by Tsybakov, 2009 for a full introduction to nonparametric estimation and Györfi et al, 2006 for a clear account on nonparametric regression). A more recent challenge regarding this statistical problem is the regression onto a functional covariate (see the books by Ramsay and Silverman, 2006 and Horváth and Kokoszka, 2012 for more precision on functional data analysis). Although very challenging, the functional regression problem in the minimax setting has little coverage up to our knowledge. In the kernel estimation setting, Mas (2012) studied the small ball probabilities over some Hilbert spaces to derive minimax lower bounds at fixed points. More recently, Chagny and Roche (2016) derived minimax lower bounds at fixed points for adaptive nonparametric estimation of the regression under some Wiener measure domination assumptions on the small ball probabilities. Based on the k -nearest neighbor approach, Biau, Cérou and Guyader (2010) used compact embedding theory to get upper bounds on the minimax risk. See also the references therein for a more complete overview.

3.1.2 Minimax regression for Poisson coprocess

In this chapter, we focus on a regression problem for which the covariate is a Poisson point process. In the spirit of Cadre and Truquet (2015), we use a method based on the chaotic decomposition of Poisson functionals.

Let X be a Poisson point process on a compact domain $\mathbb{X} \subset \mathbb{R}^d$ equipped with its Borel σ -algebra $\mathcal{X}$. Letting δ_x the Dirac measure on $x \in \mathbb{X}$, the state space is identified to $\mathcal{S} = \{s = \sum_{i=1}^m \delta_{x_i} : m \in \mathbb{N}^*, x_i \in \mathbb{X}\}$ equipped with the smallest σ -algebra making the mappings $s \mapsto s(B)$ measurable for all Borel set B in $\mathcal{X}$. We denote by $\mathbb{P}_X$ the distribution of X whereas $\mathbb{L}^2(\mathbb{P}_X)$ denotes the space of all measurable functions $g : \mathcal{S} \rightarrow \mathbb{R}$ such that

$$\|g\|_{\mathbb{L}^2(\mathbb{P}_X)}^2 = \mathbb{E}g(X)^2 < +\infty.$$

Let $\mathbb{P}$ be a distribution on $\mathcal{S} \times \mathbb{R}$ and (X, Y) with law $\mathbb{P}$. Provided $\mathbb{E}|Y| < +\infty$ where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$, we consider the regression function $r : \mathcal{S} \rightarrow \mathbb{R}$ defined by $r(s) = \mathbb{E}(Y \mid X = s)$.

In this chapter we assume that r belongs to $\mathbb{L}^2(\mathbb{P}_X)$ and we aim at estimating r on the basis of an i.i.d. sample randomly drawn from the distribution $\mathbb{P}$ of (X, Y) . In this context, any measurable map $\tilde{r} : (\mathcal{S} \times \mathbb{R})^n \rightarrow \mathbb{L}^2(\mathbb{P}_X)$ is an estimator, the accuracy of which is measured by the risk

$$R_n(\tilde{r}, r) = \mathbb{E}^n \|\tilde{r} - r\|_{\mathbb{L}^2(\mathbb{P}_X)}^2,$$

where $\mathbb{E}^n$ denotes the expectation with respect to the distribution $\mathbb{P}^{\otimes n}$. Following the minimax approach, we define the maximal risk of $\tilde{r}$ over a class $\mathcal{P}$ of distributions for the random pair (X, Y) by

$$R_n(\tilde{r}, \mathcal{P}) = \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r).$$

We are interested in finding an estimator $\hat{r}$ such that

$$R_n(\hat{r}, \mathcal{P}) \asymp \inf_{\tilde{r}} R_n(\tilde{r}, \mathcal{P}),$$

where the infimum is taken over all possible estimates of r and $u_n \asymp v_n$ stands for $0 < \liminf_n u_n v_n^{-1} \leq \limsup_n u_n v_n^{-1} < +\infty$. Such an estimate is called asymptotically minimax over $\mathcal{P}$.

3.1.3 Chaotic decomposition in the Poisson space

Roughly, Itô's famous chaos expansion (see Itô, 1956 and Nualart and Vives, 1990 for technical details) says that every square integrable and $\sigma(X)$ -measurable random variable can be decomposed as a sum of multiple stochastic integrals, called chaos.

To be more precise, we now recall some basic facts about chaos decomposition in the Poisson space. Let μ be the mean measure of the Poisson point Process X , defined by $\mu(A) = \mathbb{E}X(A)$ for $A \in \mathcal{X}$, whenever $X(A)$ is the number of points of X lying in A . Fix $k \geq 1$. Provided $g \in \mathbb{L}^2(\mu^{\otimes k})$, we can define the k -th chaos $I_k(g)$ associated with g , namely

$$I_k(g) = \int_{\Delta_k} g d(X - \mu)^{\otimes k}, \quad (3.1)$$

where $\Delta_k = \{x \in \mathbb{X}^k : x_i \neq x_j \text{ for all } i \neq j\}$. In Nualart and Vives (1990), it is proved that every square integrable $\sigma(X)$ -measurable random variable can be decomposed as an infinite sum of chaos. Applied to our regression problem, this statement writes as

$$r(X) = \mathbb{E}Y + \sum_{k \geq 1} \frac{1}{k!} I_k(f_k), \quad (3.2)$$

where equality holds in $\mathbb{L}^2(\mathbb{P}_X)$, provided $\mathbb{E}Y^2 < \infty$. In the above formula, each f_k is an element of $\mathbb{L}_{\text{sym}}^2(\mu^{\otimes k})$ –the subset of symmetric functions in $\mathbb{L}^2(\mu^{\otimes k})$ –, and the decomposition is defined in a unique way.

3.1.4 Organization of the chapter

In this chapter we introduce a new estimator of the regression function r based on independent copies of (X, Y) and we study its minimax properties. Section 3.2 is devoted to the definition of a semiparametric model i.e. the construction of the family $\mathcal{P}$ of distributions of (X, Y) . In particular, we assume that r is a sum of δ chaos. In Section 3.3, we provide a lower bound for the minimax risk over $\mathcal{P}$. When δ is known, we prove that our estimator achieves this bound up to a logarithmic term. Finally, in Section 3.4, we define an adaptive procedure when δ is unknown, the risk of which is also proved to be optimal up to a logarithmic term. Last sections contain proofs.

3.2 Model

In the rest of the chapter, we let $\Theta \subset \mathbb{R}^p$. For each $\theta \in \Theta$, $\varphi_\theta : \mathbb{X} \rightarrow \mathbb{R}_+$ is a Borel function. The family $\{\varphi_\theta\}_{\theta \in \Theta}$ contains the constant function $\mathbf{1}_{\mathbb{X}}/\lambda(X)$, and is such that there exists three positive constants $\underline{\varphi}$, $\overline{\varphi}$ and γ_1 satisfying, for all $x, y \in \mathbb{X}$ and $\theta, \theta' \in \Theta$,

$$\underline{\varphi} \leq \varphi_\theta(x) \leq \overline{\varphi}, \quad (3.3)$$

$$|\varphi_\theta(x) - \varphi_{\theta'}(x)| \leq \overline{\varphi} |\theta - \theta'|, \quad (3.4)$$

$$|\varphi_\theta(x) - \varphi_\theta(y)| \leq \gamma_1 |x - y|, \quad (3.5)$$

where, here and in the following, $|\cdot|$ stands for the euclidean norm.

Let (X, Y) be a pair of random variables taking values in $\mathcal{S} \times \mathbb{R}$ with distribution $\mathbb{P}$, where $\mathcal{S}$ is the Poisson space over the compact domain $\mathbb{X} \subset$

$\mathbb{R}^d$. Here, X is a Poisson point process on $\mathbb{X}$ with intensity φ_θ , i.e. for all Borel set $A \in \mathcal{X}$:

$$\mathbb{E}X(A) = \int_A \varphi_\theta d\lambda, \quad (3.6)$$

where λ is the Lebesgue measure and $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$. In other words, the mean measure of X , say μ , has a Radon-Nikodým derivative φ_θ with respect to λ . We assume that for all $l \geq 1$, there exists an estimator

$$\tilde{\theta}_l : (\mathcal{S} \times \mathbb{R})^l \rightarrow \Theta,$$

such that,

$$\mathbb{E}^l |\tilde{\theta}_l - \theta|^2 \leq \frac{\kappa}{l+1}, \quad (3.7)$$

where $\kappa > 0$ is an absolute constant that does not depend on l and $\mathbb{E}^l$ is the expectation with respect to $\mathbb{P}^{\otimes l}$. As shown in Birgé (1983, Proposition 3.1), the above property is satisfied by a wide class of models, provided $\tilde{\theta}_l$ is a maximum likelihood estimate.

Moreover, the real-valued random variable Y satisfies, for some $u, M > 0$, the exponential moment condition:

$$\mathbb{E}Y^2 e^{u|Y|} \leq M, \quad (3.8)$$

As seen in (3.2), the regression function $r(s) = \mathbb{E}(Y \mid X = s)$ has a chaotic decomposition. In our model, we consider the case of a finite chaotic decomposition, i.e. there exists a strictly positive integer δ and $f_1 \in \mathbb{L}_{\text{sym}}^2(\mu), \dots, f_\delta \in \mathbb{L}_{\text{sym}}^2(\mu^{\otimes \delta})$ such that

$$r(X) = \mathbb{E}Y + \sum_{k=1}^{\delta} \frac{1}{k!} I_k(f_k), \quad (3.9)$$

where the $I_k(f_k)$'s are defined in (3.1). The coefficients f_k 's of the chaos lie in a nonparametric family, for which there exists two strictly positive constants γ_2 and $\bar{f}$ such that for all $k = 1, \dots, \delta$, and $x, y \in \mathbb{X}^k$

$$|f_k(x) - f_k(y)| \leq \gamma_2 |x - y|, \quad (3.10)$$

$$|f_k(x)| \leq \bar{f}. \quad (3.11)$$

Remarks (On the finiteness of the number of chaos).

1. Finiteness of the number of chaos roughly implies that the regression function $r(X)$ is unbounded. Indeed, consider for simplicity the case where $r(X)$ is decomposed onto only one chaos, i.e.

$$r(X) = \int f d(X - \lambda).$$

Here X is a simple Poisson process on the domain $\mathbb{X} = [0, 1]$ with unit intensity and f is any λ -integrable function on $\mathbb{X}$. Observe that, if $f \geq a > 0$, then

$$r(X) \geq aX([0, 1]) - \int f d\lambda.$$

Consequently, $r(X)$ is unbounded. The same tendency may be expected whatever the number of chaos.

2. Finiteness of the chaotic decomposition relies on the distribution of (X, Y) via the Malliavin calculus. Indeed, as proved in Proposition 4.1 from Last and Penrose (2011), the decomposition in δ chaos of $r(X)$ holds if and only if the $\delta + 1$ -th Malliavin derivative of r is null.

In the rest of the chapter, the constants $\varphi, \bar{\varphi}, \gamma_1, u, M, \delta, \gamma_2, \bar{f}$ and κ will be fixed, and we shall denote by $\mathcal{P}$ the set of distributions $\mathbb{P}$ of (X, Y) such that the assumptions (3.3)-(3.11) are satisfied. In this setting, θ implicitly denotes the true value of the parameter, that is φ_θ is the intensity of X (with mean measure μ).

3.3 Minimax properties for known δ

3.3.1 Chaos estimator

Our main goal is to construct an estimate of the regression function which achieves fast rates over $\mathcal{P}$. Let $\mathbb{P} \in \mathcal{P}$ and $(X, Y) \sim \mathbb{P}$ where X has mean measure $\mu = \varphi_\theta \cdot \lambda$.

First recall some basic facts about chaos decomposition in the Poisson space. If $g \in \mathbb{L}^2(\mu^{\otimes k})$ and $h \in \mathbb{L}^2(\mu^{\otimes l})$ for $k, l \geq 1$, we have the so-called Itô Isometry Formula:

$$\mathbb{E}I_k(g)I_l(f) = k! \int_{\mathbb{X}^k} \bar{g}h d\mu^{\otimes k} \mathbf{1}_{\{k=l\}} \text{ and } \mathbb{E}I_k(g) = 0, \quad (3.12)$$

where $\bar{g}$ and $\bar{h}$ are the symmetrizations of g and h , that is:

$$\bar{g}(x_1, \dots, x_k) = \frac{1}{k!} \sum_{\sigma} g(x_{\sigma(1)}, \dots, x_{\sigma(k)}), \quad (3.13)$$

for all $(x_1, \dots, x_k) \in \mathbb{X}^k$, the sum being taken over all permutations $\sigma = (\sigma(1), \dots, \sigma(k))$ of $\{1, \dots, k\}$, and similarly for $\bar{h}$.

Now let $\bar{W}$ be a strictly positive constant and W be a density on $\mathbb{X}$ such that $\sup_{\mathbb{X}} W \leq \bar{W}$. Furthermore, let $h_k = h_k(n) > 0$ a bandwidth to be tuned later on and denote

$$W_{h_k}(\cdot) = \frac{1}{h_k^d} W\left(\frac{\cdot}{h_k}\right).$$

One may easily deduce from relations (3.2) and (3.12) that

$$\mathbb{E} Y I_k(W_{h_k}^{\otimes k}(x - \cdot)) = \int_{\mathbb{X}^k} f_k W_{h_k}^{\otimes k}(x - \cdot) \varphi_{\theta}^{\otimes k} d\lambda^{\otimes k},$$

where, here and in the following, for any real-valued function g defined on $\mathbb{X}$, the notation $g^{\otimes k}$ denotes the real-valued function on $\mathbb{X}^k$ such that

$$g^{\otimes k}(x) = \prod_{i=1}^k g(x_i), \quad x = (x_1, \dots, x_k) \in \mathbb{X}^k.$$

Thus, under the smoothness assumptions (3.3) on φ_{θ} and (3.11) on f_k , the right-hand side converges to $f_k(x) \varphi_{\theta}^{\otimes k}(x)$, provided $h_k \rightarrow 0$.

Now let $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. with distribution $\mathbb{P}$. Based on this observation, a semiparametric estimate denoted $\hat{I}_{k, h_k}(X)$ of the k -th chaos $I_k(f_k)$ of (3.1) may be defined as follows:

$$\frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} \frac{W_{h_k}^{\otimes k}(x - y)}{\varphi_{\hat{\theta}_i}^{\otimes k}(x)} (X_i - \varphi_{\hat{\theta}_i} \cdot \lambda)^{\otimes k}(dy) (X - \varphi_{\hat{\theta}_i} \cdot \lambda)^{\otimes k}(dx), \quad (3.14)$$

where $T_n > 0$ is a truncation parameter to be tuned later on and the $\hat{\theta}_i$'s are the leave-one-out estimates defined by $\hat{\theta}_i = \tilde{\theta}_{n-1}((X_j)_{j \leq n, j \neq i})$ (see Section 3.2).

3.3.2 Results

Based on the estimate (3.14) of the k -th chaos, we may define the following empirical mean type estimator of the regression function r for any strictly positive integer l

$$\hat{r}_l(X) = \bar{Y}_n + \sum_{k=1}^l \frac{1}{k!} \hat{I}_{k,h_k}(X), \quad (3.15)$$

where $\bar{Y}_n$ is the empirical mean of $Y_1, \dots, Y_n$.

In this subsection, we study the performance of the estimate $\hat{r}_\delta$ of the regression function from a minimax point of view when the number of chaos δ is known.

Theorem 3.1. *Let $\varepsilon > 0$ and set $T_n = (\ln n)^{1+\varepsilon}$ and $h_k = (T_n^2 n^{-1})^{1/(2+dk)}$. Then,*

$$\limsup_{n \rightarrow +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}} \right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}} R_n(\hat{r}_\delta, r) < \infty.$$

Remark. Thus, the optimal rate of convergence over $\mathcal{P}$ is upper bounded by $((\ln n)^{2+2\varepsilon} n^{-1})^{2/(2+d\delta)}$. Here it is noticeable that, up to a logarithmic factor, we recover the optimal rate $n^{-2/(2+d\delta)}$ corresponding to the $d\delta$ -dimensional regression with a Lipschitz regression function (see, e.g., Theorem 1 in Kohler et al., 2009).

In our next result, we provide a lower bound for the optimal rate of convergence over $\mathcal{P}$ in order to assess the tightness of the upper bound obtained in Theorem 3.1.

Theorem 3.2. *We have,*

$$\liminf_{n \rightarrow +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r) > 0,$$

where the infimum is taken over all estimates $\tilde{r}$.

Remark. Theorem 3.2 indicates that the optimal rate of convergence over $\mathcal{P}$ is lower bounded by $n^{-2/(2+d\delta)}$ which, up to a logarithmic factor, corresponds to the upper bound found in Theorem 3.1. As a conclusion, up to a logarithmic factor, the estimate $\hat{r}_\delta$ is asymptotically minimax on $\mathcal{P}$.

3.4 Adaptive properties for unknown δ

We now consider the case of an unknown number of chaos δ . For $m > 0$, we set

$$\mathcal{P}(m) = \{\mathbb{P} \in \mathcal{P} : \|f_k\| \geq m; k \in 1, \dots, \delta\},$$

where $\|\cdot\|$ stands for the $\mathbb{L}^2$ -norm relatively to the Lebesgue measure. Thus, whenever $\mathbb{P} \in \mathcal{P}(m)$,

$$\delta = \min(k : \|f_k\| = 0) - 1.$$

Based on this observation, a natural estimate of δ may be obtained as follows. Let us assume that the dataset is of size $2n$, and let $(X_i, Y_i)_{1 \leq i \leq 2n}$ be i.i.d. with distribution $\mathbb{P} \in \mathcal{P}(m)$. For $k \in 1, \dots, \delta$, we introduce the empirical counterpart of $\varphi_\theta f_k$ defined by

$$\hat{g}_k(x) = \frac{1}{n} \sum_{i=n+1}^{2n} Y_i \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) (X_i - \varphi_{\hat{\theta}} \cdot \lambda)^{\otimes k}(\mathrm{d}y), \quad (3.16)$$

where $\hat{\theta} = \tilde{\theta}_n(X_{n+1}, \dots, X_{2n})$ is defined in Section 3.2, and $b_k = b_k(n)$ is a bandwidth to be tuned later. The estimator $\hat{\delta}$ of δ is then defined by

$$\hat{\delta} = \min(k : \|\hat{g}_k\| \leq \rho_k) - 1, \quad (3.17)$$

where $\rho_k = \rho_k(n)$ is a vanishing sequence of positive numbers that we choose later on. We may now define the adaptative estimator $\hat{r}$ of r by

$$\hat{r} = \hat{r}_{\hat{\delta}},$$

where $\hat{r}_l$ is defined in (3.15) for all strictly positive integer l .

Theorem 3.3. *Let $\varepsilon > d\delta \geq 2$, $\alpha, \beta > 0$ such that $\alpha + \beta < 1$ and $2\alpha + \beta > 1/(2 + d\delta)$, and set $T_n = (\ln n)^{1+\varepsilon}$. Then, if we take for all integer k ,*

$$h_k = (T_n^2 n^{-1})^{1/(2+dk)}, \quad \rho_k = ((2k)!)^2 n^{(\alpha+\beta-1)/2} \quad \text{and} \quad b_k = n^{-\beta/(2dk)},$$

we obtain, for all $m > 0$,

$$\limsup_{n \rightarrow +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}} \right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}(m)} R_n(\hat{r}, r) < +\infty.$$

Remark. Here it is noticeable that despite the estimation of the number of chaos δ and up to a logarithmic factor, we recover the optimal rate $n^{-2/(2+d\delta)}$ of Theorems 3.1 and 3.2.

3.5 Proof of Theorem 3.1

In this section, we assume without loss of generality that the constants $\bar{\varphi}$, γ_1 , γ_2 , $\bar{f}$, $\lambda(\mathbb{X})$ and $\bar{W}$ are greater than 1 and that $\underline{\varphi}$ is smaller than 1. Moreover, C denotes a positive number that only depends on the parameters of the model, i.e. $u, \varphi, \bar{\varphi}, \gamma_1, \gamma_2, \bar{f}, \delta, \theta, \kappa, \lambda(\mathbb{X}), M$ and W , and whose value may change from line to line.

We let $\mathbb{P} \in \mathcal{P}$ and, for simplicity, we may denote $\mathbb{E} = \mathbb{E}^n$ and var stands for the variance relatively to $\mathbb{P}^{\otimes n}$. Finally, let $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. with distribution $\mathbb{P}$.

3.5.1 Technical results

Let $k \geq 1$ be fixed and denote for all $x, y \in \mathbb{X}$ and $i = 1, \dots, n$:

$$g_i(x, y) = \frac{W_{h_k}(x - y)}{\varphi_{\hat{\theta}_i}(x)} \text{ and } g(x, y) = \frac{W_{h_k}(x - y)}{\varphi_{\theta}(x)}. \quad (3.18)$$

We also let

$$d\hat{X}_i = dX_i - \varphi_{\hat{\theta}_i} d\lambda, \quad d\hat{X}'_i = dX - \varphi_{\hat{\theta}_i} d\lambda, \quad (3.19)$$

$$d\tilde{X}_i = dX_i - \varphi_{\theta} d\lambda \text{ and } d\tilde{X} = dX - \varphi_{\theta} d\lambda. \quad (3.20)$$

With this respect, we have (see (3.14)):

$$\hat{I}_{k, h_k}(X) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(dy) \hat{X}_i'^{\otimes k}(dx).$$

Furthermore, denote for all $x \in \mathbb{X}^k$:

$$Z_{i, k}(x) = Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}_i^{\otimes k}(dy). \quad (3.21)$$

Lemma 3.4. *Let $i = 1, \dots, n$ and $k \leq \delta$ be fixed. Then for all $x \in \mathbb{X}^k$:*

$$\text{var}(Z_{i, k}(x)) \leq T_n^2 \frac{k! C^k}{h_k^{dk}}, \text{ and}$$

$$|\mathbb{E} Z_{i, k}(x) - f_k(x)| \leq C^k \left(\phi \frac{k!}{h_k^{dk}} \right)^{1/2} + C^k h_k,$$

where $\phi = \mathbb{E} Y^2 \mathbf{1}_{|Y| > T_n}$.

Proof. On the one hand, by the isometry formula (3.12) over the set $\mathcal{P}$,

$$\begin{aligned} \text{var}(Z_{i,k}(x)) &\leq T_n^2 \mathbb{E} \left(\int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \\ &\leq T_n^2 k! \int_{\mathbb{X}^k} \overline{g^{\otimes k}}^2(x, y) \varphi_\theta^{\otimes k}(y) \mathrm{d}y \\ &\leq T_n^2 \frac{\overline{W}^k \underline{\varphi}^k}{\underline{\varphi}^{2k}} \frac{k!}{h_k^{dk}}, \end{aligned}$$

where $\overline{g^{\otimes k}}(x, \cdot)$ is the symmetrization –see (3.13)– of the function $g^{\otimes k}(x, \cdot)$ defined in (3.18). On the other hand, denote

$$\tilde{Z}_{i,k}(x) = Y_i \int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}_i^{\otimes k}(\mathrm{d}y),$$

then by the isometry formula (3.12):

$$\begin{aligned} \mathbb{E} \tilde{Z}_{i,k}(x) &= \mathbb{E} r(X) \int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}^{\otimes k}(\mathrm{d}y) = \int_{\mathbb{X}^k} f_k(y) \overline{g^{\otimes k}}(x, y) \varphi_\theta^{\otimes k}(y) \mathrm{d}y \\ &= \frac{1}{\varphi_\theta^{\otimes k}(x)} \int_{\mathbb{X}^k} f_k(x - h_k z) \overline{W^{\otimes k}}(z) \varphi_\theta^{\otimes k}(x - h_k z) \mathrm{d}z. \end{aligned}$$

Furthermore, by assumptions (3.3), (3.5), (3.10) and (3.11) on the model, we have for all $x, y \in \mathbb{X}^k$:

$$\begin{aligned} |f_k(x) \varphi_\theta^{\otimes k}(x) - f_k(y) \varphi_\theta^{\otimes k}(y)| &\leq \bar{\varphi}^k |f_k(x) - f_k(y)| + \bar{f} |\varphi_\theta^{\otimes k}(x) - \varphi_\theta^{\otimes k}(y)| \\ &\leq (\bar{\varphi}^k \gamma_2 + k \bar{f} \bar{\varphi}^{k-1} \gamma_1) |x - y| \\ &\leq 2k \bar{f} \bar{\varphi}^k \gamma_2 \gamma_1 |x - y|. \end{aligned}$$

Hence, letting $\omega_k = \int_{\mathbb{X}^k} |z| W^{\otimes k}(z) \mathrm{d}z$,

$$\begin{aligned} |\mathbb{E} Z_{i,k}(x) - f_k(x)| &\leq |\mathbb{E}(Z_{i,k}(x) - \tilde{Z}_{i,k}(x))| + |\mathbb{E} \tilde{Z}_{i,k}(x) - f_k(x)| \\ &\leq |\mathbb{E} Y \mathbf{1}_{|Y| > T_n} \int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}^{\otimes k}(\mathrm{d}y)| \\ &\quad + |\mathbb{E} \tilde{Z}_{i,k}(x) - f_k(x)| \\ &\leq \phi^{1/2} \left(\mathbb{E} \left(\int_{\Delta_k} g^{\otimes k}(x, y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \right)^{1/2} \\ &\quad + 2k \bar{f} \omega_k \frac{\bar{\varphi}^k \gamma_2 \gamma_1}{\underline{\varphi}^k} h_k. \end{aligned}$$

One last application of the isometry formula (3.12) to the first term on the right-hand side of above gives the Proposition. $\square$

With the help of notations (3.18), (3.19) and (3.20), define

$$R_{1k}^i = \mathbb{E} \left(\int_{\Delta_k^2} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) [\hat{X}_i^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)] \right)^2, \quad (3.22)$$

$$R_{2k}^i = \mathbb{E} \left(\int_{\Delta_k^2} g_i^{\otimes k}(x, y) [\hat{X}_i^{\otimes k}(\mathrm{d}y) - \tilde{X}_i^{\otimes k}(\mathrm{d}y)] \tilde{X}^{\otimes k}(\mathrm{d}x) \right)^2. \quad (3.23)$$

Lemma 3.5. *Let $i = 1, \dots, n$ and $k \leq \delta$ be fixed. Then, for $j = 1$ or 2 :*

$$R_{jk}^i \leq C^k \frac{(k!)^2}{nh_k^{dk}}.$$

Proof. The proofs for the bounds for R_{1k}^i and R_{2k}^i being similar, we only prove the one for R_{1k}^i . We have

$$R_{1k}^i = \mathbb{E} \mathbb{E} \left[\left(\int_{\Delta_k^2} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) [\hat{X}_i^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)] \right)^2 \middle| (X_l)_{l \leq n} \right],$$

Using the independence of X and $(X_l)_{l \leq n}$, we can apply Lemma 4.2 from Cadre and Truquet (2015), which entails

$$R_{1k}^i \leq \sum_{j=0}^{k-1} j! \binom{k}{j}^2 \bar{\varphi}^j \int_{\mathbb{X}^k} \mathbb{E} \left[V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) \right)^2 \right] \mathrm{d}x, \quad (3.24)$$

where $V_i = \|\varphi_{\hat{\theta}_i} - \varphi_{\theta}\|^2$. Now let $x \in \mathbb{X}^k$ and $j = 0, \dots, k-1$ be fixed. We have

$$\begin{aligned} \mathbb{E} V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) \right)^2 &\leq 2 \mathbb{E} V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) \tilde{X}_i^{\otimes k}(\mathrm{d}y) \right)^2 \\ &\quad + 2 \mathbb{E} V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) (\hat{X}_i^{\otimes k}(\mathrm{d}y) - \tilde{X}_i^{\otimes k}(\mathrm{d}y)) \right)^2. \end{aligned}$$

We proceed to bound the two terms on the right-hand side of above. As before, we apply Lemma 4.2 from Cadre and Truquet (2015), but conditionally on $(X_l)_{l \leq n, l \neq i}$. For notational simplicity, and since it does change the

result anymore, we do not specify the symmetrized version of the functions when using the isometry formula. By (3.12) and assumption (3.3), we then get

$$\begin{aligned} \mathbb{E}V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) \right)^2 &\leq 2 \sum_{l=0}^{k-1} l! \binom{k}{l}^2 \bar{\varphi}^l \mathbb{E}V_i^{2k-l-j} \int_{\mathbb{X}^k} g_i^{\otimes k}(x, y)^2 \mathrm{d}y \\ &\quad + 2k! \bar{\varphi}^k \mathbb{E}V_i^{k-j} \int_{\mathbb{X}^k} g_i^{\otimes k}(x, y)^2 \mathrm{d}y. \end{aligned}$$

Note that for all $m \geq 1$, by (3.4), we have

$$\begin{aligned} V_i^m &\leq (\bar{\varphi}^2 \lambda(\mathbb{X}))^{m-1} \left(\sup_{\mathbb{X}} |\varphi_{\hat{\theta}_i} - \varphi_{\theta}| \right)^2 \lambda(\mathbb{X}) \\ &\leq (\bar{\varphi}^2 \lambda(\mathbb{X}))^m |\hat{\theta}_i - \theta|^2, \end{aligned}$$

and

$$\int_{\mathbb{X}^k} g_i^{\otimes k}(x, y)^2 \mathrm{d}y \leq \frac{\bar{W}^k}{\underline{\varphi}^{2k} h_k^{dk}}.$$

Hence, since $l! \binom{k}{l} \leq k!$, we get with (3.7):

$$\mathbb{E}V_i^{k-j} \left(\int_{\Delta_k} g_i^{\otimes k}(x, y) \hat{X}_i^{\otimes k}(\mathrm{d}y) \right)^2 \leq 2k! \frac{\bar{W}^k}{\underline{\varphi}^{2k}} \frac{\kappa}{n h_k^{dk}} (\bar{\varphi}^2 \lambda(\mathbb{X}))^{k-j} (\bar{\varphi} + \bar{\varphi}^2 \lambda(\mathbb{X}))^k.$$

Finally, we deduce with similar arguments and inequality (3.24) that

$$\begin{aligned} R_{1k}^i &\leq 2(k!)^2 \frac{\bar{W}^k}{\underline{\varphi}^{2k}} \frac{\kappa}{n h_k^{dk}} (\bar{\varphi} + \bar{\varphi}^2 \lambda(\mathbb{X}))^{2k} \\ &\leq 2(k!)^2 \frac{4^k \bar{W}^k \bar{\varphi}^{4k}}{\underline{\varphi}^{2k}} \frac{\kappa}{n h_k^{dk}} \lambda(\mathbb{X})^{2k}, \end{aligned}$$

because both $\bar{\varphi}$ and $\lambda(\mathbb{X})$ are greater than 1. The Lemma is proved. $\square$

Lemma 3.6. *Let $\varepsilon > 0$ be fixed and set $T_n = (\ln n)^{1+\varepsilon}$. Then, for all $k \leq \delta$:*

$$\mathbb{E}(\hat{I}_{k, h_k}(X) - I_k(f_k))^2 \leq C^k (k!)^2 \left(\frac{(\ln n)^{2+2\varepsilon}}{n h_k^{dk}} + h_k^2 \right).$$

Proof. With the help of notations (3.18), (3.19), (3.20), we let

$$\begin{aligned} J_1 &= \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} g_i^{\otimes k}(x, y) \tilde{X}_i^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x), \\ J_2 &= \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} g^{\otimes k}(x, y) \tilde{X}_i^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x). \end{aligned}$$

Then, using notations of Lemma 3.5, by Jensen's Inequality

$$\mathbb{E}(\hat{I}_{k, h_k}(X) - J_1)^2 \leq 2 \frac{T_n^2}{n} \sum_{i=1}^n (R_{1k}^i + R_{2k}^i),$$

Hence, by Lemma 3.5

$$\mathbb{E}(\hat{I}_{k, h_k}(X) - J_1)^2 \leq T_n^2 C^k \frac{(k!)^2}{n h_k^{dk}}. \quad (3.25)$$

Moreover, sequentially conditioning on $(X_l)_{l \leq n}$, then on $(X_l)_{l \leq n, l \neq i}$, and using assumption (3.3), we find with two successive applications of the isometry formula (3.12) that

$$\mathbb{E}(J_1 - J_2)^2 \leq (k!)^2 T_n^2 \bar{\varphi}^{2k} \mathbb{E} \int_{\mathbb{X}^{2k}} (\overline{g_1^{\otimes k}(x, y)} - g^{\otimes k}(x, y))^2 \mathrm{d}x \mathrm{d}y.$$

Now let $x, y \in \mathbb{X}^k$ be fixed. We have

$$|g_1^{\otimes k}(x, y) - g^{\otimes k}(x, y)| \leq \frac{W_{h_k}^{\otimes k}(x - y)}{\varphi_{\theta}^{\otimes k}(x)} k \frac{\bar{\varphi}^{k-1}}{\underline{\varphi}^k} \sup_{\mathbb{X}} |\varphi_{\hat{\theta}_1} - \varphi_{\theta}|,$$

so that (3.4) and (3.7) give

$$\mathbb{E}(J_1 - J_2)^2 \leq C^k T_n^2 \frac{(k!)^2}{n h_k^{dk}}. \quad (3.26)$$

Finally, using notation (3.21), by the isometry formula (3.12), we have

$$\begin{aligned} \mathbb{E}(J_2 - I_k(f_k))^2 &= \mathbb{E} \left(\int_{\Delta_k} \left(\frac{1}{n} \sum_{i=1}^n Z_{i,k}(x) - f_k(x) \right) \tilde{X}^{\otimes k}(\mathrm{d}x) \right)^2 \\ &= k! \int_{\mathbb{X}^k} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n Z_{i,k}(x) - f_k(x) \right)^2 \varphi_{\theta}^{\otimes k}(x) \mathrm{d}x \\ &= k! \int_{\mathbb{X}^k} \left(\frac{1}{n} \mathrm{var}(Z_{1,k}(x)) + (\mathbb{E} Z_{1,k}(x) - f_k(x))^2 \right) \varphi_{\theta}^{\otimes k}(x) \mathrm{d}x. \end{aligned}$$

By Lemma 3.4, we thus get

$$\mathbb{E}(J_2 - I_k(f_k))^2 \leq T_n^2 \frac{(k!)^2 C^k}{n h_k^{dk}} + C^k (k!)^2 \frac{\phi^2}{h_k^{dk}} + C^k k! h_k^2.$$

Moreover, given that (3.8) gives $\phi \leq e^{-uT_n} \mathbb{E}Y^2 e^{u|Y|}$. Consequently,

$$\mathbb{E}(J_2 - I_k(f_k))^2 \leq C^k (k!)^2 \left(\frac{T_n^2}{n h_k^{dk}} + \frac{e^{-2uT_n}}{h_k^{dk}} + h_k^2 \right).$$

Finally, combining inequalities (3.25), (3.26) and above, we deduce that with the choice $T_n = (\ln n)^{1+\varepsilon}$:

$$\mathbb{E}(\hat{I}_{k,h_k}(X) - I_k(f_k))^2 \leq C^k (k!)^2 \left(\frac{(\ln n)^{2+2\varepsilon}}{n h_k^{dk}} + h_k^2 \right),$$

hence the Lemma. $\square$

3.5.2 Proof of Theorem 3.1

According to Jensen Inequality and Lemma 3.6, we have by (3.15) and (3.9):

$$\begin{aligned} \mathbb{E}(\hat{r}_\delta(X) - r(X))^2 &= \mathbb{E} \left(\bar{Y}_n - \mathbb{E}Y + \sum_{k=1}^{\delta} \frac{1}{k!} (\hat{I}_{k,h_k}(X) - I_k(f_k)) \right)^2 \\ &\leq \frac{2\text{var}(Y)}{n} + \delta \sum_{k=1}^{\delta} C^k \left(\frac{(\ln n)^{2+2\varepsilon}}{n h_k^{dk}} + h_k^2 \right). \end{aligned}$$

Setting $h_k = ((\ln n)^{2+\varepsilon} n^{-1})^{1/(2+dk)}$, we deduce that, since $\text{var}(Y) \leq M$:

$$\mathbb{E}(\hat{r}_\delta(X) - r(X))^2 \leq \frac{2M}{n} + C \left(\frac{(\ln n)^{2+2\varepsilon}}{n} \right)^{2/(2+d\delta)},$$

hence the theorem. $\square$

3.6 Proof of Theorem 3.2

In this section we assume for simplicity that $\mathbb{X}$ contains the hypercube $\mathbb{X}_0 = [0, 1]^d$.

3.6.1 Technical results

We introduce the set $\mathcal{F} = \mathcal{F}_\delta(\gamma_2, \bar{f})$ of functions $f : \mathbb{X}^\delta \rightarrow \mathbb{R}$ in $\mathbb{L}_{\text{sym}}^2(\lambda^{\otimes \delta})$ for which conditions (3.10) and (3.11) hold, and let $\mathcal{R} = \mathcal{R}_\delta(\gamma_2, \bar{f})$ be the class of functions $r_f : \mathcal{S} \rightarrow \mathbb{R}$ with $f \in \mathcal{F}$ such that

$$r_f(\cdot) = \frac{1}{\delta!} \int_{\Delta_\delta} f d(\cdot - \lambda)^{\otimes \delta}. \quad (3.27)$$

Letting P the distribution of the Poisson point process on $\mathbb{X}$ with unit intensity, we may define the following distance D on $\mathcal{R}$ by

$$D(r_{f_0}, r_{f_1}) = \|r_{f_0} - r_{f_1}\|_{\mathbb{L}^2(P)}. \quad (3.28)$$

Whenever $\mathbb{P} \in \mathcal{P}$, we can associate the regression function r . To stress the dependency on r , we now write $\mathbb{P}_r$ instead of $\mathbb{P}$. Now let $N > 0$. We define the following three conditions for any sequence of size $N + 1$ of functions $r^{(0)}, \dots, r^{(N)}$ from $\mathcal{S}$ to $\mathbb{R}$:

- R1. $r^{(j)} \in \mathcal{R}$, for $j = 0, \dots, N$;
- R2. $D(r^{(i)}, r^{(j)}) \geq 2n^{-1/(2+d\delta)}$, for $0 \leq i < j \leq N$;
- R3. $\frac{1}{N} \sum_{j=1}^N \mathcal{K}(\mathbb{P}_{r^{(j)}}^{\otimes n}, \mathbb{P}_{r^{(0)}}^{\otimes n}) \leq \alpha \log N$ for some $0 < \alpha < 1/8$ where $\mathcal{K}$ is the Kullback-Leibler divergence (see e.g. Tsybakov, 2009).

Lemma 3.7. *Introduce $f_0 \equiv 0, f_1, \dots, f_N$, $N + 1$ functions from $\mathbb{X}^\delta$ to $\mathbb{R}$ such that*

- F1. $f_j \in \mathcal{F}$;
- F2. $\|f_i - f_j\| \geq 2n^{-1/(2+d\delta)}$;
- F3. $\frac{1}{N} \sum_{i=1}^N \frac{n}{2} \|f_i\|^2 \leq \alpha \log N$ for some $0 < \alpha < 1/8$.

Then, the sequence of functions $r_{f_0}, \dots, r_{f_N}$ defined by (3.27) verify conditions R1, R2 and R3.

Proof. First of all, remark that since $f_j \in \mathcal{F}$ for $j = 0, \dots, N$, by definition of the r_{f_j} 's and the set $\mathcal{R}$, we have $r_{f_j} \in \mathcal{R}$. Hence condition R1 is satisfied by the r_{f_j} 's. Now remark that the Itô isometry (3.12) gives for any $0 \leq i, j \leq N$

$$D(r_{f_j}, r_{f_i}) = \|f_j - f_i\|.$$

This ensures that the r_{f_j} 's satisfy condition R2. Finally, for all $j = 0, \dots, N$

$$\begin{aligned} \mathcal{K}(\mathbb{P}_{r_{f_j}}^{\otimes n}, \mathbb{P}_{r_{f_0}}^{\otimes n}) &= n\mathcal{K}(\mathbb{P}_{r_{f_j}}, \mathbb{P}_{r_{f_0}}) \\ &= n\mathbb{E}_{r_{f_0}} \left(\log \frac{d\mathbb{P}_{r_{f_0}}}{d\mathbb{P}_{r_{f_j}}}(X, Y) \right) \\ &= n\mathbb{E}_{r_{f_0}} \mathbb{E}_{r_{f_0}} \left(\log \frac{d\mathbb{P}_{r_{f_0}}}{d\mathbb{P}_{r_{f_j}}}(X, Y) \mid X \right), \end{aligned}$$

where $\mathbb{E}_{r_{f_0}}$ is the expectation under $\mathbb{P}_{r_{f_0}}$. Denote by p the density of the $\mathcal{N}(0, 1)$. Then, since $f_0 \equiv 0$

$$\mathbb{E}_{r_{f_0}} \left(\log \frac{d\mathbb{P}_{r_{f_0}}}{d\mathbb{P}_{r_{f_j}}}(X, Y) \mid X \right) = \int_{\mathbb{R}} \log \left(\frac{p(u)}{p(u - r_{f_j}(X))} \right) p(u) du.$$

Simple calculus then give

$$\mathbb{E}_{r_{f_0}} \left(\log \frac{d\mathbb{P}_{r_{f_0}}}{d\mathbb{P}_{r_{f_j}}}(X, Y) \mid X \right) \leq \frac{1}{2} (r_{f_j}(X))^2.$$

Thus, by the Itô Isometry,

$$\mathcal{K}(\mathbb{P}_{r_{f_j}}^{\otimes n}, \mathbb{P}_{r_{f_0}}^{\otimes n}) \leq \frac{n}{2} \|f_j\|^2,$$

hence the lemma. $\square$

3.6.2 Proof of Theorem 3.2

Let $\mathcal{P}_0$ be the subset of distributions $\mathbb{P}_r$ of (X, Y) in $\mathcal{P}$ for which X is a Poisson point process with unit intensity (recall that the unit function lies in $\{\varphi_\theta\}_{\theta \in \Theta}$, see Section 3.2) and such that

$$Y = r(X) + \varepsilon, \tag{3.29}$$

where $r \in \mathcal{R}$ and ε is independent from X with distribution $\mathcal{N}(0, 1)$. Since $\mathcal{P}_0 \subset \mathcal{P}$, we have

$$\inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} R_n(\tilde{r}, r) \leq \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}} R_n(\tilde{r}, r).$$

As a result, in order to prove Theorem 3.2, we need only to prove that

$$\liminf_{n \rightarrow +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} R_n(\tilde{r}, r) > 0,$$

which accordingly to (3.28) may be written

$$\liminf_{n \rightarrow +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} \mathbb{E}_r^n D^2(\tilde{r}, r) > 0, \quad (3.30)$$

where $\mathbb{E}_r^n$ denotes expectation with respect to $\mathbb{P}_r^{\otimes n}$. Then, according to Lemma 3.7 and Theorem 2.5 page 99 in the book by Tsybakov (2009), in order to prove (3.30), we need only to prove the existence of a sequence of functions satisfying conditions F1, F2 and F3 defined in the previous subsection. To this end, we let $\psi \in \mathbb{L}_{\text{sym}}^2(\lambda^{\otimes \delta})$ be a nonzero function such that $\text{Supp}(\psi) = \mathbb{X}_0^\delta$ and, for all $x, y \in \mathbb{X}_0^\delta$:

$$|\psi(y) - \psi(x)| \leq \frac{\gamma_2}{2} |y - x| \text{ and } |\psi(x)| \leq \bar{f}. \quad (3.31)$$

Let $Q = \lfloor c_0 n^{d\delta/(2+d\delta)} \rfloor \geq 8$ where $c_0 > 0$ and $\lfloor \cdot \rfloor$ is the integer part and let $a_n = (1/Q)^{1/(d\delta)}$. One may easily prove that there exists $t_1, \dots, t_Q$ in $\mathbb{X}_0^\delta$ such that the functions

$$\psi_q(\cdot) = a_n \psi\left(\frac{\cdot - t_q}{a_n}\right), \text{ for } q = 1, \dots, Q,$$

verify the following assumptions

- (i) $\text{Supp}(\psi_q) \subset \mathbb{X}_0^\delta$, for $q = 1, \dots, Q$;
- (ii) $\text{Supp}(\psi_q) \cap \text{Supp}(\psi_{q'}) = \emptyset$, for $q \neq q'$;
- (iii) $\lambda^{\otimes \delta}(\text{Supp}(\psi_q)) = Q^{-1}$.

Now let, for all $\omega \in \{0, 1\}^Q$:

$$f_\omega(\cdot) = \sum_{q=1}^Q \omega_q \psi_q(\cdot),$$

According to the Varshamov-Gilbert Lemma (see the book by Tsybakov, 2009, Lemma 2.8 page 104), there exists a subset $\Omega = \{\omega^{(0)}, \dots, \omega^{(N)}\}$ of $\{0, 1\}^Q$ such that $\omega^{(0)} = (0, 0, \dots)$, $N \geq 2^{Q/8}$, and for all $j \neq k$:

$$\sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \geq \frac{Q}{8}.$$

Now fix $0 < \alpha < 1/8$ and set

$$c_0 = (4\|\psi\|^2 \alpha^{-1})^{d\delta/(2+d\delta)}.$$

We may now prove that functions $\{f_{\omega^{(j)}} : j = 0, \dots, N\}$ satisfy conditions F1, F2 and F3. First of all, let $j \neq k$ be fixed and remark that,

$$\begin{aligned} \|f_{\omega^{(j)}} - f_{\omega^{(k)}}\|^2 &\leq \int_{\mathbb{X}_0^\delta} (f_{\omega^{(j)}}(x) - f_{\omega^{(k)}}(x))^2 dx \\ &= \int_{\mathbb{X}_0^\delta} \left(\sum_{q=1}^Q (\omega_q^{(j)} - \omega_q^{(k)}) \psi_q(x) \right)^2 dx \\ &= \sum_{q=1}^Q \int_{\text{Supp}(\psi_q)} (\omega_q^{(j)} - \omega_q^{(k)})^2 a_n^2 \psi^2 \left(\frac{x - t_q}{a_n} \right) dx, \\ &= \frac{a_n^2}{Q} \sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \int_{\mathbb{X}_0^\delta} \psi^2(x) dx. \end{aligned}$$

Furthermore, by definition of the set Ω , we have

$$\frac{Q}{8} \leq \sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \leq Q,$$

so that

$$\frac{\|\psi\|^2}{8} a_n^2 \leq \|f_{\omega^{(j)}} - f_{\omega^{(k)}}\|^2 \leq \|\psi\|^2 a_n^2. \quad (3.32)$$

Now let $0 \leq j \leq N$. Since $\psi \in \mathbb{L}_{\text{sym}}^2(\lambda^{\otimes \delta})$ it is clear that $f_{\omega(j)}$ inherits that property. Then, using the first part of assumption (3.31) on ψ and assumption (ii) on the ψ_q 's, one may easily prove that $f_{\omega(j)}$ is a Lipschitz function with constant γ_2 . Finally, using the second part of assumption (3.31) on ψ and assumption (ii) on the ψ_q 's, we get that for all $x \in \mathbb{X}^k$, $|f_{\omega(j)}(x)| \leq a_n \bar{f}$ with $a_n \leq 1$ as soon as $n \geq c_0^{-(2+d\delta)/(d\delta)}$. We conclude that $f_{\omega(j)} \in \mathcal{F}$ so that condition F1 is satisfied by $f_{\omega(j)}$. Now, taking $k = 0$ in (3.32) gives

$$\frac{1}{N} \sum_{j=1}^N \frac{n}{2} \|f_{\omega(j)}\|^2 \leq \frac{n\|\psi\|^2}{2} a_n^2 \leq \frac{\|\psi\|^2}{2} c_0^{-(2+d\delta)/(d\delta)} Q.$$

Since $N \geq 2^{Q/8}$ and $c_0 \geq (4\|\psi\|^2 \alpha^{-1})^{d\delta/(2+d\delta)}$, we get

$$\frac{1}{N} \sum_{j=1}^N \frac{n}{2} \|f_{\omega(j)}\|^2 \leq 4\|\psi\|^2 c_0^{-(2+d\delta)/(d\delta)} \log N \leq \alpha \log N,$$

so that F3 is satisfied. Finally, according to (3.32), we have

$$\|f_{\omega(j)} - f_{\omega(k)}\| \geq \frac{\|\psi\|}{2\sqrt{2}} a_n = \frac{\|\psi\|}{2\sqrt{2}} (2c_0)^{-1/(d\delta)} n^{-1/(2+d\delta)}.$$

We can now conclude that conditions F1, F2 and F3 are satisfied so that according to Theorem 2.5 page 99 from the book by Tsybakov (2009) and Lemma 3.7, (3.30) is verified. Theorem follows. $\square$

3.7 Proof of Theorem 3.3

In this section, we assume without loss of generality that the constants $\bar{\varphi}$, γ_1 , γ_2 , $\bar{f}$, $\lambda(\mathbb{X})$ and $\bar{W}$ are greater than 1 and that $\underline{\varphi}$ and m are smaller than 1. Moreover, C denotes a positive number that only depends on the parameters of the model, i.e. $m, u, \underline{\varphi}, \bar{\varphi}, \delta, \gamma_1, \gamma_2, \bar{f}, \theta, \kappa, \lambda(\mathbb{X}), M$ and W , and whose value may change from line to line.

We let $\mathbb{P} \in \mathcal{P}(m)$, and for simplicity, we denote $\mathbb{E} = \mathbb{E}^{2n}$ the expectation with respect to $\mathbb{P}^{\otimes 2n}$. Recall that $(X, Y), (X_1, Y_1), \dots, (X_{2n}, Y_{2n})$ are i.i.d. with distribution $\mathbb{P}$.

3.7.1 Technical results

For $k \geq 1$, denote for all $x \in \mathbb{X}^k$:

$$g_k(x) = \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y) (X_i - \varphi_\theta \cdot \lambda)^{\otimes k}(\mathrm{d}y). \quad (3.33)$$

We also let

$$S_k = \frac{1}{n} \sum_{i=1}^n |Y_i| (X_i(\mathbb{X}) + \bar{\varphi} \lambda(\mathbb{X}))^k. \quad (3.34)$$

Lemma 3.8. *We have, for all k :*

$$\|\hat{g}_k - g_k\| \leq C^k S_k \frac{|\hat{\theta} - \theta|}{b_k^{dk/2}}.$$

Proof. Let $\hat{\mu}$ the signed measure

$$\begin{aligned} \hat{\mu}(\mathrm{d}y) &= \frac{1}{n} \sum_{i=1}^n Y_i \left[(X_i - \varphi_{\hat{\theta}} \cdot \lambda)^{\otimes k}(\mathrm{d}y) - (X_i - \varphi_\theta \cdot \lambda)^{\otimes k}(\mathrm{d}y) \right] \\ &= \frac{1}{n} \sum_{i=1}^n Y_i \sum_{l=0}^{k-1} \binom{k}{l} (-1)^{k-l} \prod_{j=1}^l X_i(\mathrm{d}y_j) \left[\prod_{j=l+1}^k \varphi_{\hat{\theta}}(y_j) \mathrm{d}y_j - \prod_{j=l+1}^k \varphi_\theta(y_j) \mathrm{d}y_j \right]. \end{aligned}$$

Then,

$$\begin{aligned} \|\hat{g}_k - g_k\|^2 &= \int_{\mathbb{X}^k} \left(\int_{\Delta_k} W_{b_k}^{\otimes k}(x - y) \hat{\mu}(\mathrm{d}y) \right)^2 \mathrm{d}x \\ &= \int_{\Delta_k^2} \left[\int_{\mathbb{X}^k} W_{b_k}^{\otimes k}(x - y^{(1)}) W_{b_k}^{\otimes k}(x - y^{(2)}) \mathrm{d}x \right] \hat{\mu}(\mathrm{d}y^{(1)}) \hat{\mu}(\mathrm{d}y^{(2)}). \end{aligned}$$

But,

$$\int_{\mathbb{X}^k} W_{b_k}^{\otimes k}(x - y^{(1)}) W_{b_k}^{\otimes k}(x - y^{(2)}) \mathrm{d}x \leq \frac{\overline{W}^k}{b_k^{dk}}.$$

Furthermore, one gets by induction that for $l = 0, \dots, k-1$:

$$\left| \prod_{j=l+1}^k \varphi_{\hat{\theta}}(y_j) - \prod_{j=l+1}^k \varphi_\theta(y_j) \right| \leq \bar{\varphi}^{k-l-1} \sum_{j=l+1}^k |\varphi_{\hat{\theta}}(y_j) - \varphi_\theta(y_j)|,$$

and by assumptions (3.3) and (3.4),

$$\begin{aligned} \int_{\mathbb{X}^{k-l}} \sum_{j=l+1}^k |\varphi_{\hat{\theta}}(y_j) - \varphi_{\theta}(y_j)| dy_{l+1} \dots dy_k &\leq (k-l) \lambda(\mathbb{X})^{k-l-1} \int_{\mathbb{X}} |\varphi_{\hat{\theta}} - \varphi_{\theta}| d\lambda \\ &\leq (k-l) \lambda(\mathbb{X})^{k-l} \bar{\varphi} |\hat{\theta} - \theta|. \end{aligned}$$

Putting all pieces together, we get

$$\|\hat{g}_k - g_k\|^2 \leq \frac{\bar{W}^k}{b_k^{dk}} \left(\frac{1}{n} \sum_{i=1}^n |Y_i| \sum_{l=0}^{k-1} k \binom{k-1}{l} X_i(\mathbb{X})^l \bar{\varphi}^{k-l-1} \lambda(\mathbb{X})^{k-l} |\hat{\theta} - \theta| \right)^2.$$

We can then conclude

$$\|\hat{g}_k - g_k\|^2 \leq \frac{\bar{W}^k k^2 \lambda(\mathbb{X})^2}{b_k^{dk}} |\hat{\theta} - \theta|^2 \left(\frac{1}{n} \sum_{i=1}^n |Y_i| (X_i(\mathbb{X}) + \bar{\varphi} \lambda(\mathbb{X}))^{k-1} \right)^2.$$

Lemma follows, since $\bar{\varphi} \lambda(\mathbb{X}) \geq 1$. $\square$

Denote for all $i, j \geq 0$

$$s_{i,j} = \mathbb{E}|Y|^i (X(\mathbb{X}) + \bar{\varphi} \lambda(\mathbb{X}))^j. \quad (3.35)$$

Moreover, $(V_n)_n$ is a sequence of real numbers, bigger than 1 and tending to infinity, to be tuned latter.

Lemma 3.9. *Let $l \geq 0$ be fixed. Then for all $x \in \mathbb{X}^k$,*

$$\mathbb{E} \left| Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(dy) \right|^l \leq s_{l,lk} \frac{C^{lk}}{b_k^{ldk}}.$$

Moreover,

$$\mathbb{E} \left(Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(dy) \right)^2 \leq C^k \sqrt{s_{4,4k}} \left(k! \frac{V_n^2}{b_k^{dk}} + \frac{e^{-uV_n/2}}{b_k^{2dk}} \right).$$

Proof. First observe that

$$\begin{aligned} \left| \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(dy) \right| &\leq \frac{\bar{W}^k}{b_k^{dk}} \sum_{l=0}^k \binom{k}{l} \bar{\varphi}^{k-l} \lambda(\mathbb{X})^{k-l} X(\mathbb{X})^l \\ &\leq \frac{\bar{W}^k}{b_k^{dk}} (X(\mathbb{X}) + \bar{\varphi} \lambda(\mathbb{X}))^k. \end{aligned}$$

Hence,

$$\mathbb{E} \left| Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right|^l \leq \frac{\overline{W}^{lk} s_{l,lk}}{b_k^{ldk}}$$

Regarding the second inequality, we observe that

$$\begin{aligned} \mathbb{E} \left(Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 &\leq V_n^2 \mathbb{E} \left(\int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \\ &\quad + \mathbb{E} \left(Y \mathbf{1}_{|Y| > V_n} \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2. \end{aligned}$$

By (3.12),

$$\mathbb{E} \left(\int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \leq k! \frac{C^k}{b_k^{dk}}.$$

Moreover, by the Cauchy-Schwarz Inequality, (3.8) and above, we get:

$$\left[\mathbb{E} \left(Y \mathbf{1}_{|Y| > V_n} \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \right]^2 \leq C^k s_{4,4k} \frac{e^{-uV_n}}{b_k^{4dk}}.$$

Putting all pieces together gives the result. $\square$

Lemma 3.10. *Let $k > \delta$ be fixed. Then,*

$$\begin{aligned} \mathbb{P}(\hat{\delta} = k) &\leq C^k \frac{((2k)! + s_{1,k})^2}{n(\rho_k b_k^{dk/2})^2} + \frac{s_{2,2k}}{n((2k)!)^2} \\ &\quad + \frac{C^k}{\rho_k^8} s_{8,8k} \max \left(\frac{n}{(nb_k^{dk})^8}, \frac{(k!)^4 V_n^8}{(nb_k^{dk})^4} + \frac{e^{-2uV_n}}{(nb_k^{2dk})^4} \right). \end{aligned}$$

Proof. Since $k > \delta$, we have with notation (3.33)

$$\begin{aligned} \mathbb{P}(\hat{\delta} = k) &\leq \mathbb{P}(\|\hat{g}_k\| > \rho_k) \\ &\leq \mathbb{P}(\|\hat{g}_k - g_k\| + \|g_k\| > \rho_k) \\ &\leq \mathbb{P}(2\|\hat{g}_k - g_k\| > \rho_k) + \mathbb{P}(2\|g_k\| > \rho_k). \end{aligned} \tag{3.36}$$

Regarding the first term on the right-hand side, by Lemma 3.8 and (3.7), we observe that:

$$\begin{aligned}
\mathbb{P}(2\|\hat{g}_k - g_k\| > \rho_k) &\leq \mathbb{P}\left(C^k S_k \frac{|\hat{\theta} - \theta|}{b_k^{dk/2}} \geq \rho_k\right) \\
&\leq \mathbb{P}\left(C^k ((2k)! + s_{1,k}) |\hat{\theta} - \theta| \geq \rho_k b_k^{dk/2}\right) \\
&\quad + \mathbb{P}\left(|S_k - s_{1,k}| \geq (2k)!\right) \\
&\leq C^k \frac{((2k)! + s_{1,k})^2}{n(\rho_k b_k^{dk/2})^2} + \frac{\mathbb{E}|S_k - s_{1,k}|^2}{((2k)!)^2}.
\end{aligned}$$

Regarding the latter term on the right-hand side of inequality (3.36), since $k > \delta$, we have

$$\mathbb{E}Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y) \tilde{X}^{\otimes k}(dy) = 0,$$

by (3.12). Thus, according to the Rosenthal Inequality (e.g. see Ibragimov and Sharakhmetov, 2002), we get

$$\begin{aligned}
\mathbb{E}\|g_k\|^8 &\leq \lambda(\mathbb{X})^{3k} \mathbb{E} \int_{\mathbb{X}^k} g_k^8 d\lambda \\
&\leq C^k \int_{\mathbb{X}^k} \max\left(\frac{1}{n^7} D_{k,8}(x), \frac{1}{n^4} (D_{k,2})^4(x)\right) dx,
\end{aligned}$$

where for all $j \geq 0$ and $x \in \mathbb{X}^k$:

$$D_{k,j}(x) = \mathbb{E}\left|Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y) \tilde{X}^{\otimes k}(dy)\right|^j.$$

Finally, according to Lemma 3.9 and since $s_{4,4k}^2 \leq s_{8,8k}$, we obtain

$$\mathbb{E}\|g_k\|^8 \leq C^k s_{8,8k} \max\left(\frac{n}{(nb_k^{dk})^8}, \frac{(k!)^4 V_n^8}{(nb_k^{dk})^4} + \frac{e^{-2uV_n}}{(nb_k^{2dk})^4}\right).$$

Lemma follows. □

From now on, denote

$$E_{k+1}^2 = 2(\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) \omega_{k+1} \lambda(\mathbb{X})^{k+2} \bar{f}, \quad (3.37)$$

where $\omega_{k+1} = \int_{\mathbb{X}^{k+1}} |x| K^{\otimes k+1}(x) dx$ and the other constants are defined in assumptions (3.10)-(3.11) on the model.

Lemma 3.11. *Let $k < \delta$ be fixed. Then,*

$$\begin{aligned} \|\mathbb{E}g_{k+1}\| &\geq m - E_{k+1}\sqrt{b_{k+1}}, \text{ and} \\ \mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 &\leq C\left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{e^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}}\right). \end{aligned}$$

Proof. By (3.12), for all $x \in \mathbb{X}^{k+1}$,

$$\begin{aligned} \mathbb{E}g_{k+1}(x) &= \int_{\mathbb{X}^{k+1}} f_{k+1}(y) \overline{W_{b_{k+1}}^{\otimes k+1}(x-y)} \varphi_{\theta}^{\otimes k+1}(y) dy \\ &= \int_{\mathbb{X}^{k+1}} (f_{k+1} \varphi_{\theta}^{\otimes k+1})(x - b_{k+1}z) \overline{W^{\otimes k+1}(z)} dz. \end{aligned}$$

By assumptions (3.3)-(3.10) on the model, we have

$$|(f_{k+1} \varphi_{\theta}^{\otimes k+1})(x - b_{k+1}z) - (f_{k+1} \varphi_{\theta}^{\otimes k+1})(x)| \leq (\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) b_{k+1} |z|.$$

Hence, since $\int_{\mathbb{X}} W d\lambda = 1$,

$$\mathbb{E}g_{k+1} \geq f_{k+1}(x) - (\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) b_{k+1} \omega_{k+1}.$$

Then,

$$\begin{aligned} \|\mathbb{E}g_{k+1}\|^2 &\geq \int_{\mathbb{X}^{k+1}} (f_{k+1}(x) - (\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) b_{k+1} \omega_{k+1})^2 dx \\ &\geq \|f_{k+1}\|^2 - 2(\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) b_{k+1} \omega_{k+1} \int_{\mathbb{X}^{k+1}} f_{k+1} d\lambda \\ &\geq \|f_{k+1}\|^2 - 2(\bar{\varphi}^{k+1} \gamma_2 + k \bar{f} \bar{\varphi}^k) b_{k+1} \omega_{k+1} \lambda(\mathbb{X})^{k+1} \|f_{k+1}\|. \end{aligned}$$

First part of the lemma follows, since $\|f_{k+1}\| \geq m$ and $|f_{k+1}| \leq \bar{f}$. Moreover, the second part is straightforward from Lemma 3.9, since

$$\begin{aligned} \mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 &= \int_{\mathbb{X}^{k+1}} \text{var}(g_{k+1}(x)) dx \\ &= \frac{1}{n} \int_{\mathbb{X}^{k+1}} \text{var}\left(Y \int_{\Delta_{k+1}} W_{b_{k+1}}^{\otimes k+1}(x-y) \tilde{X}^{\otimes k+1}(dy)\right) dx \\ &\leq \frac{1}{n} \int_{\mathbb{X}^{k+1}} \mathbb{E}\left(Y \int_{\Delta_{k+1}} W_{b_{k+1}}^{\otimes k+1}(x-y) \tilde{X}^{\otimes k+1}(dy)\right)^2 dx. \end{aligned}$$

□

Lemma 3.12. *We have, for all $k < \delta$:*

$$\mathbb{P}(\hat{\delta} = k) \leq C \left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{e^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}} \right).$$

Proof. Since $\|f_{k+1}\| \neq 0$,

$$\begin{aligned} \mathbb{P}(\hat{\delta} = k) &\leq \mathbb{P}(\|\hat{g}_{k+1}\| \leq \rho_{k+1}) \\ &\leq \mathbb{P}(\|\hat{g}_{k+1} - g_{k+1}\| + \|g_{k+1} - \mathbb{E}g_{k+1}\| \geq \|\mathbb{E}g_{k+1}\| - \rho_{k+1}) \\ &\leq \mathbb{P}(\|\hat{g}_{k+1} - g_{k+1}\| \geq \frac{\|\mathbb{E}g_{k+1}\| - \rho_{k+1}}{2}) \\ &\quad + \mathbb{P}(\|g_{k+1} - \mathbb{E}g_{k+1}\| \geq \frac{\|\mathbb{E}g_{k+1}\| - \rho_{k+1}}{2}). \end{aligned} \quad (3.38)$$

According to Lemma 3.8 and using the lower bound from Lemma 3.11, we find that the first term on the right-hand side of (3.38), denoted by p_1 , is upper bounded by

$$\mathbb{P}\left(C^k S_{k+1} \frac{|\hat{\theta} - \theta|}{b_{k+1}^{d(k+1)/2}} \geq \frac{1}{2}(m - E_{k+1}\sqrt{b_{k+1}} - \rho_{k+1})\right).$$

Since by assumption b_{k+1}, ρ_{k+1} tend to 0 as n tends to $+\infty$, for all $k < \delta$:

$$p_1 \leq \mathbb{P}(S_{k+1}|\hat{\theta} - \theta| \geq b_{k+1}^{d(k+1)/2} I),$$

for some $I > 0$ that does not depend on n and $k < \delta$. Thus, by the Markov Inequality,

$$\begin{aligned} p_1 &\leq \mathbb{P}((1 + s_{1,k+1})|\hat{\theta} - \theta| \geq b_{k+1}^{d(k+1)/2} I) + \mathbb{P}(|S_{k+1} - s_{1,k+1}| \geq 1) \\ &\leq \frac{(1 + s_{1,k+1})^2 \mathbb{E}|\hat{\theta} - \theta|^2}{I^2 b_{k+1}^{d(k+1)}} + \mathbb{E}(S_{k+1} - s_{1,k+1})^2 \\ &\leq \frac{C}{nb_{k+1}^{d(k+1)}} + \frac{C}{n}, \end{aligned} \quad (3.39)$$

by assumption (3.7) on the model, and since $s_{2,2(k+1)} \leq C$. In a similar fashion, regarding the latter term on the right-hand side of (3.38), further denoted by p_2 , we obtain with Lemma 3.11:

$$p_2 \leq \frac{4}{I^2} \mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 \leq C \left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{e^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}} \right).$$

We conclude the proof combining (3.38), (3.39) and above, since $V_n \geq 1$. $\square$

3.7.2 Proof of Theorem 3.3

First observe that for all $k \geq 1$, $\hat{\delta}$ is independent from $\hat{r}_k$, so that

$$\begin{aligned} \mathbb{E}(\hat{r}_{\hat{\delta}}(X) - r(X))^2 &\leq \mathbb{E}(\hat{r}_{\hat{\delta}}(X) - r(X))^2 + \sum_{k < \delta} \mathbb{E}(\hat{r}_k(X) - r(X))^2 \mathbb{P}(\hat{\delta} = k) \\ &\quad + \sum_{k > \delta} \mathbb{E}(\hat{r}_k(X) - r(X))^2 \mathbb{P}(\hat{\delta} = k). \end{aligned} \quad (3.40)$$

By Lemma 3.6, it is clear that for $k < \delta$,

$$\mathbb{E}(\hat{r}_k(X) - r(X))^2 \leq C.$$

Moreover, following the arguments of the proof of Theorem 3.1 (see Section 3.5.2), we can prove that for all $k \geq \delta$ (recall that $C > 0$ does not depend on k nor n):

$$\mathbb{E}(\hat{r}_k(X) - r(X))^2 \leq C^k \left(\frac{T_n^2}{n} \right)^{2/(2+dk)}.$$

Thus, by Theorem 3.1, (3.40) and above,

$$\begin{aligned} \mathbb{E}(\hat{r}_{\hat{\delta}}(X) - r(X))^2 &\leq C \left(\frac{T_n^2}{n} \right)^{2/(2+d\delta)} + C \sum_{k < \delta} \mathbb{P}(\hat{\delta} = k) \\ &\quad + \sum_{k > \delta} C^k \mathbb{P}(\hat{\delta} = k). \end{aligned} \quad (3.41)$$

Let ρ_k and b_k be defined in Theorem 3.3, and let $V_n = 2(\ln n)^\xi / u$, where $\xi > 1$. Then by Lemma 3.12, we get for $k < \delta$:

$$\mathbb{P}(\hat{\delta} = k) \leq C \frac{(\ln n)^{2\xi}}{n^{1-\beta/2}}.$$

Provided $\varepsilon > d\delta$, $\xi \leq (2 + \varepsilon)/(2 + d\delta)$ and $\beta \leq 2d\delta/(2 + d\delta)$, we have

$$\sum_{k < \delta} \mathbb{P}(\hat{\delta} = k) \leq C \left(\frac{T_n^2}{n} \right)^{2/(2+d\delta)}. \quad (3.42)$$

Note that when $\beta < 1$, the previous condition on β holds if $d\delta \geq 2$. Our task now is to bound $\mathbb{P}(\hat{\delta} = k)$ when $k > \delta$. First, we observe that for all

$i, j \geq 0$:

$$\begin{aligned} s_{i,j} &\leq (\mathbb{E}Y^{2i})^{1/2} \left[\left(\mathbb{E}(X(\mathbb{X})^{2j}) \right)^{1/2} + \bar{\varphi}^j \lambda(\mathbb{X})^j \right] \\ &\leq (\mathbb{E}Y^{2i})^{1/2} \left[e^{\bar{\varphi} \lambda(\mathbb{X})} \sqrt{(2j)!} + \bar{\varphi}^j \lambda(\mathbb{X})^j \right] \end{aligned}$$

according to the Cauchy-Schwarz Inequality. Consequently, by Lemma 3.10, if $k > \delta$:

$$\begin{aligned} \mathbb{P}(\hat{\delta} = k) &\leq \frac{C^k}{((2k)!)^2 n^{\alpha+\beta/2}} + \frac{C^k \sqrt{(4k)!}}{n((2k)!)^2} \\ &\quad + \frac{C^k \sqrt{(16k)!}}{((2k)!)^{16}} \max \left(\frac{1}{n^{4\alpha+3}}, \frac{(k!)^4 V_n^8}{n^{4\alpha+2\beta}} + \frac{e^{-2uV_n}}{n^{4\alpha}} \right). \end{aligned}$$

Noticing that by the Stirling Formula,

$$\sum_{k \geq 1} C^k \left(\frac{\sqrt{(4k)!}}{((2k)!)^2} + \frac{(k!)^4 \sqrt{(16k)!}}{((2k)!)^{16}} \right) < \infty,$$

we deduce, whenever $2\alpha + \beta > 1/(2 + d\delta)$:

$$\sum_{k > \delta} \mathbb{P}(\hat{\delta} = k) \leq C \left(\frac{T_n^2}{n} \right)^{2/(2+d\delta)}.$$

Theorem is now a straightforward consequence of (3.41), (3.42) and above.

□

Kernel estimation of the intensity of Cox processes

Sommaire

4.1	Introduction	107
4.2	Estimation strategy and results	109
4.2.1	Estimation strategy	109
4.2.2	Results	111
4.2.3	Examples	113
4.3	Simulation study	116
4.3.1	Experimental design	116
4.3.2	Results	117
4.3.3	Cross-Validation	121
4.4	Application to real data	123
4.5	Proofs	125
4.5.1	Proof of Theorem 4.1	126
4.5.2	Proof of Proposition 4.2	131
4.5.3	Proof of Theorem 4.3	132
4.6	Auxiliary results	134

4.1 Introduction

Counting processes and in particular Cox processes have been used for many years to model a large variety of situations from neuroscience (see Bialek et al., 1991, Brette, 2008, and Krumin and Shoham, 2009) to seismic (see Ogata, 1988), financial (see Merton, 1976), insurance (see Asmussen and Albrecher, 2010) or biophysical data (see Kou et al., 2005). Recall that a Cox process $N = (N_t)_{t \in [0,1]}$ with random intensity $\lambda = (\lambda(t))_{t \in [0,1]}$ is a

counting process such that the conditional distribution of N given λ is a Poisson process with intensity λ . In all the previous situations one of the main problem can be summarized as the estimation of the intensity λ of the process (see Zhang and Kou, 2010).

Note that when Cox process data arise the intensity of the process is mainly not directly observed but covariates are observed instead. Returning to one of the previous example, in single-molecule experiments only the peaks inducing the counting process and an underlying process are observed (see Kou et al., 2005). Another example can be found in car insurance (see Asmussen and Albrecher, 2010) where the counting process models the occurrence of car crash that are subject to weather conditions. In these cases the counting process $N = (N_t)_{t \in [0,1]}$ that naturally raises is accompanied with a covariate Z such that the conditional law of N given Z is a Poisson process with intensity $\lambda(t, Z)$.

From a statistical point of view one of the major issue is to estimate λ using n independent copies $(N^1, Z^1), \dots, (N^n, Z^n)$ of (N, Z) . For example, in Hansen, et al. (2015) and Gaïffas and Guillaux (2012), the authors consider the case where Z takes values in some Hilbert space and use Lasso procedure to get oracle inequalities on the risk of their estimator.

In examples above, it is however more natural to suppose that Z is a co-process and that the frequency of the jumping times of N at time $t \in [0, 1]$ depends on the past of Z . The intensity would then write $\lambda(t, (Z_s)_{s \leq t})$. That is the case we propose to study.

When dealing with practical problems it is often unnecessary, or at least not strictly required for the modeling, to observe the full trajectory of the co-process. One can instead observe the values taken by the co-process at some well chosen random times that cover most of the information in the co-process. In this model the co-process is observed at a finite number of random times thereby circumventing the curse of dimensionality that would occur otherwise since the co-process Z takes its values in an infinite dimension space as seen in O'Sullivan (1993).

In this paper we treat this original point of view. Let us now detail the model more precisely. Let $N = (N_t)_{t \in [0,1]}$ be a counting process and $Z = (Z_t)_{t \in [0,1]}$ be a $\mathbb{R}^d$ -valued co-process. We assume that N admits a random intensity which depends on t and on the observations of Z at random times $S_1 < S_2 < \dots$

More precisely, given the σ -algebra $\mathcal{S}$ generated by these times, N is a Cox process with intensity

$$\lambda_S(t, Z) = \theta_S\left(t, \vec{Z}_S(t)\right), \quad (4.1)$$

where M is the counting process associated to $S = (S_1, S_2, \dots)$, for any function $z : [0, 1] \rightarrow \mathbb{R}$, $\vec{z}_S(t)$ denotes the projection $(z_{S_1}, \dots, z_{S_{M_t}}) \in \mathbb{R}^{dM_t}$ and $\theta_S(t, \cdot)$ is a function from $\mathbb{R}^{dM_t}$ into $\mathbb{R}_+$.

In the sequel we consider that given $\mathcal{S}$, $(N^1, Z^1), \dots, (N^n, Z^n)$ are independent and identically distributed (*i.i.d.*) copies of (N, Z) . The goal of this paper is to construct and study the statistical properties of a kernel-type estimator of λ_S using these data. Note that the dimension of our estimation problem, which depends on the counting process $(M_t)_{t \in [0, 1]}$, increases with t . This potentially leads to a deterioration of the accuracy of any estimation procedure as the time variable increases.

It is noticeable that, for two theoretical examples, the obtained rate of convergence is in-between the traditional finite-dimensional rate (see Györfi et al., 2006), and the rates obtained by Biau et al. (2010) in the infinite-dimensional setting. This is explained by the particularity of our model which is itself in-between the finite and infinite-dimensional settings.

The paper is organized as follows. Section 4.2 presents the estimator we propose and its asymptotic properties. In Section 4.3 we proceed to a simulation study. In particular, we propose a cross-validation procedure to construct data-driven bandwidths that can be used for practical purposes. Technical proofs of the asymptotic properties are postponed to the last sections.

4.2 Estimation strategy and results

4.2.1 Estimation strategy

Let $t \in [0, 1]$ and $z : [0, 1] \rightarrow \mathbb{R}$ be fixed. In this section we present the main ideas behind the construction of our estimator $\tilde{\lambda}_S(t, z)$ of $\lambda_S(t, z)$.

As an introduction to our methodology we consider the ideal case where we observe $\theta_S\left(t, \vec{Z}_S^k(t)\right)$ for all $k = 1, \dots, n$. Then our problem of estimation

can simply be viewed as a regression estimation problem where $\lambda_S(t, \cdot)$ is the regression function. In this context, the Nadaraya-Watson estimator writes

$$\hat{\lambda}_{NW}(t, z) = \frac{\sum_{k=1}^n \theta_S \left(t, \vec{Z}_S^k(t) \right) H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right)}{\sum_{\ell=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^\ell(t) \right)},$$

where H_η denotes the multivariate product kernel $\mathcal{H}_\eta^{\otimes dM_t}$ with $\mathcal{H}$ a kernel, that is $\mathcal{H} \in \mathbb{L}^1(\mathbb{R})$ such that $\int_{\mathbb{R}} \mathcal{H}(u) du = 1$, η is an $\mathcal{S}$ -measurable positive random variable (called a bandwidth) and $\mathcal{H}_\eta(\cdot) = \eta^{-1} \mathcal{H}(\eta^{-1} \cdot)$.

In practice $\theta_S \left(t, \vec{Z}_S^k(t) \right)$ can be estimated using the observations. Indeed, conditionally to $\mathcal{S}$ and the σ -algebra $\mathcal{Z}$ generated by the co-processes $(Z^1, \dots, Z^n)$, N^k is a non-homogeneous Poisson process with intensity function $t \mapsto \theta_S \left(t, \vec{Z}_S^k(t) \right)$, a natural estimator of this intensity is given by

$$\int_0^t K_h(t-u) dN_u^k = \sum_{i=1}^{N_t^k} K_h(t - T_i^k),$$

where $T_1^k, T_2^k, \dots$ denote the jumping times of the trajectory N^k , $K : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a kernel and h is a bandwidth. Denoting

$$\begin{aligned} \hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) &= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^{N_t^k} K_h(t - T_i^k) H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right), \\ \hat{f}_{S,h,\eta}(\vec{z}_S(t)) &= \frac{1}{n} \sum_{\ell=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^\ell(t) \right), \end{aligned}$$

we define the plug-in estimator by

$$\hat{\lambda}_{S,h,\eta}(t, z) = \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{\hat{f}_{S,h,\eta}(\vec{z}_S(t))}.$$

For the sake of stability (see Bickel, 1982) we consider a trimmed version of the previous estimator

$$\tilde{\lambda}_{S,h,\eta}(t, z) = \tilde{\theta}_{S,h,\eta}(t, \vec{z}_S(t)) = \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{\tilde{f}_{S,h,\eta}(\vec{z}_S(t))}, \quad (4.2)$$

where $\tilde{f}_{S,h,\eta}(\vec{z}_S(t)) = \max(\hat{f}_{S,h,\eta}(\vec{z}_S(t)), a_n)$ and $(a_n)_{n \in \mathbb{N}}$ is an $\mathcal{S}$ -measurable real-valued positive decreasing sequence defined by $a_n = (n\eta^{dM_t})^{-1/2}$.

4.2.2 Results

We are interested in the local behaviour of our estimator. We thus consider the pointwise mean squared error defined by

$$\text{MSE}(t, z) = \mathbb{E} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right)^2, \quad (4.3)$$

and we make local regularity assumptions on the model.

Remark that, $\mathbb{P}$ -a.s., t belongs to the random interval $I_t =]S_{M_t}, S_{M_t+1}[$, so that for any $u \in I_t$, the function $\theta_S(u, \cdot)$ maps $\mathbb{R}^{dM_t}$ into $\mathbb{R}_+$. In what follows, we only make local assumptions on the neighborhood I_t of t :

- (H1) For any $u \in I_t$, $\vec{Z}_S(u)$ admits a conditional density $f : \mathbb{R}^{dM_t} \rightarrow \mathbb{R}_+$ with respect to $\mathbb{P}(\cdot | S)$;
- (H2) $\theta_S : I_t \times \mathbb{R}^{dM_t} \rightarrow \mathbb{R}_+$ and f are two positive continuous functions;
- (H3) θ_S and f are twice continuously differentiable and there exists a positive constant Q such that for all $1 \leq k, \ell \leq dM_t$,

$$\forall y \in \mathbb{R}^{dM_t}, \left| \frac{\partial^2}{\partial x_k \partial x_\ell} f(y) \right| \leq Q^{M_t},$$

and for all $1 \leq k, \ell \leq dM_t + 1$,

$$\forall (u, y) \in I_t \times \mathbb{R}^{dM_t}, \left| \frac{\partial^2}{\partial x_k \partial x_\ell} (f(y)\theta_S(u, y)) \right| \leq Q^{M_t};$$

- (H4) There exists positive constants F_0, F_∞ and Θ such that $\|\theta_S\|_\infty < \Theta^{M_t}$ and $0 < F_0^{M_t} \leq f \leq F_\infty^{M_t} < \infty$;
- (H5) For any $\kappa > 0$, $\mathbb{E} e^{\kappa M_t} < +\infty$.

Remark. Assumptions (H1), (H2), (H3) and (H4) are very similar usual regularity assumptions for the estimation of the regression but made locally on the neighborhood I_t of t . Assumption (H5) comes from the particularity of the model (4.1). Indeed, since regularity assumptions have been made conditionally on $\mathcal{S}$, the good asymptotic properties of estimator (4.2) depend on the integrability of the process $(M_t)_{t \in [0,1]}$. The conditions defined

here are nonetheless usual and not very restrictive on the family of processes that can be considered.

We also make technical assumptions on the kernels:

- (**T1**) $\mathcal{H}$ is a kernel such that $\text{supp}\mathcal{H} = [-1, 1]$ and $\mathcal{H} \in \mathbb{L}_\infty([-1, 1])$ and K is a kernel such that $\text{supp}K = [0, 1]$ and $K \in \mathbb{L}^2([0, 1])$;
- (**T2**) $\mathcal{H}$ and K such that $\int_{\mathbb{R}} u\mathcal{H}(u)du = 0$ and $\int_{\mathbb{R}} uK(u)du = 0$.

We are now in position to state our main results. Define the pointwise conditional mean squared error by

$$\text{MSE}_S(t, z) = \mathbb{E} \left[\left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right)^2 \mid S \right].$$

Theorem 4.1. *Assume that **(H1)** to **(H5)**, **(T1)** and **(T2)** are satisfied. Set $h = \eta = n^{-\frac{1}{5+dM_t}}$ two $\mathcal{S}$ -measurable bandwidths. There exists a positive constant τ depending only on F_0 , F_∞ , Θ and Q such that*

$$\text{MSE}_S(t, z) \leq e^{\tau M_t} n^{-\frac{4}{5+dM_t}}.$$

Hence,

$$\text{MSE}(t, z) \leq \mathbb{E} \left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}} \right). \quad (4.4)$$

Theorem 4.1 gives the consistency of estimator (4.2). We can nevertheless get the consistency of our estimator under weaker assumptions as shown in the following proposition.

Proposition 4.2. *Assume that **(H1)**, **(H2)** and **(H4)** are satisfied. Let h and η be two $\mathcal{S}$ -measurable bandwidths such that $h \rightarrow 0$, $\eta \rightarrow 0$, and $nh\eta^{dM_t} \rightarrow +\infty$ $\mathbb{P}$ -a.s. as $n \rightarrow +\infty$, then*

$$\tilde{\lambda}_{S,h,\eta}(t, z) \xrightarrow{\mathbb{P}} \lambda_S(t, z).$$

Now, for any kernel $\mathcal{K}$, denote by $m_j(\mathcal{K}) = \int u^j \mathcal{K}(u)du$ its j -th moment. Theorem 4.3 shows the asymptotic normality of our estimator.

Theorem 4.3. Assume that (H1) to (H5), (T1) and (T2) are satisfied. Let h and η be two $\mathcal{S}$ -measurable bandwidths such that $h \rightarrow 0$, $\eta \rightarrow 0$, $nh^5\eta^{dM_t} = O(1)$, $nh\eta^{dM_t+4} = O(1)$ and $nh\eta^{dM_t} \rightarrow +\infty$ $\mathbb{P}$ -a.s. as $n \rightarrow +\infty$ then for any $z : [0, 1] \rightarrow \mathbb{R}$ such that $\theta_S(t, \vec{z}_S(t)) \neq 0$

$$(nh\eta^{dM_t})^{1/2}\sigma^{-1} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) + h^2 B_1 + \eta^2 B_2 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where

$$\begin{aligned} \sigma^2 &= f(\vec{z}_S(t))\theta_S(t, \vec{z}_S(t))\|K\|_2^2\|\mathcal{H}\|_2^{2dM_t}, \\ B_1 &= \frac{1}{2}m_2(K)\frac{1}{f(\vec{z}_S(t))}\partial_1^2(f(\vec{z}_S(t))\theta_S(t, \vec{z}_S(t))), \\ B_2 &= \frac{1}{2}m_2(\mathcal{H})\sum_{i=1}^{dM_t}\left(\frac{1}{f(\vec{z}_S(t))}\partial_{i+1}^2(f(\vec{z}_S(t))\theta_S(t, \vec{z}_S(t))) - \frac{\theta(t, \vec{z}_S(t))}{f(\vec{z}_S(t))}\partial_i^2 f(\vec{z}_S(t))\right). \end{aligned}$$

Remark. Note that taking the bandwidth of optimal order in Theorem 4.3, that is $h = \eta = n^{-\frac{1}{5+dM_t}}$ yields

$$(nh\eta^{dM_t})^{1/2}\sigma^{-1} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right) + \sigma^{-1}(B_1 + B_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

for then $(nh\eta^{dM_t})^{1/2}h^2 = (nh\eta^{dM_t})^{1/2}\eta^2 = 1$. This says that our estimator is asymptotically normal, with a non-zero asymptotic bias and variance.

4.2.3 Examples

Theorem 4.1 gives the tools to define optimal bandwidths in terms of point-wise asymptotic mean squared error. Assumptions must however be made on the process M to conclude on the convergence rate since the MSE depends on the quantity $\mathbb{E}\left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}}\right)$.

Assume that (H1) to (H4) are satisfied. Assume also that M is a renewal process with inter-arrival times distributed accordingly to a strictly increasing cumulative distribution function F . The behaviour of the considered expectation is linked to the local behaviour of F around 0. The two following examples give incentive on the performances of our estimator for F close to 0 around 0. In both situations, it is easily seen that (H5) is fulfilled.

Example 1. Let ε be a positive constant and assume that $F(x) = 0$ for any $x \in [0, \varepsilon]$. Then, for any $\kappa > 0$,

$$\mathbb{E}(e^{\kappa M_t}) = \sum_{k \geq 0} e^{\kappa k} \mathbb{P}(M_t = k).$$

Since $\mathbb{P}(M_t = k) = \mathbb{P}(S_k \leq t) - \mathbb{P}(S_{k-1} \leq t)$ by definition of M , we get:

$$\begin{aligned} \mathbb{E}(e^{\kappa M_t}) &= \mathbb{P}(S_0 \leq t) + \sum_{k \geq 1} (e^{\kappa k} - e^{\kappa(k-1)}) \mathbb{P}(S_k \leq t) \\ &\leq 1 + \sum_{k=1}^{\lfloor \frac{t}{\varepsilon} \rfloor} (e^{\kappa k} - e^{\kappa(k-1)}) \\ &\leq e^{\kappa \lfloor \frac{t}{\varepsilon} \rfloor}. \end{aligned}$$

Assumption **(H5)** follows. The same sketch of proof gives

$$\mathbb{E}\left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}}\right) \leq e^{\tau \lfloor \frac{t}{\varepsilon} \rfloor} n^{-\frac{4}{5+d \lfloor \frac{t}{\varepsilon} \rfloor}}.$$

So that (4.4) gives

$$\text{MSE}(t, z) \leq e^{\tau \lfloor \frac{t}{\varepsilon} \rfloor} n^{-\frac{4}{5+d \lfloor \frac{t}{\varepsilon} \rfloor}}.$$

This is the optimal rate of convergence for the nonparametric regression with a twice continuously differentiable regression function that maps $\mathbb{R}^{d \lfloor \frac{t}{\varepsilon} \rfloor + 1}$ into $\mathbb{R}$ (see Györfi et al., 2006). In some sense, $d \lfloor \frac{t}{\varepsilon} \rfloor + 1$ can be viewed as an upper bound of the “effective” dimension of the model.

Example 2. Let ε and α be two positive constants such that $\alpha > 1$ and assume that $F(x) \leq \exp\{-(\varepsilon x^{-1})^\alpha\}$, around 0. Then, in a similar fashion as for Example 1, for any $\kappa > 0$, we have

$$\mathbb{E}(e^{\kappa M_t}) = \mathbb{P}(S_0 \leq t) + \sum_{k \geq 1} (e^{\kappa k} - e^{\kappa(k-1)}) \mathbb{P}(S_k \leq t).$$

Now let $k^* = \left(\frac{t}{\varepsilon}\right)^{\frac{\alpha}{\alpha-1}} (2\kappa)^{\frac{1}{\alpha-1}}$, $\varepsilon_k = \varepsilon(2k\kappa)^{-\frac{1}{\alpha}}$ and

$$A_k = \{\forall \ell = 1, \dots, k; S_\ell - S_{\ell-1} \geq \varepsilon_k\},$$

so that, $k^* \varepsilon_{k^*} = t$, $\exp\{-(\varepsilon x^{-1})^\alpha\} = \exp(-2\kappa k)$ and

$$\begin{aligned} \mathbb{P}(S_k \leq t) &\leq \mathbb{P}(S_k \leq t; A_k) + \mathbb{P}(A_k^c) \\ &\leq \mathbb{P}(k\varepsilon_k \leq t; A_k) + kF(\varepsilon_k). \end{aligned}$$

We deduce that,

$$\begin{aligned} \mathbb{E}(e^{\kappa M_t}) &\leq 1 + \sum_{k=1}^{k^*} (e^{\kappa k} - e^{\kappa(k-1)}) + \sum_{k \geq 1} (e^{\kappa k} - e^{\kappa(k-1)}) kF(\varepsilon_k) \\ &\leq C e^{\kappa k^*}, \end{aligned}$$

where C is an absolute positive constant. Assumption [\(H5\)](#) follows. Letting $k_1^* = \left(\frac{t}{\varepsilon}\right)^{\frac{\alpha}{\alpha-1}} 3^{\frac{1}{\alpha-1}}$, $\varepsilon_k^1 = \varepsilon(3 \log(k))^{-\frac{1}{\alpha}}$ and

$$A_k^1 = \{\forall \ell = 1, \dots, k; S_\ell - S_{\ell-1} \geq \varepsilon_k^1\},$$

the same sketch of proof gives

$$\mathbb{E}\left(n^{-\frac{8}{5+dM_t}}\right) \leq n^{-\frac{8}{5+dk_1^*}},$$

so that combining above with Cauchy's inequality, we get

$$\mathbb{E}\left(e^{\tau M_t} n^{-\frac{4}{5+dM_t}}\right) \leq C e^{\tau k^*} n^{-\frac{4}{5+dk_1^*}}.$$

Finally, according to [\(4.4\)](#), we get

$$\text{MSE}(t, z) \leq C e^{\left(\frac{t}{\varepsilon}\right)^{\frac{\alpha}{\alpha-1}} (2\kappa)^{\frac{1}{\alpha-1}}} n^{-\frac{4}{5+d\left(\frac{t}{\varepsilon}\right)^{\frac{\alpha}{\alpha-1}} 3^{\frac{1}{\alpha-1}}}}.$$

Remark that if we formally take $\alpha = +\infty$, we get back to the situation of [Example 1](#) and the upper bounds coincide as the previous upper bound writes

$$\text{MSE}(t, z) \leq C e^{\tau \left(\frac{t}{\varepsilon}\right)} n^{-\frac{4}{5+d\frac{t}{\varepsilon}}}.$$

Note that the rate of convergence of the mean squared error is in-between the traditional finite-dimensional rate (see Györfi et al., [2006](#)), and the rates obtained by Biau et al. ([2010](#)) in the infinite-dimensional setting. This is explained by the particularity of our model which is itself in-between the finite and infinite-dimensional settings.

4.3 Simulation study

In this section we aim at studying the performances of our estimator from a practical point of view. To this end we study our estimator over n_{MC} replications of Monte Carlo simulations. The squared error, its mean (MSE) defined in (4.3), median, first and third empirical quartiles as well as the normalized root mean squared error (NRMSE) defined as follows

$$\text{NRMSE}(t, z) := \frac{\sqrt{\text{MSE}(t, z)}}{\lambda_S(t, z)}, \quad (4.5)$$

are used as indicators of the performances of our estimator and are calculated over a grid of n_t times t in $[0, 1]$.

In model (4.1) the intensity of the Cox process N depends on a co-process observed at random times which implies that the estimation problem lies in-between a classical multivariate problem and a functional one. To illustrate these properties we propose a specific form of the intensity $\lambda_S(t, Z)$ that depends on two tuning parameters ε and β , such that the smaller ε is, the greater the effective dimension is and the smaller β is, the closer the model is to the inhomogeneous Poisson model.

4.3.1 Experimental design

Let us consider the following experimental design. On the one hand the inter-arrival times of the counting process M have the same distribution function than $U + \varepsilon$ with $U \sim \mathcal{E}(1/\varepsilon)$ for $\varepsilon > 0$, putting ourselves in the situation of Example 1. Recall that, in this setting, the effective dimension of the estimation problem is upper bounded by $d\lfloor \frac{t}{\varepsilon} \rfloor + 1$. On the other hand

$$\theta_S(t, \vec{Z}_S(t)) = \lambda_0(t) \exp \left(\sin \left(\beta \sum_{i=1}^{M_t} Z_{S_i} \right) \right), \quad (4.6)$$

where $\lambda_0(t) = \frac{b}{a} \left(\frac{t}{a} \right)^{b-1}$ with $a, b, \beta > 0$. For simplicity, the co-process Z is simulated according to a univariate Brownian motion.

Note that the intensity function θ_S is a modified version of the proportional hazards intensity function that models the dependence of our counting process on the past of the co-process Z . For $\beta > 0$, $\theta_S(t, \vec{Z}_S(t))$ can be

viewed as a stochastic perturbation of the intensity of a Weibull process as presented in Figure 4.1. Remark that for $\beta \geq 0.3$ the replications look too irregular for a kernel estimator to capture their behaviour properly. It is due to the increasing impact of the covariates which we chose to be Brownian.

To simulate the data, we may use the Poisson nature of $(N_t)_{t \in [0,1]}$, that is, $(N_t)_{t \in [0,1]}$ is an inhomogeneous Poisson process conditionally on $\mathcal{Z}$ and $\mathcal{S}$. We can then simulate its jumping times by applying the inverse function of $\Lambda(\cdot) = \int_0^\cdot \theta_S(u, \vec{Z}_S(u)) du$ to the jumping times of a homogeneous Poisson process with intensity 1. In our case this inverse function writes

$$\Lambda^{-1}(u) = \Lambda_0^{-1} \left(\Lambda_0(S_{j_u}) + \frac{u - \Lambda(S_{j_u})}{\exp \left(\sin \left(\beta \sum_{i=1}^{M_t} Z_{S_i} \right) \right)} \right),$$

where j_u is such that $\Lambda(S_{j_u}) \leq u < \Lambda(S_{j_u+1})$ and $\Lambda_0(\cdot) = \int_0^\cdot \lambda_0(u) du$.

We finally take

$$\begin{aligned} \mathcal{H}(u) &= \left(\frac{1}{2} - \frac{5}{8}(3u^2 - 1) \right) \mathbf{1}_{|u| \leq 1}, \\ K(u) &= (30u^2 - 36u + 9) \mathbf{1}_{0 \leq u \leq 1}, \end{aligned}$$

so that assumptions **(T1)** and **(T2)** are fulfilled.

4.3.2 Results

In the following, according to the definition of Z , $d = 1$ and we take $a = 0.2$, $b = 3$, and $\varepsilon = 0.005$ in model (4.6).

Figure 4.2 represents the theoretical intensity (black line) versus the first and third empirical quartiles (gray lines) of 100 Monte Carlo replications of our estimator for $n = 500$ and $\beta = 0$ (top left), $\beta = 0.1$ (top right), $\beta = 0.3$ (bottom left) and $\beta = 0.5$ (bottom right). The case $\beta = 0$ is the one of the estimation of the intensity of a Weibull process. As β increases, the counting process N deviates from this simple case to a point where the signal is almost chaotic due to the influence of the co-process Z for $\beta = 0.5$ (see Figure 4.1). As expected our estimator is less accurate for high values of t (*i.e.* high dimensionality) and quickly varying objective function (*e.g.* $\beta = 0.5$). We also note an artifact for the estimation around zero. It is a

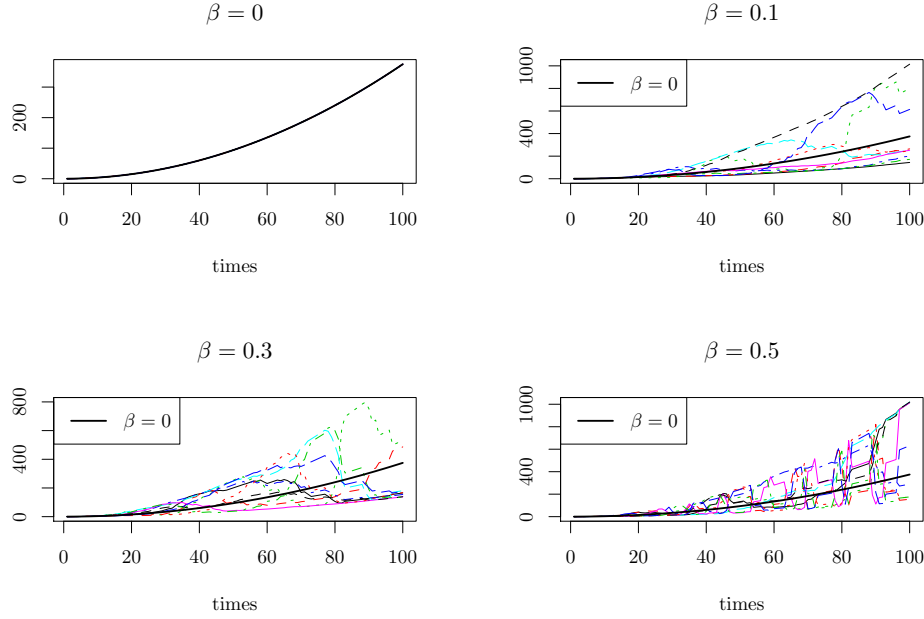


Figure 4.1: 10 replications of the studied intensity (4.6) for $a = 0.2$, $b = 3$, $\varepsilon = 0.005$, and $\beta = 0$ (top left), $\beta = 0.1$ (top right), $\beta = 0.3$ (bottom left) and $\beta = 0.5$ (bottom right). The bold line represents the corresponding intensity for a Weibull process ($\beta = 0$).

well known issue with kernel estimation on the edges of the support of the objective function.

In the following we fix $\beta = 0.1$.

Figure 4.3 represents the median (black line) and the first and third empirical quartiles (gray lines) of the squared error computed on 10,000 Monte Carlo replications of our estimator for $n = 500$ (Figure 4.3a) and $n = 10,000$ (Figure 4.3b). As expected, the results are far better for $n = 10,000$ where the third quartile does not exceed 14,000 compared to a maximum of 50,000 for $n = 500$. Remark that these maxima are always attained near to $t = 1$. That is explained by the fact that as t increases, the dimension of the estimation problem increases.

In Table 4.1 the results are obtained for 10,000 Monte Carlo replications of

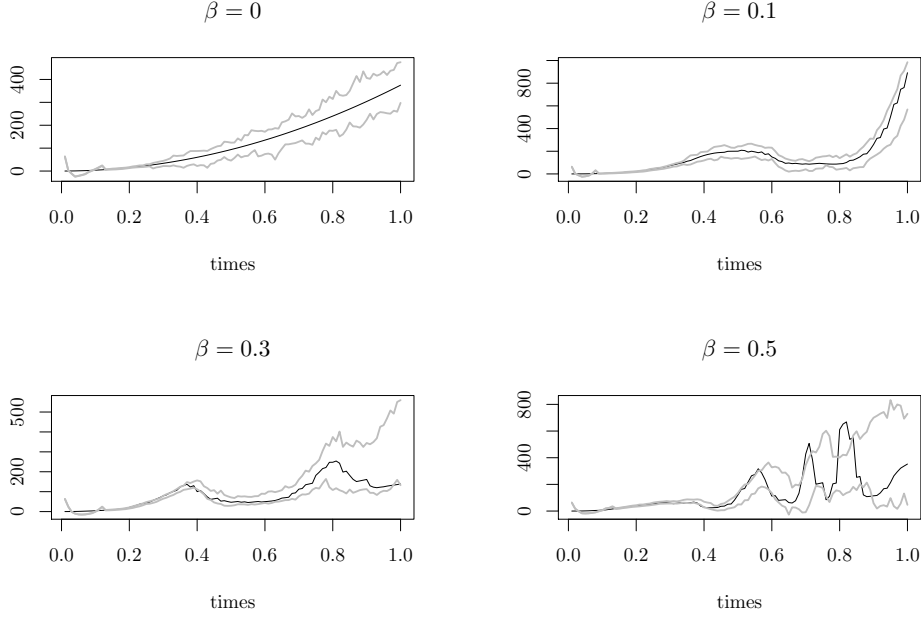


Figure 4.2: Objective intensity (black line) versus first and third (gray lines) empirical quartiles for 500 observations, $n_{MC} = 100$, $n_t = 100$ and $\beta = 0$ (top left), $\beta = 0.1$ (top right), $\beta = 0.3$ (bottom left) and $\beta = 0.5$ (bottom right).

our model for 3 different times and 5 different values of n . We observe that for fixed n , the performances of our estimator decrease as t increases. This is explained by the fact that the effective dimensionality of the estimation problem increases with t (see Example 1). The dimensionality of our estimation problem increases quickly towards 59 at time $t = 0.9$. This shows the difficulty of the estimation for small n . We observe nevertheless an increase in performance for bigger values of n . For $n = 10,000$ the NRMSE indicator stays below 0.2 after time $t = 0.3$. At this point we seem to have attained the asymptotic property for the MSE described in Section 4.2.2.

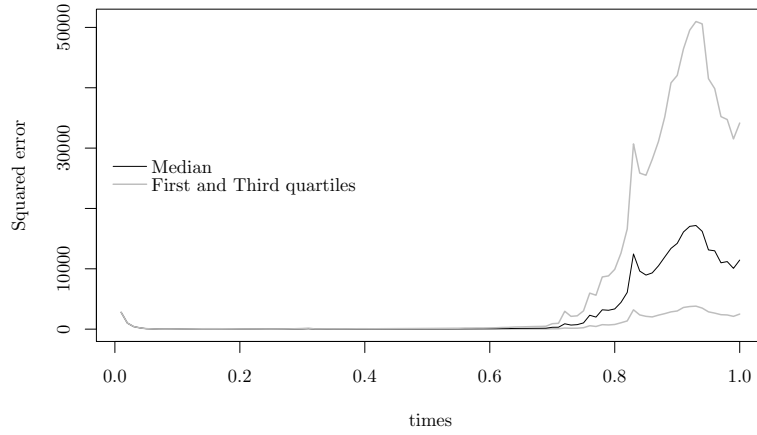
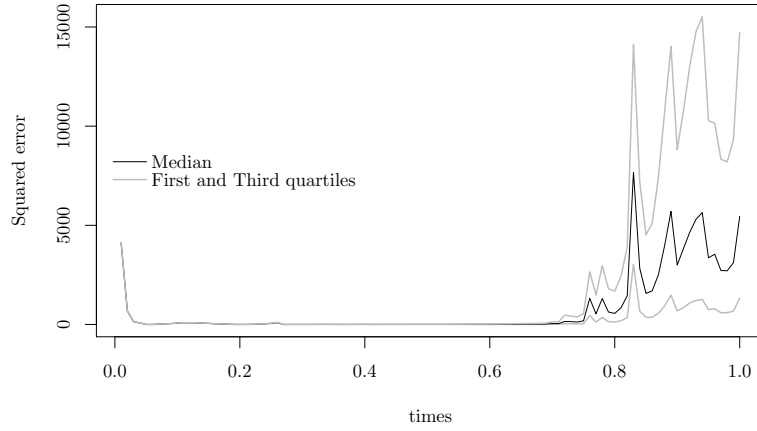
(a) $n = 500$ (b) $n = 10,000$

Figure 4.3: Median, first and third empirical quartiles of the squared error of our estimator for $n_{MC} = 10,000$, $n_t = 100$, $\beta = 0.1$ and (a) $n = 500$ and (b) $n = 10,000$.

n	t	M_t	$\lambda(t)$	Estimate	MSE	NRMSE
100	0.5	33	36.21	38.15	4.7E+03	1.89
	0.7	47	110.17	97.50	1.2E+05	3.14
	0.9	59	763.96	675.75	6.5E+06	3.34
250	0.5	33	36.21	38.26	3.4E+02	0.51
	0.7	47	110.17	92.28	1.3E+03	0.33
	0.9	59	763.96	717.24	3E+06	2.27
500	0.5	33	36.21	38.37	96	0.27
	0.7	47	110.17	91.51	9.4E+02	0.28
	0.9	59	763.96	649.71	4.8E+07	9.06
1,000	0.5	33	36.21	38.03	54	0.20
	0.7	47	110.17	92.05	7.1E+02	0.24
	0.9	59	763.96	732.71	7.2E+04	0.35
10,000	0.5	33	36.21	37.84	11	0.09
	0.7	47	110.17	92.39	3.8E+02	0.18
	0.9	59	763.96	743.70	2.9E+03	0.07

Table 4.1: Mean value of the estimator, mean squared error (4.3) and normalized root mean squared error (4.5) for $n_{MC} = 10,000$ and $\beta = 0.1$.

4.3.3 Cross-Validation

The practical use of our estimator relies on the calibration of two bandwidths. In kernel regression, choosing the bandwidth that minimizes the prediction error is usually the way to go and cross-validation is probably the simplest and most widely used method for estimating this error (see e.g., Friedman, Hastie and Tibshirani, 2001 and the references within). In a similar way, we construct a K -fold cross-validation estimate of the following integrated estimation error

$$\mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{dM_s}} \left(\tilde{\theta}_{S,h,\eta}(u, y) - \theta_S(u, y) \right)^2 f(y) dy du \right]. \quad (4.7)$$

Furthermore, for more simplicity, we may consider the case where $h = \eta$.

That is we split the data into K roughly equal-sized parts. For the k -th part, we fit the model to the other $K - 1$ parts of the data and calculate the prediction error of the fitted model when predicting the k -th part of

the data. We do this for $k = 1, 2, \dots, K$ and combine the K estimates of integrated estimation error (4.7).

Let $\kappa : \{1, \dots, N\} \mapsto 1, \dots, K$ be an indexing function that indicates the partition to which observation i is allocated by the randomization. Denote by $\tilde{\theta}_{S,h}^{-k}$ the estimator computed with the k -th part of the data removed. Then a K -fold cross-validation estimate of integrated estimation error (4.7) is

$$CV(\tilde{\theta}_{S,h}, h) = \frac{1}{n} \sum_{i=1}^n \left(\int_0^t \tilde{\theta}_{S,h}^{-\kappa(i)}(u, \vec{Z}_S^i(u))^2 du - 2 \int_0^t \tilde{\theta}_{S,h}^{-\kappa(i)}(u, \vec{Z}_S^i(u)) dN_u^i \right). \quad (4.8)$$

Remarks.

1. Let us first study the theoretical properties of estimator (4.8). To this end note that it is sufficient to minimize the following difference with respect to h in order to solve the minimization problem of the integrated estimation error (4.7) when $h = \eta$:

$$\mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{d_{M_S}}} \left(\tilde{\theta}_{S,h}^{-\kappa(1)}(u, y)^2 - 2 \tilde{\theta}_{S,h}^{-\kappa(1)}(u, y) \theta_S(u, y) \right) f(y) dy du \right]. \quad (4.9)$$

Recall now that for any $1 \leq k \leq n$, conditionally to $\mathcal{S}$ and the σ -algebra generated by the co-process $(Z^1, \dots, Z^n)$ minus its k -th part, N^k is a non-homogeneous Poisson process with intensity function $t \mapsto \theta_S(t, \vec{Z}_S^k(t))$. Using martingale properties of the Poisson process and by smart conditioning, one easily gets that estimator (4.8) is an unbiased estimator of

$$\mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{d_{M_S}}} \left(\tilde{\theta}_{S,h,\eta}^{-\kappa(1)}(u, y) - \theta_S(u, y) \right)^2 f(y) dy du \right],$$

which is close to the aimed estimation error (4.9). So that minimizing (4.7) is theoretically satisfying to solve our problem.

2. On a more practical point of view, Figure 4.4 represents the 10-fold cross-validation estimate (4.8) of integrated estimation error (4.7) for one replicate of model (4.6), with $\beta = 0.1$. We deduce from the resulting curve that the choice $h = 0.78$ minimizes the empirical prediction error for this replicate.

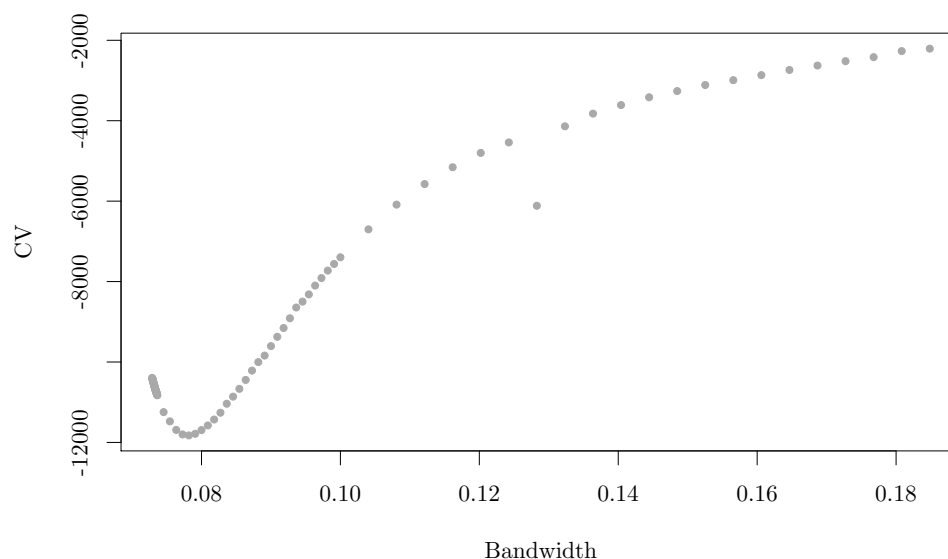


Figure 4.4: One replicate of 10-fold cross-validation estimate of estimation error (4.7) under model (4.6).

4.4 Application to real data

We study a data set constituted of historical prices of $n = 495$ companies as well as the crude oil prices over a period of roughly one year and two months (from 17th April, 2014 to 23rd June, 2015). The companies data are taken from the website Yahoo Finance so that every company considered composed the S&P500 index on the 23rd June, 2015. The crude oil prices are taken from the website Investing.com. The Cox process data consist of the count of the number of times when the percent returns of said companies go below a certain threshold with the counting rate depending on the stochastic dynamics of the company market capitalization. In our case the company market capitalization is represented by the action's trade volume normalized increments and is observed when the percent return of the crude oil action below another threshold. By analyzing this count, we aim to learn the financial properties of this 495 companies system.

Date	Open	High	Low	Close	Volume	Adj.Close
2015-06-23	39.89	39.95	39.42	39.60	2053600	39.60
2015-06-22	39.81	40.01	39.73	39.81	3901700	39.81
2015-06-19	39.80	39.94	39.49	39.49	2581000	39.49
2015-06-18	39.80	40.06	39.72	39.90	1865000	39.90
2015-06-17	39.76	39.80	39.32	39.60	1519400	39.60
2015-06-16	39.59	39.81	39.38	39.79	1422600	39.79
2015-06-15	39.63	39.63	39.25	39.52	2320100	39.52
2015-06-12	40.33	40.49	39.74	39.84	2764200	39.84
2015-06-11	40.57	40.60	40.29	40.53	1566000	40.53
2015-06-10	40.40	40.59	40.27	40.52	1787900	40.52
...						

Table 4.2: First 10 rows of raw data for Agilent Technologies Inc. taken from Yahoo Finance.

Date	Price	Open	High	Low	Vol.	Change
2015-06-23	61.01	60.21	61.49	59.55	336.22K	1.04%
2015-06-22	60.38	59.75	60.63	59.27	255.31K	1.29%
2015-06-19	59.97	60.88	60.93	59.24	299.89K	-1.40%
2015-06-18	60.82	60.10	61.33	59.67	171.48K	0.81%
2015-06-17	60.33	60.52	61.81	59.34	232.09K	-0.20%
2015-06-16	60.45	60.01	60.81	59.88	129.30K	0.75%
2015-06-15	60.00	60.33	60.42	59.19	128.26K	-0.66%
2015-06-12	60.40	60.92	61.06	60.18	91.96K	-1.34%
2015-06-11	61.22	61.56	61.91	60.65	150.62K	-0.97%
2015-06-10	61.82	61.00	62.22	60.88	188.78K	2.00%
...						

Table 4.3: First 10 rows of raw data for the crude oil action taken from Investing.com.

Tables 4.2 and 4.3 present the layout of the raw data directly taken from the websites Yahoo Finance and Investing.com.

We denote $((Y_t^1)_{t \in [0,1]}, \dots, (Y_t^n)_{t \in [0,1]})$ the `Open` columns of the Yahoo Finance raw data. It represents the daily open prices of the actions of the companies. We define the percent returns as follows $X_t^k := \frac{Y_t^k - Y_{t-1}^k}{Y_{t-1}^k}$ for $1 \leq k \leq n$. Denote $(\zeta_t)_{t \in [0,1]}$ the `Open` column of the Investing.com raw data. In the same way as for the Yahoo Finance raw data it represents the daily open prices of the crude oil action. We define its percent returns by $\xi_t := \frac{\zeta_t - \zeta_{t-1}}{\zeta_{t-1}}$. The `Volume` columns of the Yahoo Finance data are denoted $((W_t^1)_{t \in [0,1]}, \dots, (W_t^n)_{t \in [0,1]})$. Its normalized increments are defined for $1 \leq k \leq n$ by $Z_t^k := \frac{W_t^k - W_{t-1}^k}{W_{t-1}^k}$.

Random times S as well as the counting process N are deduced from these transformed data sets. They are defined such that S_1 is the first time at which ξ goes below $\alpha := -0.01$, S_2 is the second time, *etc.* and the process N counts the number of times X goes below $\beta := -0.015$. Figure 4.5 represents the trajectory of ξ and gives an illustration of the construction of S . With this construction we get a total of 50 times of observation. Remark that these thresholds represent a 1% drop for the percent returns of the crude oil action and a 1.5% drop for the percent returns of the S&P500 companies action.

We aim to compare the inhomogeneous Poisson model with our model (4.1). To this end we compute our estimator over the time span defined by the data and for 10 chosen trajectories of the covariate process Z . The resulting estimated intensities are given in Figure 4.6. In most cases (7 out of 10), we estimate the same intensity as in the inhomogeneous Poisson model. In the second graph we observe that for 3 trajectories of Z , taking covariates into consideration in the model provides estimations that stays close to the inhomogeneous Poisson model at first and deviates from it after a short moment.

4.5 Proofs

For the sake of readability we introduce some notations. $\tilde{\mathbb{P}}(\cdot)$ stands for $\mathbb{P}(\cdot|S)$ whereas $\tilde{\tilde{\mathbb{P}}}(\cdot)$ stands for $\mathbb{P}(\cdot|S, Z^1, \dots, Z^n)$. In this context, $\tilde{\mathbb{E}}$ and $\tilde{\tilde{\mathbb{E}}}$ denote the expectations under $\tilde{\mathbb{P}}$ and $\tilde{\tilde{\mathbb{P}}}$ respectively. We also assume that

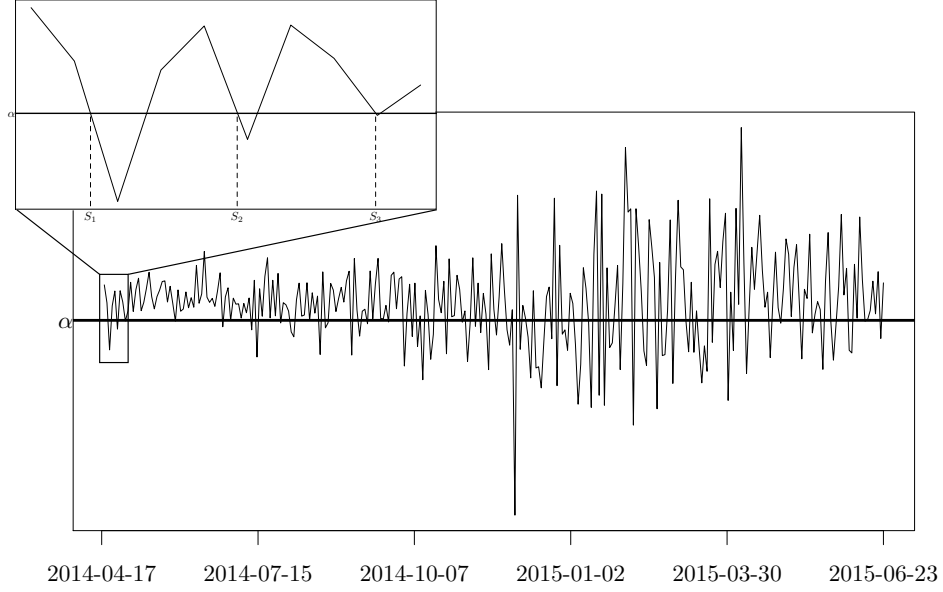


Figure 4.5: Crude oil action percent returns plus a zoom on a small window of time to demonstrate the construction of the random times S .

$t \in [0, 1]$ and function $z : [0, 1] \rightarrow \mathbb{R}$ are fixed so that the dependence on these parameters can be omitted for brevity.

4.5.1 Proof of Theorem 4.1

We aim at bounding the following quantity

$$\text{MSE}_S(t, z) = \mathbb{E} \left(\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) \right)^2.$$

To do so, introduce for any $k = 1, \dots, n$, the following weights:

$$W_k = \frac{\frac{1}{n} H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right)}{\max \left(\frac{1}{n} \sum_{\ell=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^\ell(t) \right), a_n \right)},$$

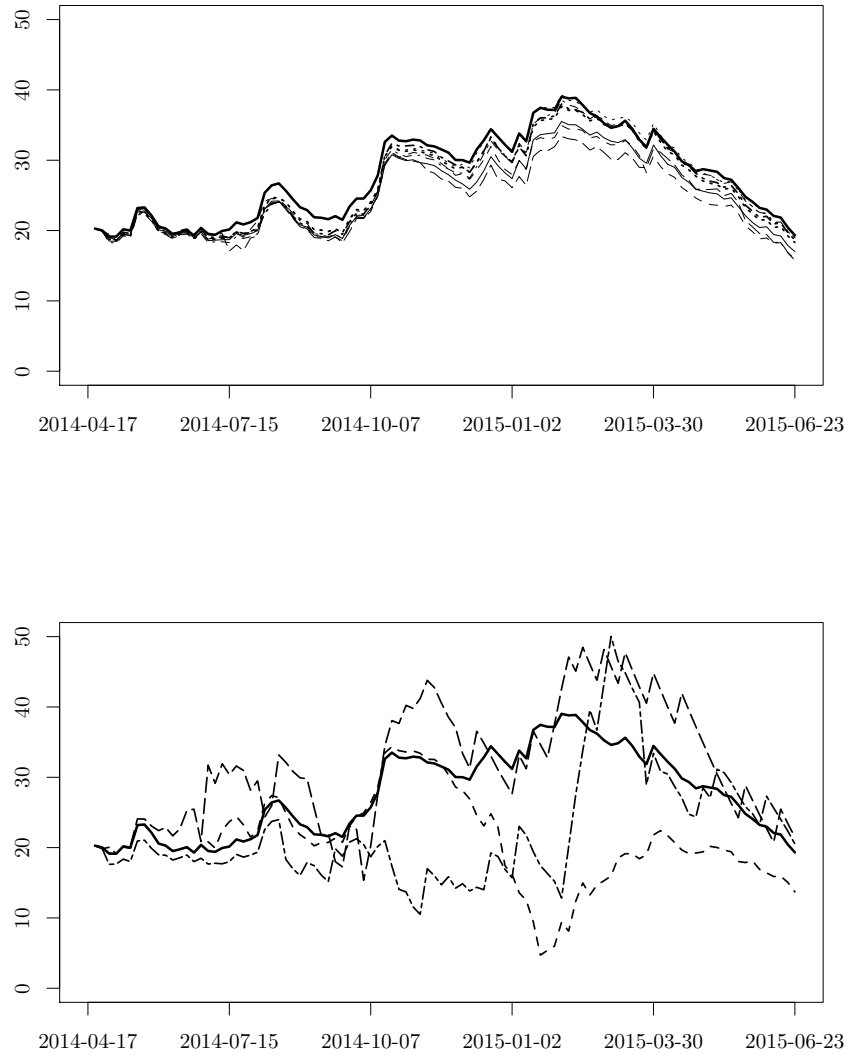


Figure 4.6: Estimation of the intensity function λ in the Cox process model (4.1) for 10 chosen trajectories of the covariate process Z compared to the estimation for an inhomogeneous Poisson model (bold solid line)

where the sequence $(a_n)_{n \in \mathbb{N}}$ is defined by $a_n = (n\eta^{dM_t})^{-1/2}$. We also consider the “compensated” processes $\bar{N}^k = (\bar{N}_t^k)_{t \in [0,1]}$ defined by

$$\bar{N}_t^k = N_t^k - \tilde{\mathbb{E}} N_t^k = N_t^k - \int_0^t \lambda_S(u, Z^k) du.$$

Finally we define:

$$A = \sum_{k=1}^n W_k \int_0^t K_h(t-u) d\bar{N}_u^k,$$

and

$$B = \sum_{k=1}^n W_k \int_0^t K_h(t-u) \lambda_S(u, Z^k) du - \lambda_S(t, z).$$

Using these notations we obtain the decomposition $\tilde{\lambda}_{S,h,\eta}(t, z) - \lambda_S(t, z) = A + B$ which readily implies that:

$$\text{MSE}_S(t, z) \leq 2(\tilde{\mathbb{E}}(A^2) + \tilde{\mathbb{E}}(B^2)). \quad (4.10)$$

We now proceed to control the two terms in the right hand-side of (4.10).

Control of the first term in (4.10). To do so, remark that:

$$\begin{aligned} \tilde{\mathbb{E}}(A^2) &= \tilde{\mathbb{E}} \left(\sum_{k=1}^n W_k \int_0^t K_h(t-u) d\bar{N}_u^k \right)^2 \\ &= \tilde{\mathbb{E}} \left(\sum_{1 \leq k, \ell \leq n} \left(W_k \int_0^t K_h(t-u) d\bar{N}_u^k \right) \left(W_\ell \int_0^t K_h(t-u) d\bar{N}_u^\ell \right) \right) \\ &= \sum_{k=1}^n \tilde{\mathbb{E}} \left(W_k \int_0^t K_h(t-u) d\bar{N}_u^k \right)^2. \end{aligned}$$

Indeed, since W_k is measurable with respect to the σ -algebra generated by S and the process Z^k , we have:

$$\begin{aligned} &\sum_{1 \leq k, n \leq n; k \neq \ell} \tilde{\mathbb{E}} \left(\left(W_k \int_0^t K_h(t-u) d\bar{N}_u^k \right) \left(W_\ell \int_0^t K_h(t-u) d\bar{N}_u^\ell \right) \right) \\ &= \sum_{1 \leq k, n \leq n; k \neq \ell} \tilde{\mathbb{E}} \left(\left(W_k \tilde{\mathbb{E}} \left(\int_0^t K_h(t-u) d\bar{N}_u^k \right) \right) \tilde{\mathbb{E}} \left(W_\ell \left(\int_0^t K_h(t-u) d\bar{N}_u^\ell \right) \right) \right) \\ &= 0, \end{aligned}$$

where last equality comes from the martingale nature of $\bar{N}^k$ under $\tilde{\mathbb{P}}$ which implies that $\int_0^t K_h(t-u)d\bar{N}_u^k$ is a martingale. In particular,

$$\tilde{\mathbb{E}} \left(\int_0^t K_h(t-u)d\bar{N}_u^k \right) = 0.$$

Now recall that under $\tilde{\mathbb{P}}$, N^k is a Poisson process with intensity function $\lambda_S(t, Z^k)$. This implies that the quadratic variation of $\bar{N}^k$ is N^k which also implies, using the Itô isometry, that:

$$\begin{aligned} \tilde{\mathbb{E}}(A^2) &= \sum_{k=1}^n \tilde{\mathbb{E}} \left(W_k^2 \tilde{\mathbb{E}} \left(\int_0^t K_h(t-u)d\bar{N}_u^k \right)^2 \right) \\ &= \sum_{k=1}^n \tilde{\mathbb{E}} \left(W_k^2 \int_0^t K_h^2(t-u)\lambda_S(u, Z^k)du \right) \\ &= \sum_{k=1}^n \tilde{\mathbb{E}} \left(W_k^2 \int_0^t K_h^2(t-u)\theta_S(u, \bar{Z}_S^k(u))du \right). \end{aligned}$$

Now using (H4) and (T1) we obtain:

$$\begin{aligned} \tilde{\mathbb{E}}(A^2) &\leq \sum_{k=1}^n \tilde{\mathbb{E}} \left(W_k^2 \frac{\Theta^{M_t} \|K\|_2^2}{h} \right) \\ &\leq \frac{\Theta^{M_t} \|K\|_2^2}{h} \tilde{\mathbb{E}} \left(\sup_{k=1, \dots, n} W_k \right), \end{aligned}$$

where last inequality comes from the facts that M_t and h are measurable with respect to the σ -algebra generated by S and that, moreover, $\sum_{k=1}^n W_k \leq 1$. Using the definition of W_k combined with (H1), (H4) and (T1), we obtain:

$$\begin{aligned} \tilde{\mathbb{E}}(A^2) &\leq \frac{\Theta^{M_t} \|K\|_2^2 \|\mathcal{H}\|_\infty^{dM_t}}{nh\eta^{dM_t}} \tilde{\mathbb{E}} \left(\frac{1}{\tilde{f}(\bar{z}_S(t))} \right) \\ &\leq \frac{\Theta^{M_t} \|K\|_2^2 \|\mathcal{H}\|_\infty^{dM_t}}{nh\eta^{dM_t}} \left(F_0^{-M_t} + \tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}(\bar{z}_S(t))} - \frac{1}{f(\bar{z}_S(t))} \right| \right). \end{aligned}$$

Using Lemma 4.4 we obtain, for n large enough:

$$\tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}(\bar{z}_S(t))} - \frac{1}{f(\bar{z}_S(t))} \right| \leq 1,$$

which implies:

$$\tilde{\mathbb{E}}(A^2) \leq \frac{(1 + F_0^{-M_t}) \|K\|_2^2 (\Theta \|\mathcal{H}\|_\infty^d)^{M_t}}{nh\eta^{dM_t}}. \quad (4.11)$$

Control of the second term in (4.10). To do so we introduce some notations that are used here and after:

$$\begin{aligned} \tilde{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) &= \frac{1}{n} \sum_{k=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right) \int_0^t K_h(t-u) \lambda_S(u, Z^k) du \\ &= \frac{1}{n} \sum_{k=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right) \int_0^t K_h(t-u) \theta_S(u, \vec{Z}_S^k(u)) du, \end{aligned}$$

and

$$\phi(t, \vec{z}_S(t)) = f(\vec{z}_S(t)) \theta_S(t, \vec{z}_S(t)).$$

Using the above notations and (H1), B can be written as:

$$B = \frac{\tilde{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{\tilde{f}_{S,h,\eta}(\vec{z}_S(t))} - \frac{\phi(t, \vec{z}_S(t))}{f(\vec{z}_S(t))}.$$

For the sake of readability we omit the subscripts and the parameters of the functions. Hence $\tilde{\phi}$ denotes $\tilde{\phi}_{S,h,\eta}(t, \vec{z}_S(t))$ and so on. We then obtain:

$$B^2 \leq 3 \left(\left(\frac{\tilde{\phi} - \phi}{f} \right)^2 + (\tilde{\phi} - \phi)^2 \left(\frac{1}{\tilde{f}} - \frac{1}{f} \right)^2 + \phi^2 \left(\frac{1}{\tilde{f}} - \frac{1}{f} \right)^2 \right).$$

Now, combining the fact that for any real numbers x and y we have $x^2 y^2 \leq (x^4 + y^4)/2$ with (H4), we obtain:

$$B^2 \leq C_1(M_t) \left((\tilde{\phi} - \phi)^2 + (\tilde{\phi} - \phi)^4 + \left(\frac{1}{\tilde{f}} - \frac{1}{f} \right)^2 + \left(\frac{1}{\tilde{f}} - \frac{1}{f} \right)^4 \right),$$

where $C_1(M_t) = 3 \max(F_0^{-2M_t}, (\Theta F_\infty)^{2M_t}, 1/2)$. It remains to upper bound $\tilde{\mathbb{E}}(\tilde{\phi} - \phi)^q$ and $\tilde{\mathbb{E}}(\tilde{f}^{-1} - f^{-1})^q$ for $q = 2$ and $q = 4$. Combining the fact

that $h = \eta = n^{-\frac{1}{5+dM_t}}$ with **(H1)** to **(H4)** and **(T1)** and using Lemmas 4.4 and 4.5 and their notations, we obtain:

$$\tilde{\mathbb{E}}(B^2) \leq C_2(M_t) n^{-\frac{4}{5+dM_t}}, \quad (4.12)$$

where $C_2(M_t) = C_1(M_t) (\Delta_{1,2}(M_t) + \Delta_{1,4}(M_t) + \Delta_{2,2}(M_t) + \Delta_{2,4}(M_t))$.

Finally, combining (4.10), and (4.11) with $h = \eta = n^{-\frac{1}{5+dM_t}}$ and (4.12) we obtain:

$$\text{MSE}_S(t, z) \leq C_3(M_t) n^{-\frac{4}{5+dM_t}},$$

where $C_3(M_t) = 2 \left((2 + F_0^{-M_t}) \|K\|_2^2 (\Theta \|\mathcal{H}\|_\infty^d)^{M_t} + C_2(M_t) \right)$. Combining above with the fact that there exists $\tau > 0$ depending only on F_0, F_∞, Θ, Q and $\mathbf{k}$ such that $C_3(M_t) \leq e^{\tau M_t}$ gives the first bound of Theorem 4.1. The second bound follows using **(H5)**.

4.5.2 Proof of Proposition 4.2

Using the dominated convergence theorem we have to prove that:

$$\tilde{\lambda}_{S,h,\eta}(t, z) \xrightarrow[n \rightarrow +\infty]{\tilde{\mathbb{P}}} \lambda_S(t, z).$$

Using the condensed notations introduced above and **(H1)**, we use the following decomposition:

$$\tilde{\lambda}_S - \lambda_S = \frac{(\hat{\phi} - \phi)}{\tilde{f}} + \frac{\phi(f - \tilde{f})}{\tilde{f}f}. \quad (4.13)$$

This, combined with Slutsky's Lemma, shows that it is sufficient to prove the following properties

$$\hat{\phi} \xrightarrow[n \rightarrow +\infty]{\tilde{\mathbb{P}}} \phi, \quad \text{and} \quad \tilde{f} \xrightarrow[n \rightarrow +\infty]{\tilde{\mathbb{P}}} f \quad (4.14)$$

which are also implied by the convergence in the $\mathbb{L}^2$ -norm. Using the first item of Lemma 4.7 combined with Lemma 4.8 and **(H2)** and **(H4)**, we easily check that:

$$\tilde{\mathbb{E}}(\hat{\phi} - \phi)^2 = \tilde{\mathbb{E}}(\tilde{\phi} - \phi)^2 \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{and} \quad \tilde{\mathbb{E}}(\tilde{f} - f)^2 \leq 2(a_n^2 + \tilde{\mathbb{E}}(\tilde{f} - f)^2) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Proposition follows.

4.5.3 Proof of Theorem 4.3

Note that, using (H1), (4.13) and (4.14) combined with Slutsky's Lemma, we only have to study the convergence of the following random variables:

$$(nh\eta^{dM_t})^{1/2}(\hat{\phi} - \phi) \quad \text{and} \quad (nh\eta^{dM_t})^{1/2}(\tilde{f} - f).$$

To this end, let us define

$$\Xi_1 = (nh\eta^{dM_t})^{1/2} \left(\hat{\phi} - \phi - (\tilde{\mathbb{E}}\hat{\phi} - \phi) \right),$$

and

$$\Xi_2 = (nh\eta^{dM_t})^{1/2} \left(\tilde{f} - f - (\tilde{\mathbb{E}}\hat{f} - f) \right).$$

First, notice that

$$\begin{aligned} \tilde{\mathbb{E}}\Xi_2^2 &= nh\eta^{dM_t} \tilde{\mathbb{E}} \left(\tilde{f} - \hat{f} + \hat{f} - \tilde{\mathbb{E}}\hat{f} \right)^2 \\ &\leq 2nh\eta^{dM_t} \left(\tilde{\mathbb{E}}(\tilde{f} - \hat{f})^2 + \tilde{\mathbb{E}}(\hat{f} - \tilde{\mathbb{E}}\hat{f})^2 \right). \end{aligned} \quad (4.15)$$

Now, since $\tilde{f} = \max(\hat{f}, a_n)$, we deduce that $\tilde{\mathbb{E}}(\tilde{f} - \hat{f})^2 \leq a_n^2$. Then, remark that the second term of the right-hand side of equation (4.15) may be written as follows:

$$\tilde{\mathbb{E}}(\hat{f} - \tilde{\mathbb{E}}\hat{f})^2 = \tilde{\mathbb{E}} \left(\frac{1}{n} \sum_{k=1}^n \zeta_k \right)^2, \quad (4.16)$$

where $\zeta_k = H_\eta(\vec{z}_S(t) - \vec{Z}_S^k(t)) - \tilde{\mathbb{E}}H_\eta(\vec{z}_S(t) - \vec{Z}_S^k(t))$ are independent centered variables under $\tilde{\mathbb{P}}$. Combining (H4), (T1) and Lemma 4.8 with random variables ζ_k and bounds on the first and second moments of ζ_k respectively equal to $2\|\mathcal{H}\|_\infty^{dM_t}\eta^{-dM_t}$ and $F_\infty^{M_t}\|\mathcal{H}\|_2^{2dM_t}\eta^{-dM_t}$ to bound this term, we get:

$$\tilde{\mathbb{E}}(\hat{f} - \tilde{\mathbb{E}}\hat{f})^2 \leq F_\infty^{M_t}\|\mathcal{H}\|_2^{2dM_t}(n\eta^{dM_t})^{-1}. \quad (4.17)$$

Using the definition of $a_n = (n\eta^{dM_t})^{-1/2}$ combined with (4.15) and above, we deduce that

$$\tilde{\mathbb{E}}\Xi_2^2 \leq 2(1 + F_\infty^{M_t}\|\mathcal{H}\|_2^{2dM_t})h, \quad (4.18)$$

so that $\tilde{\mathbb{E}}\Xi_2^2 \rightarrow 0$ which also implies $\Xi_2 \xrightarrow{\tilde{\mathbb{P}}} 0$.

Let us now study the convergence of Ξ_1 . Notice that

$$\Xi_1 = (nh\eta^{dM_t})^{1/2}(\hat{\phi} - \tilde{\mathbb{E}}\hat{\phi}).$$

Then, using Lemma 4.6 with **(H1)**, **(H2)**, **(H4)**, **(T1)** and **(T2)**, we get under $\tilde{\mathbb{P}}$:

$$\Xi_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \phi(t, \vec{z}_S(t)) \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t}$.

Using Slutsky's Lemma and combining (4.13), (4.18) and above, we get, under $\tilde{\mathbb{P}}$:

$$(nh\eta^{dM_t})^{1/2} \left(\tilde{\lambda}_S - \lambda_S + \frac{\tilde{\mathbb{E}}\hat{\phi} - \phi}{f} - \frac{\phi(\tilde{\mathbb{E}}\hat{f} - f)}{f^2} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2). \quad (4.19)$$

Now, under **(H3)**, **(T1)** and **(T2)**, according to (4.26) in Lemma 4.7, we also know that

$$\tilde{\mathbb{E}}\hat{f} - f = \frac{\eta^2}{2} m_2(\mathcal{H}) \sum_{i=1}^{dM_t} \partial_i^2 f(\vec{z}_S(t)) + o(\eta^2),$$

and

$$\tilde{\mathbb{E}}\hat{\phi} - \phi = \frac{h^2}{2} m_2(K) \partial_1^2 \phi(t, \vec{z}_S(t)) + \frac{\eta^2}{2} m_2(\mathcal{H}) \sum_{i=2}^{dM_t+1} \partial_i^2 \phi(t, \vec{z}_S(t)) + o(h^2 + \eta^2),$$

so that, since $nh^5\eta^{dM_t} = O(1)$ and $nh\eta^{dM_t+4} = O(1)$, another way of writing (4.19) is, under $\tilde{\mathbb{P}}$:

$$(nh\eta^{dM_t})^{1/2} \left(\tilde{\lambda}_S - \lambda_S + h^2 B_1 + \eta^2 B_2 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$B_1 = \frac{1}{2} m_2(K) \frac{1}{f(\vec{z}_S(t))} \partial_1^2 \phi(t, \vec{z}_S(t)),$$

and,

$$B_2 = \frac{1}{2} m_2(\mathcal{H}) \sum_{i=1}^{dM_t} \left(\frac{1}{f(\vec{z}_S(t))} \partial_{i+1}^2 \phi(t, \vec{z}_S(t)) - \frac{\phi(t, \vec{z}_S(t))}{f^2(\vec{z}_S(t))} \partial_i^2 f(\vec{z}_S(t)) \right).$$

Theorem follows.

4.6 Auxiliary results

Lemma 4.4. Set $\eta = n^{-\frac{1}{5+dM_t}}$. Under assumptions **(H1)** to **(H4)**, **(T1)** and **(T2)** we have, for any integer $q \geq 1$:

$$\tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}_{S,h,\eta}(\tilde{z}_S(t))} - \frac{1}{f(\tilde{z}_S(t))} \right|^q \leq \Delta_{1,q}(M_t) n^{-\frac{2q}{5+dM_t}},$$

where $\Delta_{1,q}(M_t) = 2^{3q} C_{1,q}(M_t)(2 + C_{4,q}(M_t) + C_{4,2q}(M_t) + C_{4,3q}(M_t))$, and constant $C_{4,q}(M_t)$ is defined in (4.20).

Proof of Lemma 4.4 Note that, using condensed notations similar to those introduced above and **(H1)**, we have:

$$\frac{1}{\tilde{f}} - \frac{1}{f} = \frac{f - \tilde{f}}{f^2} + \frac{(f - \tilde{f})^2}{f^3} + \frac{(f - \tilde{f})^3}{f^3 \tilde{f}}.$$

Using **(H4)** combined with the fact that $\tilde{f} \geq a_n$ we deduce

$$\tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}} - \frac{1}{f} \right|^q \leq C_{1,q}(M_t) \left(\tilde{\mathbb{E}}|f - \tilde{f}|^q + \tilde{\mathbb{E}}|f - \tilde{f}|^{2q} + a_n^{-q} \tilde{\mathbb{E}}|f - \tilde{f}|^{3q} \right)$$

where $C_{1,q}(M_t) = 3^{q-1} \max(F_0^{-2}, F_0^{-3})^{qM_t}$. Now, remark that $|\tilde{f} - f| \leq |\tilde{f} - \hat{f}| + |\hat{f} - f|$. Since $\tilde{f} = \max(\hat{f}, a_n)$ we deduce that the quantity $|\tilde{f} - \hat{f}|$ is bounded by a_n . Taking all together we obtain:

$$\tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}} - \frac{1}{f} \right|^q \leq 2^{3q} C_{1,q}(M_t) \left(a_n^q + \tilde{\mathbb{E}}|f - \hat{f}|^q + a_n^{2q} + \tilde{\mathbb{E}}|f - \hat{f}|^{2q} + a_n^{-q} \tilde{\mathbb{E}}|f - \hat{f}|^{3q} \right)$$

It remains to bound $\tilde{\mathbb{E}}|f - \hat{f}|^\ell$ for any integer ℓ . To do so we use the classical decomposition of this risk into a bias term and a stochastic term:

$$\begin{aligned} \tilde{\mathbb{E}}|f - \hat{f}|^\ell &\leq 2^{\ell-1} \left(\left| \tilde{\mathbb{E}}\hat{f} - f \right|^\ell + \tilde{\mathbb{E}} \left| \hat{f} - \tilde{\mathbb{E}}\hat{f} \right|^\ell \right) \\ &\leq 2^{\ell-1} (I_1(\ell) + J_1(\ell)) \end{aligned}$$

where

$$I_1(\ell) = \left| \int_{\mathbb{R}^{dM_t}} H_\eta(\tilde{z}_S(t) - u) (f(u) - f(\tilde{z}_S(t))) du \right|^\ell$$

and

$$J_1(\ell) = \tilde{\mathbb{E}} \left| \frac{1}{n} \sum_{k=1}^n \zeta_k \right|^\ell,$$

where ζ_k are independent centered variables under $\tilde{\mathbb{P}}$ defined in (4.16).

The bias term $I_1(\ell)$ can be bounded using (H3), (T1), (T2) and (4.25) in Lemma 4.7 with kernel H , bandwidth η and function f whereas, as seen in (4.17), (H4), (T1) and Lemma 4.8 can be combined to bound the stochastic term $J_1(\ell)$. Recalling that $\eta = n^{-\frac{1}{5+dM_t}}$, we thus obtain:

- if $1 \leq \ell \leq 2$,

$$\tilde{\mathbb{E}} |f - \hat{f}|^\ell \leq C_{2,\ell}(M_t) n^{-\frac{2\ell}{5+dM_t}},$$

$$\text{where } C_{2,\ell}(M_t) = 2^{\ell-1} \left((dM_t)^{2\ell} Q^{\ell M_t} \|\mathcal{H}\|_1^{\ell d M_t} + F_\infty^{\ell M_t/2} \|\mathcal{H}\|_2^{\ell d M_t} \right),$$

- if $\ell > 2$

$$\tilde{\mathbb{E}} |f - \hat{f}|^\ell \leq C_{3,\ell}(M_t) n^{-\frac{2\ell}{5+dM_t}},$$

$$\text{where } C_{3,\ell}(M_t) = \max(\mathfrak{K}_\ell, 1) C_{2,\ell}(M_t) + 2^{2\ell-3} \mathfrak{K}_\ell F_\infty^{M_t} (\|\mathcal{H}\|_2^2 \|\mathcal{H}\|_\infty^{l-2})^{dM_t}.$$

Since $a_n = (n\eta^{dM_t})^{-1/2}$, we get:

$$\tilde{\mathbb{E}} \left| \frac{1}{\tilde{f}(\vec{z}_S(t))} - \frac{1}{f(\vec{z}_S(t))} \right|^q \leq \Delta_{1,q}(M_t) n^{-\frac{2q}{5+dM_t}},$$

where $\Delta_{1,q}(M_t) = 2^{3q} C_{1,q}(M_t) (2 + C_{4,q}(M_t) + C_{4,2q}(M_t) + C_{4,3q}(M_t))$ and

$$C_{4,q}(M_t) = \begin{cases} C_{2,q}(M_t) & \text{if } 1 \leq q \leq 2 \\ C_{3,q}(M_t) & \text{if } q > 2. \end{cases} \quad (4.20)$$

Lemma follows.

Lemma 4.5. Set $h = \eta = n^{-\frac{1}{5+dM_t}}$. Under assumptions (H1) to (H4), (T1) and (T2) we have, for any integer $q \geq 1$:

$$\tilde{\mathbb{E}} \left| \tilde{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) - \phi(t, \vec{z}_S(t)) \right|^q \leq \Delta_{2,q}(M_t) n^{-\frac{2q}{5+dM_t}},$$

where for $1 \leq q \leq 2$, $\Delta_{2,q}(M_t) = C_{5,q}(M_t)$ and for $q > 2$, $\Delta_{2,q}(M_t) = C_{6,q}(M_t)$ with $C_{5,q}(M_t)$ and $C_{6,q}(M_t)$ respectively defined in (4.21) and (4.22).

Proof of Lemma 4.5 Once again, let us use the classical decomposition of the risk into a bias term and a stochastic term, using condensed notations similar to those introduced above and (H1), we get:

$$\begin{aligned}\tilde{\mathbb{E}}|\tilde{\phi} - \phi|^q &\leq 2^{q-1} \left(\left| \tilde{\mathbb{E}}\tilde{\phi} - \phi \right|^q + \tilde{\mathbb{E}} \left| \tilde{\phi} - \tilde{\mathbb{E}}\tilde{\phi} \right|^q \right) \\ &\leq 2^{q-1} (I_2(q) + J_2(q))\end{aligned}$$

where

$$I_2(q) = \left| \tilde{\mathbb{E}}\tilde{\phi} - \phi \right|^q,$$

and

$$J_2(q) = \tilde{\mathbb{E}} \left| \tilde{\phi} - \tilde{\mathbb{E}}\tilde{\phi} \right|^q.$$

Note that, for h small enough $(t - u)/h \in \text{supp}(K)$ implies that $u \in I_t$. This implies that $\vec{Z}_S^k(u) = \vec{Z}_S^k(t)$ which also implies:

$$\tilde{\phi}(t, z) = \frac{1}{n} \sum_{k=1}^n H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right) \int_0^t K_h(t - u) \theta_S(u, \vec{Z}_S^k(t)) du,$$

and

$$\tilde{\mathbb{E}}\tilde{\phi}(t, z) = \int_{\mathbb{R}^{dM_t}} \int_{\mathbb{R}} H_\eta(\vec{z}_S(t) - y) K_h(t - u) \theta_S(u, y) f(y) dy du.$$

Thus

$$I_2(q) = \left| \int_{\mathbb{R}^{dM_t}} \int_{\mathbb{R}} H_\eta(\vec{z}_S(t) - y) K_h(t - u) (\theta_S(u, y) f(y) - \phi(t, \vec{z}_S(t))) dy du \right|^q,$$

and

$$J_2(q) = \tilde{\mathbb{E}} \left| \frac{1}{n} \sum_{k=1}^n \xi_k \right|^q.$$

where $\xi_k = H_\eta(\vec{z}_S(t) - \vec{Z}_S^k(t)) \int_0^t K_h(t - s) \lambda_S(s, Z^k) ds - \tilde{\mathbb{E}}\tilde{\phi}(t, \vec{z}_S(t))$ are independent centered variables under $\tilde{\mathbb{P}}$.

The bias term $I_2(q)$ can be bounded using (H3), (T1), (T2) and (4.25) in Lemma 4.7 with kernel $H \otimes K$, bandwidths h and η and function ϕ whereas (H4), (T1), (T2) and Lemma 4.8 with random variables ξ_k and bounds on the first and second moments of ξ_k respectively equal to $2\|\mathcal{H}\|_\infty^{dM_t} \eta^{-dM_t}$ and $F_\infty^{M_t} \Theta^{2M_t} \|\mathcal{H}\|_2^{2dM_t} \eta^{-dM_t}$ can be combined to bound the stochastic term $J_2(q)$. Recalling that $h = \eta = n^{-\frac{1}{5+dM_t}}$, we thus obtain:

- if $1 \leq q \leq 2$,

$$\tilde{\mathbb{E}}|\tilde{\phi} - \phi|^q \leq C_{5,q}(M_t)n^{-\frac{2q}{5+dM_t}}, \quad (4.21)$$

where

$$C_{5,q}(M_t) = 2^{q-1} \left((dM_t + 1)^{2q} Q^{qM_t} \|K\|_1^q \|\mathcal{H}\|_1^{qdM_t} + F_\infty^{qM_t/2} \Theta^{qM_t} \|\mathcal{H}\|_2^{qdM_t} \right),$$

- if $q > 2$

$$\tilde{\mathbb{E}}|\tilde{\phi} - \phi|^q \leq C_{6,q}(M_t)n^{-\frac{2q}{5+dM_t}}, \quad (4.22)$$

where

$$C_{6,q}(M_t) = \max(\mathfrak{K}_q, 1) C_{5,q}(M_t) + 2^{2q-3} \mathfrak{K}_q F_\infty^{M_t} \Theta^{2M_t} (\|\mathcal{H}\|_2^2 \|\mathcal{H}\|_\infty^{q-2})^{dM_t}.$$

Lemma follows.

Lemma 4.6. Assume that **(H1)**, **(H2)**, **(H4)**, **(T1)** and **(T2)** are satisfied. Under $\tilde{\mathbb{P}}$, we have

$$(nh\eta^{dM_t})^{1/2} \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) - \tilde{\mathbb{E}}\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{(\phi(t, \vec{z}_S(t)) \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t})^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Proof of Lemma 4.6 Denote

$$\hat{L}_k^n = (nh\eta^{dM_t})^{1/2} \frac{H_\eta \left(\vec{z}_S(t) - \vec{Z}_S^k(t) \right) \int_0^t K_h(t-s) dN_S^k - \tilde{\mathbb{E}}\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{n (\phi(t, \vec{z}_S(t)) \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t})^{1/2}},$$

and

$$\hat{L}^n = (nh\eta^{dM_t})^{1/2} \frac{\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t)) - \tilde{\mathbb{E}}\hat{\phi}_{S,h,\eta}(t, \vec{z}_S(t))}{(\phi(t, \vec{z}_S(t)) \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t})^{1/2}},$$

so that $\hat{L}^n = \sum_{k=1}^n \hat{L}_k^n$. For any, $n \in \mathbb{N}^*$, the $\hat{L}_1^n, \dots, \hat{L}_n^n$ are independent centered random variables under $\tilde{\mathbb{P}}$ and

$$\tilde{\mathbb{E}}(\hat{L}^n - \tilde{\mathbb{E}}\hat{L}^n)^2 = \sum_{k=1}^n \tilde{\mathbb{E}}(\hat{L}_k^n - \tilde{\mathbb{E}}\hat{L}_k^n).$$

As a result, according to Lyapunov's Theorem (see Billingsley, 2013), we need only to prove that the Lyapunov condition

$$\tilde{\mathbb{E}}^{-2}(\hat{L}^n - \tilde{\mathbb{E}}\hat{L}^n)^2 n \tilde{\mathbb{E}} \left| \hat{L}_1^n \right|^4 \rightarrow 0, \quad (4.23)$$

is satisfied. To this end, we remark that

$$\tilde{\mathbb{E}}(\hat{L}^n - \tilde{\mathbb{E}}\hat{L}^n)^2 = \frac{nh\eta^{dM_t}}{\phi \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t}} \tilde{\mathbb{E}}(\hat{\phi} - \tilde{\mathbb{E}}\hat{\phi})^2.$$

Under assumptions **(H1)**, **(H2)** and **(H4)** on the model, using the Poisson nature of N under $\tilde{\mathbb{P}}$, it is easily seen that

$$nh\eta^{dM_t} \tilde{\mathbb{E}}(\hat{\phi} - \tilde{\mathbb{E}}\hat{\phi})^2 \xrightarrow{n \rightarrow +\infty} \phi \|K\|_2^2 \|\mathcal{H}\|_2^{2dM_t},$$

which also implies $\tilde{\mathbb{E}}(\hat{L}^n - \tilde{\mathbb{E}}\hat{L}^n)^2 \rightarrow 1$. Basic martingale properties as well as the Burkholder-Davis-Gundy inequality (see Burkholder et al., 1972) and Lemma 4.8 then give $n\tilde{\mathbb{E}}(\hat{L}_1^n)^4 \rightarrow 0$. This ensures that the Lyapunov condition (4.23) is satisfied for $\delta = 2$. Lemma follows.

Lemma 4.7. *Let δ be a positive integer. Let $\mathcal{K}^{(1)}, \dots, \mathcal{K}^{(\delta)}$ be univariate bounded kernels with compact support included in $[-1, 1]$ and define $b = (b_1, \dots, b_\delta) \in [0, 1]^\delta$ a bandwidth. Define the kernel product*

$$\mathcal{K}_b(u) = \prod_{i=1}^{\delta} b_i^{-1} \mathcal{K}^{(i)}(b_i^{-1} u_i).$$

For any function $\psi : \mathbb{R}^\delta \rightarrow \mathbb{R}$ and $x \in \mathbb{R}^\delta$, define

$$I_b(\psi, x) = \int_{\mathbb{R}^\delta} \mathcal{K}_b(x - u)(\psi(u) - \psi(x)) du.$$

The following properties hold:

- If ψ is a bounded continuous function then $|I_b(\psi, x)| \xrightarrow{b \rightarrow 0} 0$.
- If ψ is twice continuously differentiable and is such that, for any $y \in \mathbb{R}^\delta$ the entries $\partial_{i,j}^2 \psi(y)$ of the Hessian matrix are bounded by M , and assume that for any $i = 1, \dots, \delta$:

$$\int_{\mathbb{R}} \mathcal{K}^{(i)}(v) v dv = 0, \quad (4.24)$$

then

$$|I_b(\psi, x)| \leq M \|K\|_1 \sum_{1 \leq i, j \leq \delta} b_i b_j, \quad (4.25)$$

and,

$$I_h(\psi, x) = \sum_{i=1}^{\delta} \frac{b_i^2}{2} m_2(\mathcal{K}^{(i)}) \partial_i^2 \psi(x) + \sum_{1 \leq i, j \leq \delta} b_i b_j \varepsilon_{i,j}(x, b), \quad (4.26)$$

where

$$\varepsilon_{i,j}(x, b) = (1 + \mathbf{1}_{i \neq j}) \int_{\mathbb{R}^d} \mathcal{K}(u) \int_0^1 (1 - \tau) (\partial_{i,j}^2 \psi(x - \tau b u) - \partial_{i,j}^2 \psi(x)) d\tau u_i u_j du,$$

tends to 0 as $b \rightarrow 0$.

Proof of Lemma 4.7 First item follows immediately from Bochner's Lemma. To prove second item, we use the following Taylor's expansions:

$$\psi(x - bu) - \psi(x) = \sum_{i=1}^{\delta} (-b_i u_i) \partial_i \psi(x) + R_1(x, b, u) \quad (4.27)$$

$$\begin{aligned} &= \sum_{i=1}^{\delta} (-b_i u_i) \partial_i \psi(x) + \sum_{i=1}^{\delta} \frac{b_i^2}{2} \partial_i^2 \psi(x) u_i^2 \\ &\quad + \sum_{i \neq j} b_i b_j \partial_{i,j}^2 \psi(x) u_i u_j + R_2(x, b, u) \end{aligned} \quad (4.28)$$

where

$$R_1(x, b, u) = \sum_{1 \leq i, j \leq \delta} b_i b_j (1 + \mathbf{1}_{i \neq j}) \int_0^1 (1 - \tau) \partial_{i,j}^2 \psi(x - \tau b u) d\tau u_i u_j,$$

and,

$$R_2(x, b, u) = \sum_{1 \leq i, j \leq \delta} b_i b_j (1 + \mathbf{1}_{i \neq j}) \int_0^1 (1 - \tau) (\partial_{i,j}^2 \psi(x - \tau b u) - \partial_{i,j}^2 \psi(x)) d\tau u_i u_j.$$

Combining the product form of the kernel, Fubini's theorem, bounding assumption on the entries of the Hessian matrix of ψ and (4.24) with (4.27),

item (4.25) follows. Item (4.26) follows using the same arguments combined with (4.28).

The following lemma is stated for the convenience of the reader (see Bretagnolle and Huber, 1979).

Lemma 4.8. *Let $Y_1, \dots, Y_n$ be n independent and identically distributed random variables such that $\mathbb{E}(Y_1) = 0$, $\mathbb{E}(Y_1^2) \leq \sigma^2$ and $|Y_1| \leq M$ almost surely. Then, for any $r \geq 1$ we have:*

- if $1 \leq r \leq 2$,

$$\mathbb{E} \left| \frac{1}{n} \sum_{k=1}^n Y_k \right|^r \leq \left(\frac{\sigma^2}{n} \right)^{r/2},$$

- if $r > 2$,

$$\mathbb{E} \left| \frac{1}{n} \sum_{k=1}^n Y_k \right|^r \leq \mathfrak{K}_r \left(\frac{\sigma^2 M^{r-2}}{n^{r-1}} + \left(\frac{\sigma^2}{n} \right)^{r/2} \right),$$

where $\mathfrak{K}_r$ is a positive constant depending solely on r .

Bibliography

- P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics, 1993. (Cited on page 36.)
- S. Asmussen. *Applied probability and queues*, volume 51. Springer Science & Business Media, 2008. (Cited on page 60.)
- S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 14. World Scientific Publishing Company, 2010. (Cited on pages 37, 107, and 108.)
- V. E. Bening and V. Y. Korolev. *Generalized Poisson models and their applications in insurance and finance*. Walter de Gruyter, 2002. (Cited on page 20.)
- W. Bialek, F. Rieke, R. De Ruyter van Steveninck, and D. Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991. (Cited on page 107.)
- G. Biau, F. Cérou, and A. Guyader. Rates of convergence of the functional-nearest neighbor estimate. *Information Theory, IEEE Transactions on*, 56(4):2034–2040, 2010. (Cited on pages 28, 39, 43, 80, 109, and 115.)
- P. J. Bickel. On adaptive estimation. *The Annals of Statistics*, pages 647–671, 1982. (Cited on page 110.)
- P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013. (Cited on pages 69 and 138.)
- L. Birgé. Approximation dans les espaces métriques et théorie de l’estimation. *Probability Theory and Related Fields*, 65(2):181–237, 1983. (Cited on page 83.)

- T. Björk and J. Grandell. Exponential inequalities for ruin probabilities in the cox case. *Scandinavian Actuarial Journal*, 1988(1-3):77–111, 1988. (Cited on page 46.)
- D. Bohning. A note on a test for poisson overdispersion. *Biometrika*, 81(2):418–419, 1994. (Cited on pages 25 and 50.)
- A. N. Borodin and P. Salminen. *Handbook of Brownian motion-facts and formulae*. Birkhäuser, 2012. (Cited on page 53.)
- P. Brémaud. *Point processes and queues*. Springer Series in Statistics, 1981. (Cited on pages 12 and 21.)
- J. Bretagnolle and C. Huber. Estimation des densités : risque minimax. *Probability Theory and Related Fields*, 47(2):119–137, 1979. (Cited on page 140.)
- R. Brette. Generation of correlated spike trains. *Neural computation*, 21(1):188–215, 2008. (Cited on page 107.)
- L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association*, 100(469):36–50, 2005. (Cited on pages 60 and 61.)
- D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proc. Sixth Berkeley Symp. Math. Statist. Prob*, volume 2, pages 223–240, 1972. (Cited on page 138.)
- B. Cadre and L. Truquet. Nonparametric regression estimation onto a poisson point process covariate. *ESAIM: Probability and Statistics*, 19: 251–267, 2015. (Cited on pages 29, 80, and 90.)
- B. Cadre, N. Klutchnikoff, and G. Massiot. Minimax regression estimation for poisson coprocess. *ESAIM: Probability and Statistics*, 2017a. (Cited on page 79.)
- B. Cadre, G. Massiot, and L. Truquet. Nonparametric tests for cox processes. *Journal of Statistical Planning and Inference*, 184:48–61, 2017b. (Cited on page 45.)

- B. W. Carroll and D. A. Ostlie. *An introduction to modern astrophysics and cosmology*, volume 1. 2007. (Cited on page 46.)
- G. Chagny and A. Roche. Adaptive estimation in the functional nonparametric regression model. *Journal of Multivariate Analysis*, 146:105–118, 2016. (Cited on pages 28 and 80.)
- D. D. Cox. Multivariate smoothing spline functions. *SIAM Journal on Numerical Analysis*, 21(4):789–813, 1984. (Cited on page 36.)
- D. D. Cox and F. O’Sullivan. Asymptotic analysis of penalized likelihood and related estimators. *The Annals of Statistics*, pages 1676–1695, 1990. (Cited on page 36.)
- D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955. (Cited on pages 19, 20, 22, and 36.)
- D. R. Cox and V. Isham. *Point processes*, volume 12. CRC Press, 1980. (Cited on page 46.)
- R. Davidson and J. G. MacKinnon. The power of bootstrap and asymptotic tests. *Journal of Econometrics*, 133(2):421–441, 2006. (Cited on pages 25 and 50.)
- M. Denuit, X. Maréchal, S. Pitrebois, and J.-F. Walhin. *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons, 2007. (Cited on pages 25 and 50.)
- L. Devroye and G. Lugosi. A universally acceptable smoothing factor for kernel density estimates. *The Annals of Statistics*, pages 2499–2512, 1996. (Cited on page 78.)
- R. F. Engle. Wald, likelihood ratio, and lagrange multiplier tests in econometrics. *Handbook of econometrics*, 2:775–826, 1984. (Cited on pages 25 and 50.)
- J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001. (Cited on page 121.)

- S. Gaïffas and A. Guillaoux. High-dimensional additive hazard models and the lasso. *Electronic Journal of Statistics*, 6:522–546, 2012. (Cited on pages 37 and 108.)
- W. Gerstner and W. M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002. (Cited on page 46.)
- J. Grandell. *Aspects of risk theory*. Springer Science & Business Media, 2012. (Cited on page 46.)
- L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Science & Business Media, 2006. (Cited on pages 27, 39, 43, 80, 109, 114, and 115.)
- N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015. (Cited on pages 37 and 108.)
- A. Heuer, C. Mueller, and O. Rubner. Soccer: Is scoring goals a predictable poissonian process? *EPL (Europhysics Letters)*, 89(3):38007, 2010. (Cited on pages 61 and 62.)
- L. Horváth and P. Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012. (Cited on pages 27 and 80.)
- R. Ibragimov and S. Sharakhmetov. The exact constant in the rosenthal inequality for random variables with mean zero. *Theory of Probability & Its Applications*, 46(1):127–132, 2002. (Cited on page 102.)
- K. Itô. Spectral type of the shift transformation of differential processes with stationary increments. *Transactions of the American Mathematical Society*, 81(2):253–263, 1956. (Cited on pages 30 and 81.)
- J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, 2nd Ed., volume 288. Springer Science & Business Media, 2013. (Cited on pages 36, 48, 65, and 66.)
- A. F. Karr. *Point processes and their statistical applications*, 2nd ed., 1991. (Cited on page 46.)

- J. Kerstan, K. Matthes, and J. Mecke. *Infinitely divisible point processes*. John Wiley & Sons, 1978. (Cited on page 12.)
- J. F. C. Kingman. *Poisson processes*. Wiley Online Library, 1993. (Cited on pages 23, 46, and 48.)
- M. Kohler, A. Krzyżak, and H. Walk. Optimal global rates of convergence for nonparametric regression with unbounded data. *Journal of Statistical Planning and Inference*, 139(4):1286–1296, 2009. (Cited on pages 29, 34, and 86.)
- S. C. Kou. Stochastic networks in nanoscale biophysics: modeling enzymatic reaction of a single protein. *Journal of the American Statistical Association*, 103(483):961–975, 2008. (Cited on page 46.)
- S. C. Kou, X. Sunney Xie, and J. S. Liu. Bayesian analysis of single-molecule experimental data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):469–506, 2005. (Cited on pages 37, 46, 107, and 108.)
- M. Krumin and S. Shoham. Generation of spike trains with controlled auto- and cross-correlation functions. *Neural Computation*, 21(6):1642–1664, 2009. (Cited on page 107.)
- G. Last. Stochastic analysis for poisson processes. In *Stochastic Analysis for Poisson Point Processes*, pages 1–36. Springer, 2016. (Cited on page 12.)
- G. Last and M. D. Penrose. Poisson process fock space representation, chaos expansion and covariance inequalities. *Probability Theory and Related Fields*, 150(3):663–690, 2011. (Cited on pages 16, 32, and 84.)
- G. Last and M. D. Penrose. Lectures on the poisson process, 2016. (Cited on page 12.)
- A. Mandelbaum, A. Sakov, and S. Zeltyn. Empirical analysis of a call center. URL <http://iew3.technion.ac.il/serveng/References/ccdata.pdf>. Technical Report, 2000. (Cited on page 60.)
- A. Mas. Lower bound in regression for functional data by representation of small ball probabilities. *Electronic Journal of Statistics*, 6:1745–1778, 2012. (Cited on pages 27, 28, 39, and 80.)

- J. Mecke. Stationäre zufällige maße auf lokalkompakten abelschen gruppen. *Probability Theory and Related Fields*, 9(1):36–58, 1967. (Cited on page 16.)
- R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1):125–144, 1976. (Cited on page 107.)
- D. Nualart and J. Vives. Anticipative calculus for the poisson process based on the fock space. *Séminaire de Probabilités de Strasbourg*, 24:154–165, 1990. (Cited on pages 30, 31, 81, and 82.)
- Y. Ogata. Statistical models for earthquakes occurrences and residual analysis for point processes. *Journal of the Royal Statistical Society*, 44: 102–107, 1988. (Cited on page 107.)
- F. O’Sullivan. Nonparametric estimation in the cox model. *The Annals of Statistics*, pages 124–145, 1993. (Cited on pages 36, 37, and 108.)
- J. O. Ramsay. *Functional data analysis*. Wiley Online Library, 2006. (Cited on pages 27 and 80.)
- C. R. Rao and I. Chakravarti. Some small sample tests of significance for a poisson distribution. *Biometrics*, 12(3):264–282, 1956. (Cited on pages 25 and 50.)
- D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013. (Cited on pages 52, 62, 63, 65, 66, and 67.)
- P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4(1):3, 2014. (Cited on pages 22, 46, and 47.)
- J. D. Scargle. Studies in astronomical time series analysis. v. bayesian blocks, a new method to analyze structure in photon counting data. *The Astrophysical Journal*, 504(1):405, 1998. (Cited on page 46.)
- H. Schmidli. Lundberg inequalities for a cox model with a piecewise constant intensity. *Journal of Applied Probability*, 33(01):196–210, 1996. (Cited on page 46.)

- G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. SIAM, 2009. (Cited on pages 74, 75, and 76.)
- D. L. Snyder and M. I. Miller. *Random point processes in time and space*. Springer Science & Business Media, 2012. (Cited on page 46.)
- A. B. Tsybakov. Introduction to nonparametric estimation. revised and extended from the 2004 french original. translated by vladimir zaiats, 2009. (Cited on pages 27, 80, 94, 96, 97, and 98.)
- A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. (Cited on pages 25 and 50.)
- S. Watanabe. On discontinuous additive functionals and lévy measures of a markov process. In *Japanese journal of mathematics: transactions and abstracts*, volume 34, pages 53–70. The Mathematical Society of Japan, 1964. (Cited on page 20.)
- T. Zhang and S. C. Kou. Nonparametric inference of doubly stochastic poisson process data via the kernel method. *The annals of applied statistics*, 4(4):1913, 2010. (Cited on pages 36, 46, and 108.)

Résumé

L'objectif principal de cette thèse est de développer des méthodologies statistiques adaptées au traitement de données issues de processus stochastiques et plus précisément de processus de Cox.

Les problématiques étudiées dans cette thèse sont issues des trois domaines statistiques suivants : les tests non paramétriques, l'estimation non paramétrique à noyaux et l'estimation minimax.

Dans un premier temps, nous proposons, dans un cadre fonctionnel, des statistiques de test pour détecter la nature Poissonienne d'un processus de Cox.

Nous étudions ensuite le problème de l'estimation minimax de la régression sur un processus de Poisson ponctuel. En se basant sur la décomposition en chaos d'Itô, nous obtenons des vitesses comparables à celles atteintes pour le cas de la régression Lipschitz en dimension finie.

Enfin, dans le dernier chapitre de cette thèse, nous présentons un estimateur non-paramétrique de l'intensité d'un processus de Cox lorsque celle-ci est une fonction déterministe d'un co-processus.

Mots-clés

Statistique fonctionnelle, processus de Cox, tests statistiques, théorie Martingale, processus ponctuels de Poisson, estimation de la régression, estimation Minimax, estimation de l'intensité, lissage à noyaux.

Abstract

The main purpose of this thesis is to develop statistical methodologies for stochastic processes data and more precisely Cox process data.

The problems considered arise from three different contexts: nonparametric tests, nonparametric kernel estimation and minimax estimation.

We first study the statistical test problem of detecting whether a Cox process is Poisson or not.

Then, we introduce a semiparametric estimate of the regression over a Poisson point process. Using Itô's famous chaos expansion for Poisson functionals, we derive asymptotic minimax properties of our estimator.

Finally, we introduce a nonparametric estimate of the intensity of a Cox process whenever it is a deterministic function of a known coprocess.

Keywords

Functional Statistic, Cox process, test Statistic, Martingale theory, Poisson point process, regression estimate, Minimax estimation, intensity estimation, Kernel smoothing.