Statistical inference across time scales

Céline Duval

UNIVERSITÉ PARIS–EST

THÈSE

pour l'obtention du titre de

DOCTEUR EN MATHÉMATIQUES

présentée et soutenue par

Céline DUVAL

le 7 décembre 2012

INFÉRENCE STATISTIQUE À TRAVERS LES ÉCHELLES

Jury

Cristina BUTUCEA	Université Paris-Est	Examinateur
Fabienne COMTE	Université Paris-Descartes	Rapporteur
Marc HOFFMANN	Université Paris Dauphine	Directeur de thèse
Dominique PICARD	Université Paris-Diderot	Président de Jury
Peter SPREIJ	Universiteit van Amsterdam	Rapporteur
Alexandre TSYBAKOV	Université Pierre et Marie Curie	Examinateur

Remerciements

J'exprime en premier lieu ma plus profonde gratitude à Marc Hoffmann, qui m'a dirigé durant ces trois années. Je lui suis infiniment reconnaissante pour sa disponibilité, ses conseils, son soutien et sa confiance. Il m'a encadré scientifiquement tout en me laissant une autonomie de travail, et m'a confortée dans mon envie de continuer dans la recherche.

Je remercie sincèrement Fabienne Comte et Peter Spreij qui ont rapporté ma thèse. Je leur suis très reconnaissante de leur lecture attentive de ce manuscrit, leurs précieux commentaires et l'intérêt qu'ils ont manifesté pour mon travail.

Je suis très honorée que Cristina Butucea, Dominique Picard et Alexandre Tsybakov aient accepté de faire partie de mon jury.

Je remercie le GIS qui m'a financée durant ces trois ans et le CREST qui m'a accueillie. Je remercie tous les membres du CREST pour leur disponibilité et leurs conseils, et plus particulièrement Pierre Alquier, Nicolas Chopin, Éric Gautier et Judith Rousseau. Je remercie aussi tous les chercheurs que j'ai rencontrés lors de séminaires ou de conférences et qui m'ont conseillée.

Je remercie ensuite tous les jeunes chercheurs du CREST qui ont animé ces trois années. Plus particulièrement les participants des fameux “vendredi bières” ; les membres du F14, Pierre, Christian et Jean-Bernard, mais aussi Robin préposé aux pauses thé et Philippe, Guillaume organisateur de “lundirouges”, Julyan, Giuseppe et Victor-Emmanuel.

Je remercie aussi toutes les personnes qui ont suivi toutes les étapes de cette thèse de l'extérieur, ceux qui ont aussi suivi le chemin de la thèse Alexandra, Carole, Émilie mais surtout Claire et Marie, et les autres Kerstin, Nathalie, Pierre-François, l'équipe du Havre et Charles pour les avoir organisés. Merci à mes proches pour leur soutien et à mes parents qui m'ont donné le goût des mathématiques et de la recherche. Je termine par Nathanaël qui m'a soutenue et encouragée au quotidien.

Résumé

Cette thèse porte sur le problème d'estimation à travers les échelles pour un processus stochastique. Nous étudions comment le choix du pas d'échantillonnage impacte les procédures statistiques. Nous nous intéressons à l'estimation de processus à sauts à partir de l'observation d'une trajectoire discrétisée sur $[0, T]$. Lorsque la longueur de l'intervalle d'observation T va à l'infini, le pas d'échantillonnage tend soit vers 0 (échelle microscopique), soit vers une constante positive (échelle intermédiaire) soit encore vers l'infini (échelle macroscopique). Dans chacun de ces régimes nous supposons que le nombre d'observations tend vers l'infini.

Dans un premier temps le cas particulier d'un processus de Poisson composé d'intensité inconnue avec des sauts symétriques $\{-1, 1\}$ est étudié. Le **Chapitre 2** illustre la notion d'estimation statistique dans les trois échelles définies ci-dessus. Dans ce modèle, on s'intéresse aux propriétés des expériences statistiques. On montre la propriété de Normalité Asymptotique Locale dans les trois échelles microscopique, intermédiaire et macroscopique. L'information de Fisher est alors connue pour chacun de ces régimes. Ensuite nous analysons comment se comporte une procédure d'estimation de l'intensité qui est efficace (de variance minimale) à une échelle donnée lorsqu'on l'applique à des observations venant d'une échelle différente. On regarde l'estimateur de la variation quadratique empirique, qui est efficace dans le régime macroscopique, et on l'utilise sur des données provenant des régimes intermédiaire ou microscopique. Cet estimateur reste efficace dans les échelles microscopiques, mais montre une perte substantielle d'information aux échelles intermédiaires. Une procédure unifiée d'estimation est proposée, efficace dans tous les régimes.

Les **Chapitres 3** et **4** étudient l'estimation nonparamétrique de la densité de saut d'un processus de renouvellement composé dans les régimes microscopiques, lorsque le pas d'échantillonnage tend vers 0. Un estimateur de cette densité utilisant des méthodes d'ondelettes est construit. Il est adaptatif et minimax pour des pas d'échantillonnage qui décroissent en $T^{-\alpha}$, pour $\alpha > 0$. La procédure d'estimation repose sur l'inversion de l'opérateur de composition donnant la loi des incréments comme une transformation non linéaire de la loi des sauts que l'on cherche à estimer. L'opérateur inverse est explicite dans le cas du processus de Poisson composé (**Chapitre 3**), mais n'a pas d'expression analytique pour les processus de renouvellement composés (**Chapitre 4**). Dans ce dernier cas, il est approchée via une technique de point fixe.

Le **Chapitre 5** étudie le problème de perte d'identifiabilité dans les régimes macroscopiques. Si un processus à sauts est observé avec un pas d'échantillonnage grand, certaines approximations limites, telles que l'approximation gaussienne, deviennent valides. Ceci peut

entraîner une perte d'identifiabilité de la loi ayant généré le processus, dès lors que sa structure est plus complexe que celle étudiée dans le **Chapitre 2**. Dans un premier temps un modèle jouet à deux paramètres est considéré. Deux régimes différents émergent de l'étude : un régime où le paramètre n'est plus identifiable et un où il reste identifiable mais où les estimateurs optimaux convergent avec des vitesses plus lentes que les vitesses paramétriques habituelles. De l'étude de cas particulier, nous dérivons des bornes inférieures montrant qu'il n'existe pas d'estimateur convergent pour les processus de Lévy de saut pur ou pour les processus de renouvellement composés dans les régimes macroscopiques tels que le pas d'échantillonnage croît plus vite que racine de T. Enfin nous identifions des régimes macroscopiques où les incréments d'un processus de Poisson composé ne sont pas distinguables de variables aléatoires gaussiennes, et des régimes où il n'existe pas d'estimateur convergent pour les processus de Poisson composés dépendant de trop de paramètres.

Summary

This thesis studies the problem of statistical inference across time scales for a stochastic process. More particularly we study how the choice of the sampling parameter affects statistical procedures. We narrow down to the inference of jump processes from the discrete observation of one trajectory over $[0, T]$. As the length of the observation interval T tends to infinity, the sampling rate either goes to 0 (microscopic scale) or to some positive constant (intermediate scale) or grows to infinity (macroscopic scale). We set in a case where there are infinitely many observations.

First we specialise in a toy model : a compound Poisson process of unknown intensity with symmetric Bernoulli jumps. **Chapter 2** highlights the concept of statistical estimation in the three regimes defined above and the phenomena at stake. We study the properties of the statistical experiments in each regime, we show that the *Local Asymptotic Normality* property holds in every regime (microscopic, intermediate and macroscopic). We also provide the formula of the associated Fisher information in each regime. Then we study how a statistical procedure which is optimal (of minimal variance) at a given scale is affected when we use it on data coming from another scale. We focus on the empirical quadratic variation estimator, it is an optimal procedure at macroscopic scales. We apply it on data coming from intermediate and microscopic regimes. Although the estimator remains efficient at microscopic scales, it shows a substantial loss of information when used on data coming from an intermediate regime. That loss can be explicitly related to the sampling rate. We provide an unified procedure, efficient in all regimes.

Chapters 3 and **4** focus on microscopic regimes, when the sampling rate decreases to 0. The nonparametric estimation of the jump density of a renewal reward process is studied. We propose an adaptive wavelet threshold density estimator. It achieves minimax rates of convergence for sampling rates that vanish polynomially with T, namely in $T^{-\alpha}$ for $\alpha > 0$. The estimation procedure is based on the inversion of the compounding operator in the same spirit as Buchmann and Grübel (2003), which specialise in the study of discrete compound laws. The inverse operator is explicit in the case of a compound Poisson process (see **Chapter 3**), but has no closed form expression for renewal reward processes (see **Chapter 4**). In that latter case the inverse operator is approached with a fixed point technique.

Finally **Chapter 5** studies at which rate identifiability is lost in macroscopic regimes. Indeed when a jump process is observed at an arbitrarily large sampling rate, limit approximations, like Gaussian approximations, become valid and the specificities of the jumps may be lost, as long as the structure of the process is more complex than the one introduced in **Chapter 2**. First we study a toy model depending on a 2-dimensional parameter. We distinguish two different regimes : fast (macroscopic) regimes where all information on the parameter is lost and slow regimes where the parameter remains identifiable but where optimal estimators converge with slower rates than the expected usual parametric ones. From this toy model lower bounds are derived, they ensure that consistent estimation of Lévy processes or renewal reward processes is not possible when the sampling rate grows faster than the square root of T. Finally we identify regimes where an experiment consisting in increments of a compound Poisson process is asymptotically equivalent to an experiment consisting in Gaussian random variables. We also give regimes where there is no consistent estimator for compound Poisson processes depending on too many parameters.

Table des matières

Chapitre 1

Introduction

1.1 Introduction

1.1.1 Motivation

Le problème suivant est souvent rencontré dans la pratique : un phénomène modélisé par un processus à temps continu X est observé sur $[0,T]$ à un pas d'échantillonnage Δ, autrement dit on obtient le vecteur d'observations

$$\boldsymbol{X} = \big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big).$$

Si l'on veut inférer sur la loi de X à partir des observations $\boldsymbol{X}$, il faut d'abord choisir le pas d'échantillonnage Δ avant de pouvoir implémenter une quelconque procédure statistique. Le schéma d'observation doit donc être défini, soulevant ainsi le problème du choix de Δ. On considère un processus à temps continu X de la forme suivante :

$$X_t = X_0 + \sum_{i=1}^{N_t} \varepsilon_i, \quad t \geq 0,$$

où les (ε_i) sont des variables aléatoires indépendantes et identiquement distribuées et N est un processus de comptage supposé indépendant des (ε_i). On note c l'intensité du processus N, c'est la fréquence (moyenne) d'occurrence des sauts. Par exemple X ainsi défini peut être un processus de Poisson composé ou un processus de renouvellement composé (*cf.* Feller [37] ou Sato [91]).

Les propriétés du vecteur des observations $\boldsymbol{X}$ peuvent varier avec la taille du pas d'échantillonnage Δ. En effet, si l'on prend l'exemple du prix d'un actif, si Δ est petit devant $1/c$ (par exemple Δ égal à une microseconde) la plupart des changements de prix seront observés, mais le vecteur d'observations contiendra une grande quantité d'observations redondantes (*cf.* Russell et Engle [89] ou Masoliver *et al.* [71] et la Figure 1.1 ci-dessous). Au contraire si Δ

est grand devant $1/c$ (par exemple Δ égal à une heure, une journée ou une semaine) le contexte est très différent, certaines approximations par des diffusions peuvent s'appliquer (*cf.* Masoliver *et al.* [71] ou Hong et Satchell [55] et la Figure 1.2 ci-dessous). Cependant, utiliser de telles approximations peut masquer certains aspects du processus X sous-jacent. Enfin, si Δ est choisi du même ordre que $1/c$, le processus observé présente trop de sauts pour pouvoir localiser les sauts de X correctement, mais trop peu pour permettre d'appliquer des approximations limites, comme illustré sur la Figure 1.3 ci-dessous.

A première vue, il semble intuitif de choisir Δ du même ordre que $1/c$: assez grand pour limiter le nombre d'observations redondantes, mais suffisamment petit pour que les approximations limites ne soient pas vérifiées, et ainsi préserver dans les données la structure du processus X. Un tel choix devrait établir un équilibre entre la quantité d'information apportées par les données sur X (*e.g.* au sens de l'information de Fisher) et le nombre réel d'observations. Cependant si le choix précédent semble se justifier en prenant le point de vue de l'information, choisir Δ très grand ou très petit présente l'avantage de simplifier le choix des procédures statistiques. Si Δ est grand l'usage d'approximations gaussiennes permet de se ramener à un problème largement étudié et si Δ est petit les sauts sont localisés à partir des observations, on peut appliquer des procédures statistiques du cadre indépendant et identiquement distribué. Ce travail s'intéresse au choix du pas d'échantillonnage Δ. Nous essayons de plus de construire des procédures d'estimations adaptées au pas d'échantillonnage Δ, si dernier est imposé par les données.

1.1.2 Cadre mathématique

Construction des observations. Nous considérons un processus 1-dimensionnel X défini par

$$X_t = X_0 + \sum_{i=1}^{N_t} \varepsilon_i, \quad t \geq 0, \tag{1.1}$$

où les (ε_i) sont des variables aléatoires indépendantes et identiquement distribuées et où N est un processus de comptage indépendant des (ε_i). Par exemple, N peut être un processus de Poisson ou un processus de renouvellement (*cf.* Feller [37] ou Sato [91]). On considère le cadre suivant : le processus X est discrètement observé sur $[0, T]$ aux temps $i\Delta$, pour $\Delta > 0$

$$\big(X_0, X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor \Delta}\big). \tag{1.2}$$

Expérience statistique associée. Dans le contexte habituel d'inférence statistique, on dispose d'une (ou d'une suite d') expérience(s) statistique(s) :

$$\mathcal{E}^n := \big\{\mathbb{P}^n_\varpi, \varpi \in \Pi\big\},$$

où la mesure de probabilité $\mathbb{P}_{\varpi}^{n}$ décrit la loi des observations, n le *paramètre d'information*, qui correspond généralement à la taille de l'échantillon et ϖ le *paramètre d'intérêt*. L'espace des paramètres Π est un sous espace de $\mathbb{R}^d$ ou plus généralement d'un espace fonctionnel.

Nous introduisons une *famille d'expériences statistiques* indexée par le *paramètre d'échelle* Δ, générée par l'observation de (1.2)

$$\left(\mathcal{E}_{\Delta}^{T}\right)_{\Delta>0} = \left\{\mathbb{P}_{\varpi}^{T,\Delta},\ \varpi \in \Pi\right\}_{\Delta>0}. \tag{1.3}$$

Ici le paramètre d'intérêt ϖ caractérise soit la distribution des sauts (ε_i) et/ou la distribution du temps écoulé entre deux sauts, que l'on appellera aussi distribution des interarrivés.

Définition des différents régimes. Nous introduisons les régimes suivant, on y fera référence de façon récurrente dans ce travail.

- Des observations de loi $\mathbb{P}_{\varpi}^{T,\Delta}$ sont dans un régime *microscopique* si $\Delta = \Delta_T \to 0$ lorsque $T \to \infty$.
- Des observations de loi $\mathbb{P}_{\varpi}^{T,\Delta}$ sont dans un régime *intermédiaire* si $\Delta = \Delta_T \to \Delta_\infty > 0$ lorsque $T \to \infty$.
- Des observations de loi $\mathbb{P}_{\varpi}^{T,\Delta}$ sont dans un régime *macroscopique* si $\Delta = \Delta_T \to \infty$ et $T/\Delta_T \to \infty$ lorsque $T \to \infty$.

Intuitivement, dans un régime *microscopique*, lorsque Δ est suffisamment petit, tous les sauts $\left(\varepsilon_i\right)$ devraient être observés, mais de nombreuses données de (1.2) seraient redondantes, comme illustré par la Figure 1.1.

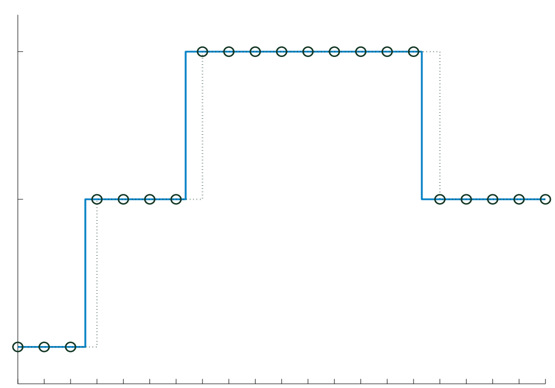

FIGURE 1.1 – Un processus de Poisson composé (trait plein) et le processus observé dans un régime microscopique (pointillé), $\Delta \ll$ intensité.

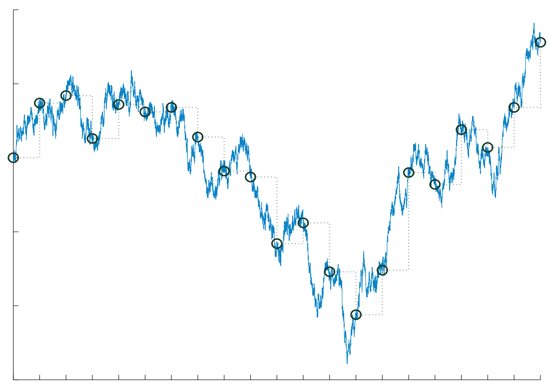

FIGURE 1.2 – Un processus de Poisson composé (trait plein) et le processus observé dans un régime macroscopique (pointillé), $\Delta \gg$ intensité.

Au contraire, si X est observé dans un régime macroscopique, lorsque $\Delta = \Delta_T \to \infty$ sous la contrainte[1] $T/\Delta_T \to \infty$ quand $T \to \infty$, le processus observé se comporte comme celui d'un processus continu (voir la Figure 1.2). En effet, entre deux observations de X consécutives beaucoup de sauts se sont produits rendant les approximations limites valides (en un sens à préciser). Enfin dans un régime intermédiaire, le processus observé présente de trop nombreux sauts pour les localiser précisément à partir des observations (1.2), mais pas assez nombreux pour vérifier une approximation limite (voir la Figure 1.3).

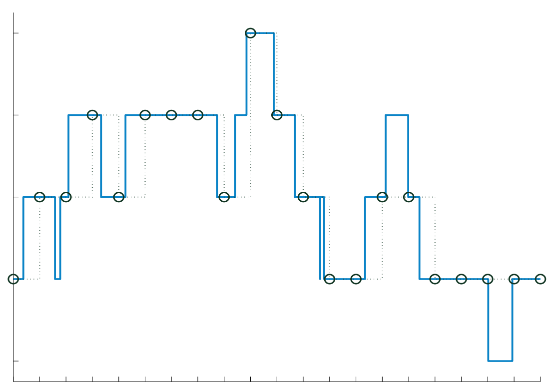

FIGURE 1.3 – Un processus de Poisson composé (trait plein) et le processus observé dans un régime intermédiaire (pointillé), $\Delta \sim$ intensité.

1. Cette contrainte assure qu'il y a asymptotiquement une infinité d'observations.

Remarque 1. *Le paramètre d'échelle, ou pas d'échantillonnage,* Δ *modifie les propriétés de* $\mathbb{P}_{\varpi}^{T,\Delta}$ *: dans le régime macroscopique* $\mathbb{P}_{\varpi}^{T,\Delta}$ *semble être absolument continu par rapport à la mesure de Lebesgue, contrairement aux régimes microscopique et intermédiaire. Mais l'interprétation du paramètre d'intérêt* ϖ *demeure le même à toutes les échelles, il caractérise la loi du processus* X*.*

Remarque 2. *Faisons maintenant les remarques suivantes :*
- *Dans la littérature, on parle de données hautes fréquences, ce terme est (le plus souvent) utilisé pour désigner un régime où* $\Delta \to 0$ *et* $T = 1$ *(voir par exemple Genon-Catalot et al. [40], Genon-Catalot et Jacod [41], Jacod [58] ou Sørensen [98]). Ce cadre est différent du régime microscopique introduit et étudié dans ce travail.*
- *De nombreux travaux s'intéressent aux régimes microscopiques et intermédiaires comme définis ci-dessus ; voir par exemple Kessler et Sørensen [65] ou Gobet et al. [46] pour l'étude de processus de diffusion et Neumann et Reiß [83] ou Comte et Genon-Catalot [22, 23, 24, 25] pour l'étude de processus de Lévy (pour plus de références, voir la Section 1.1.4).*
- *Le régime macroscopique, comme défini ci-dessus, est lui beaucoup moins étudié que les régimes microscopique et intermédiaire.*

Compromis entre taille de l'échantillon et taux d'information. Dans ce contexte, la taille de l'échantillon est toujours égale à $\lfloor T\Delta^{-1}\rfloor$, mais le *taux d'information* lui semble dépendre du choix de l'échelle d'observation (microscopique, intermédiaire ou macroscopique). Par taux d'information, nous désignons la quantité d'information réellement transmise par les observations, *e.g.* au sens de l'information de Fisher. L'information est contenue dans les sauts et leur localisation ; le cadre idéal correspond au cas où chacun des N_T sauts sont observés. Le taux maximum d'information est donc N_T, qui est en général de l'ordre de T (à une constante multiplicative près), pour T grand. Le taux d'information correspond donc au minimum entre T (ou de façon équivalente N_T) et la taille de l'échantillon $\lfloor T\Delta^{-1}\rfloor$, ce qui correspond à un cas ou le choix du pas d'échantillonnage ne permet pas d'identifier tous les sauts de X. Cette discussion est résumée dans le Tableau suivant.

Régime	Taille de l'échantillon	Taux d'information
Régime microscopique $\Delta \to 0$	$\lfloor T\Delta^{-1}\rfloor \gg T$	$N_T \approx T$
Régime intermédiaire $\Delta \to \Delta_\infty \in (0,\infty)$	$\lfloor T\Delta^{-1}\rfloor \approx T$	$\lfloor T\Delta^{-1}\rfloor \approx T \approx N_T$
Régime macroscopique $\Delta \to \infty$	$\lfloor T\Delta^{-1}\rfloor \ll T$	$\lfloor T\Delta^{-1}\rfloor$

1.1.3 Problématique

D'après les remarques précédentes, d'un côté il semble plus confortable de travailler dans un régime microscopique ou macroscopique. En effet, à la limite microscopique le processus X est continûment observé, tous les sauts et leurs tailles sont exactement connus. Les outils usuels valables dans le cadre indépendant et identiquement distribué s'appliquent. Ensuite, dans la limite macroscopique la loi des incréments devrait être une loi stable, déduite des théorèmes limites généraux (*cf.* Gnedenko et Kolmogorov [44]). L'expérience macroscopique ne devrait alors dépendre que d'un nombre restreint de paramètres. Mais si ces paramètres sont en trop petit nombre, relativement à la taille du paramètre ϖ à estimer, il devient peut-être impossible de retrouver le paramètre d'intérêt ϖ. D'un autre côté, il semble préférable d'échantillonner le processus X à la fréquence de ses sauts (dans un régime intermédiaire) : pas trop petit pour limiter le nombre d'observations inutiles (incréments nuls) mais pas non plus trop grand pour éviter l'attraction des lois limites stables.

Dans cette thèse, nous étudions et proposons quelques explications sur comment le choix du paramètre d'échelle Δ affecte les procédures statistiques. Nous tentons de répondre aux questions suivantes.

i) Le paramètre d'intérêt ϖ peut-il être identifié à partir des observations (1.2) dans chaque régime ; microscopique, intermédiaire ou macroscopique ?

ii) Si la réponse à la question **i)** est positive, est-il possible de construire une procédure d'estimation qui soit optimale dans tous les régimes ?

iii) Est-il possible d'avoir une unique procédure statistique, qui s'adapte à l'échelle d'observation ?

iv) Que se passe-t-il si l'on utilise une procédure d'estimation, qui est optimale dans une échelle microscopique (ou macroscopique), sur des données venant d'une échelle intermédiaire mais où Δ est petit (ou grand) ?

1.1.4 Liens avec d'autres travaux

Étudier des processus échantillonnés de façon discrète n'est pas nouveau, nous citons ici quelques travaux ayant un cadre statistique similaire. La liste proposée ci-dessous n'est pas exhaustive.

Un cadre important ou la discrétisation est utilisée concerne l'estimation de processus de diffusion. Les processus de diffusion ont des trajectoires continues et sont donc relativement éloignés des processus à sauts considérés dans ce travail. De nombreux articles se placent dans un cadre dit haute fréquence où la longueur de l'intervalle d'observation est fixée ($T = 1$ par exemple), ce cadre est différent de celui considéré ici (voir aussi la Remarque 2). Parmi les nombreux auteurs on peut citer les travaux de Genon-Catalot *et al.* [40], Genon-Catalot et Jacod [41], Jacod [58] ou Sørensen [98]. Plus proche du cadre de cette thèse, l'estimation nonparamétrique d'un processus de diffusion aux échelles intermédiaires ($\Delta > 0$

et $T \to \infty$) est étudié dans Kessler et Sørensen [65], Gobet *et al.* [46] ou Reiß [86]. Dans les échelles microscopiques ($\Delta \to 0$ et $T \to \infty$) l'estimation paramétrique pour des processus de diffusion est analysée dans Yosida [107], Kessler [64], Gobet [45], Masuda [72] or Jacod [59] et l'estimation nonparamétrique dans Pham [85], Hoffmann [54], Comte *et al.* [20, 21] ou van Es *et al.* [104, 105].

L'estimation d'un processus de Lévy [2] à partir de l'observation discrète d'une de ses trajectoire est plus fortement liée à la problématique de ce travail. Dans un premier temps l'estimation de la loi de saut d'un processus de Poisson composé, comme étudiée dans le Chapitre 3 dans les échelles microscopiques, à été considérée dans un régime intermédiaire ($\Delta = 1$ et $T \to \infty$) par Buchmann et Grübel [16, 17] et van Es *et al.* [103]. Buchmann et Grübel se sont intéressés plus particulièrement à l'estimation de lois discrètes contrairement à van Es *et al.* qui proposent un estimateur nonparamétrique. Plus de détails sur les différences entre ces travaux et le Chapitre 3 sont donnés Section 1.2.2.

L'estimation nonparamétrique de processus de Lévy de saut pur dans les échelles microscopiques a été étudiée de façon approfondie par Bec et Lacour [9], Comte et Genon-Catalot [22, 23] et Figueroa-López [38]. Comte et Genon-Catalot [23] se placent dans un cadre plus général où les observations sont irrégulièrement espacées et bruitées. Dans les échelles intermédiaires Comte et Genon-Catalot [24] fournissent un estimateur adaptatif et minimax de la densité de Lévy. Les processus de Poisson composés étudiés dans le Chapitre 3 sont des processus de Lévy de saut pur particuliers, dans la Section 1.2.2 nous donnons plus de détails sur ces résultats. Enfin pour les processus de Lévy (ayant une partie continue et des sauts), des estimateurs nonparamétriques du triplet caractéristique dans les régimes microscopiques sont proposés dans Shimizu [97], Gugushvili [48] ou Comte et Genon-Catalot [25], et dans les régimes intermédiaires dans Jongbloed *et al.* [61] ou Neumann et Reiß [83].

Enfin le problème de l'estimation d'un processus de renouvellement ou d'un processus de renouvellement composé à partir de l'observation discrète d'une de ses trajectoire n'a, à la connaissance de l'auteur, pas été étudié. Des estimateurs pour des processus de renouvellement basés sur l'observation de trajectoires indépendantes sur un intervalle de temps fixe sont donnés dans Vardi [102], Gill et Keiding [43] ou Guédon et Cocozza-Thivent [47].

1.2 Résultats principaux

1.2.1 Inférence statistique à travers les échelles : cas paramétrique

Le Chapitre 2 étudie les questions **i)** à **iv)** sur un modèle jouet dans le but d'illustrer et d'éclairer le concept d'inférence statistique à travers les échelles.

2. Pour les processus de Lévy présentant des sauts, l'hypothèse $T \to \infty$ est nécessaire pour garantir qu'il y ait asymptotiquement une infinité de sauts observés.

Cadre statistique. Considérons le modèle jouet suivant : soit un processus de Poisson composé X défini par

$$X_t = X_0 + \sum_{i=1}^{N_t} \varepsilon_i, \quad t \geq 0,$$

où les (ε_i) sont des variables aléatoires indépendantes et identiquement distribuées selon $\mathbb{P}(\epsilon_i = 1) = \mathbb{P}(\epsilon_i = -1) = 1/2$ et où N est un processus de Poisson indépendant des variables (ε_i) d'intensité inconnue $\vartheta \in \Theta = (0, \infty)$. Dans ce modèle jouet le paramètre d'intérêt est l'intensité inconnue ϑ du processus de Poisson : le cadre est paramétrique. Supposons que l'on dispose d'observations discrètes de X sur l'intervalle $[0, T]$ aux instants $i\Delta$

$$\boldsymbol{X} = \big(X_{i\Delta} - X_{(i-1)\Delta}, i = 1, \ldots, \lfloor T\Delta^{-1} \rfloor\big).$$

On définit la famille d'expériences suivante comme dans (1.3)

$$\big(\mathcal{E}_{\Delta}^{T}\big)_{\Delta > 0} = \big(\mathbb{Z}^{\lfloor T\Delta^{-1} \rfloor}, \mathcal{P}(\mathbb{Z}^{\lfloor T\Delta^{-1} \rfloor}), \{\mathbb{P}_{\vartheta}^{T,\Delta}, \vartheta \in \Theta\}\big),$$

où $\mathbb{P}_{\vartheta}^{T,\Delta}$ désigne la loi de $\boldsymbol{X}$.

Comportement dans les différents régimes. Nous considérons les trois régimes introduits dans la Section 1.1.2. Intuitivement, si X est observé microscopiquement, lorsque $\Delta = \Delta_T \to 0$ avec $T \to \infty$, essentiellement tous les sauts du processus N sont localisés, ce sont ces sauts qui véhiculent toute l'information sur le paramètre ϑ. A l'inverse si X est observé macroscopiquement, c'est-à-dire si $\Delta_T \to \infty$ sous la contrainte que $T/\Delta_T \to \infty$ quand $T \to \infty$, on a l'approximation gaussienne suivante

$$X_{i\Delta_T} - X_{(i-1)\Delta_T} \approx \sqrt{\vartheta}\big(W_{i\Delta_T} - W_{(i-1)\Delta_T}\big),$$

où W est un processus de Wiener. Estimer ϑ revient à estimer la variance de variables gaussiennes. Enfin, si X est observé dans un régime intermédiaire tel que $\Delta_T \to \Delta_\infty$ lorsque $T \to \infty$, le processus observé présente trop de sauts pour pouvoir les localiser à partir de $\boldsymbol{X}$ et trop peu pour que l'approximation gaussienne précédente soit justifiée. A partir de ces remarques, on aimerait pouvoir dire que si Δ est suffisamment petit alors $\mathbb{P}_{\vartheta}^{T,\Delta}$ apporte la même information sur ϑ que lorsque tous les sauts de X sont observés et si Δ est suffisamment grand que $\mathbb{P}_{\vartheta}^{T,\Delta}$ véhicule la même information que l'observation discrète d'un mouvement Brownien de variance ϑ.

Résultat 1. *La propriété de Normalité Asymptotique Locale*[3] *est vérifiée dans chacun des régimes. L'expression de l'information de Fisher $I_{T,\Delta_T}(\vartheta)$ est explicitement connue : aux échelles intermédiaires lorsque $\Delta_T \to \Delta_\infty$ avec $T \to \infty$ on a*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,\Delta_\infty}(\vartheta) = T\Delta_\infty\Big(\mathbb{E}_{\vartheta}^{T,\Delta_\infty}\big[\big(h_{\Delta_\infty}(\vartheta, X_{\Delta_\infty}) + (\vartheta\Delta_\infty)^{-1}|X_{\Delta_\infty}| - 1\big)^2\big]\Big),$$

3. Pour plus de détails sur la propriété de Normalité Asymptotique Locale voir le livre de Ibragimov et Hasminskii [57].

où

$$h_{\Delta_\infty}(\vartheta, k) = \frac{\mathcal{I}_{|k|+1}(\vartheta\Delta_\infty)}{\mathcal{I}_{|k|}(\vartheta\Delta_\infty)},$$

et où pour $x \in \mathbb{R}$ *et* $\nu \in \mathbb{N}$,

$$\mathcal{I}_\nu(x) = \sum_{m\geq 0} \frac{(x/2)^{2m+\nu}}{m!(\nu+m)!}$$

est la fonction de Bessel modifiée de première espèce, aux échelles microscopiques lorsque $\Delta_T \to 0$ *avec* $T \to \infty$ *on a*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,0}(\vartheta) := \lim_{\Delta\to 0} I_{T,\Delta}(\vartheta) = \frac{T}{\vartheta}$$

et dans les échelles macroscopiques lorsque $\Delta_T \to \infty$ *avec* $T \to \infty$ *on a*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,\infty}(\vartheta) := \frac{T\Delta_T^{-1}}{2\vartheta^2}.$$

Ce premier résultat permet de comprendre comment les expériences $\left(\mathcal{E}_\Delta^T\right)_{\Delta>0}$ se comportent dans chacun des régimes. Il nous donne aussi la valeur de l'information de Fisher qui dépend continûment du paramètre d'échelle Δ. De plus dans les échelles microscopiques l'information de Fisher est la même que pour une expérience Poissonienne d'intensité ϑ et dans les régimes macroscopiques elle coïncide avec l'information de Fisher d'une expérience gaussienne de variance ϑ. La valeur de l'information de Fisher aux régimes intermédiaires n'a pas d'interprétation aussi évidente. La connaissance de cette quantité permet, dans chaque régime, de déterminer si une procédure d'estimation de ϑ est ou non optimale.

Estimation de l'intensité ϑ. Nous étudions ici les conséquences de l'utilisation d'une procédure statistique qui est optimale à une échelle donnée sur des données venant d'une échelle différente. On considère plus particulièrement l'estimateur de la variation quadratique empirique

$$\widehat{\vartheta}_T^{QV} = \frac{1}{T} \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \left(X_{i\Delta_T} - X_{(i-1)\Delta_T}\right)^2.$$

C'est l'estimateur du maximum de vraisemblance d'un modèle où l'on observe des variables gaussiennes indépendantes centrées et de même variance (régime limite macroscopique). De plus puisque la loi composante de X est à valeur dans $\{-1, 1\}$, cet estimateur imite l'estimateur du maximum de vraisemblance du régime limite microscopique : le nombre de sauts divisé par la longueur de l'intervalle d'observation.

Résultat 2. *L'estimateur* $\widehat{\vartheta}_T^{QV}$ *est asymptotiquement efficace, i.e. il est asymptotiquement normal et sa variance asymptotique est (équivalente à) l'inverse de l'information de Fisher, dans les régimes microscopique et macroscopique. Aux échelles intermédiaires il converge à la bonne vitesse mais ne parvient pas à atteindre la variance optimale. Cette perte, allant jusqu'à 23%, peut être reliée à la valeur* Δ_∞.

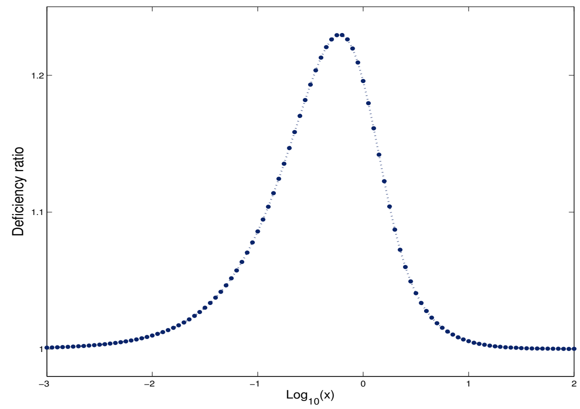

FIGURE 1.4 – Rapport de l'information de Fisher information $I_{T,\Delta_T}(\vartheta)$ et de l'inverse de la variance de $\widehat{\vartheta}_T^{QV}$.

La Figure 1.4 illustre le Résultat 2, on y retrouve que l'estimateur $\widehat{\vartheta}_T^{QV}$ est efficace à la fois dans les régimes microscopique et macroscopique, mais qu'il n'attaint l'information de Fisher dans aucuns des régimes intermédiaires. Cette procédure permet alors de donner une réponse positive aux questions **i)** et **ii)**, mais pas à la question **iii)**. L'estimateur de la variation quadratique empirique est consistent dans chacun des trois régimes mais n'est pas optimal dans le régime intermédiaire. A partir de cette estimateur, il est possible de construire un estimateur corrigé, en utilisant une méthode de correction à un pas (*cf.* van der Vaart [101]), qui répond à la question **iii)** de façon positive.

Résultat 3. *L'estimateur corrigé à un pas, construit à partir de $\widehat{\vartheta}_T^{QV}$ est efficace dans toutes les échelles.*

Si l'on interprète le paramètre ϑ comme l'intensité de trading dans les échelles microscopiques et comme une volatilité dans les régimes macroscopiques, le Résultat 2 nous dit que si le processus X est échantillonné à la même fréquence que les changements de prix, ce qui est habituellement le cas en pratique, alors l'estimateur $\widehat{\vartheta}_T^{QV}$ n'est pas efficace. L'estimateur corrigé à un pas devrait lui être préféré.

1.2.2 Estimation nonparamétrique minimax dans le régime microscopique

Les Chapitres 3 et 4 s'intéressent au régime microscopique lorsque $\Delta = \Delta_T \to 0$ avec $T \to \infty$. Plus particulièrement on y estime la densité f des sauts (ε_i) (*cf.* (1.1)), où N est soit un processus de Poisson d'intensité λ inconnue (Chapitre 3) soit un processus de renouvellement dont la densité τ des interarrivées (temps écoulé entre les sauts) est inconnue (Chapitre

4). L'estimateur proposé utilise des méthodes d'estimation de la densité par ondelettes (*cf.* Cohen [19], Donoho *et al.* [30] ou Kerkyacharian et Picard [63]).

Procédure d'estimation. Dans les échelles microscopiques de nombreuses observations sont redondantes (voir la Figure 1.1) : seuls les incréments non nuls de X apportent de l'information sur la densité f que l'on cherche à estimer : on ne s'intéresse qu'aux incréments observés non nuls. Deux sortes de difficultés surviennent :

- Le nombre de données utilisé pour l'estimation est aléatoire : c'est le nombre d'incréments non nuls observés.
- La densité des incréments non nuls n'est pas f mais la transformation non linéaire

$$\mathbf{P}_\Delta[f](x) = \sum_{m=1}^{\infty} \mathbb{P}\big(R_\Delta = m \big| N_\Delta \neq 0\big) f^{\star m}(x), \qquad \text{for } x \in \mathbb{R}, \tag{1.4}$$

où $\star$ est le produit de convolution et $f^{\star m} = f \star \ldots \star f$, m fois.

Remarque 3. *Lorsque Δ est suffisamment petit, la plupart des incréments non nuls correspondent effectivement à des réalisations de f. On a la décomposition suivante de l'opérateur*

$$\mathbf{P}_\Delta[f] = f + r(\Delta),$$

où $r(\Delta)$ est un reste déterministe qui vérifie $r(\Delta) = O(\Delta)$.

Nous allons aussi regarder l'effet d'une approximation microscopique qui consiste à ignorer le reste $r(\Delta)$, cela revient à appliquer les estimateurs nonparamétriques habituels aux incréments non nuls de X comme si tous étaient des réalisations de f.

Pour estimer f on inverse l'opérateur précédent, il s'agit d'un problème inverse bien posé, et d'utiliser

$$f = \mathbf{P}_\Delta^{-1}\big[\mathbf{P}_\Delta[f]\big].$$

On introduit une procédure d'estimation en deux étapes.

Étape 1 L'inverse de l'opérateur $\mathbf{P}_\Delta$ défini par (1.4) est calculée. On prend le développement de Taylor à l'ordre K de cette inverse, où K est choisi de manière adéquate (*cf.* le Résultat 4 ci-dessous). Ce développement fait intervenir les quantités $\mathbf{P}_\Delta[f]^{\star m}$, $m = 1, \ldots, K+1$.

Étape 2 On construit des estimateurs adaptatifs des densités $\mathbf{P}_\Delta[f]^{\star m}$, pour m allant de 1 à $K+1$, que l'on injecte dans la formule de l'inverse obtenue à l'**Étape 1**.

On appelle l'estimateur ainsi obtenu *estimateur corrigé à l'ordre K*. Il est adaptatif. Nous étudions sa vitesse de convergence pour la perte L_p, $p \geq 1$ pour des densités f qui appartiennent à une classe de régularité Besov (f est de régularité s mesurée pour la norme L_π, $\pi > 0$).

Résultat 4. *On obtient la borne supérieure suivante sur la vitesse de convergence de l'estimateur corrigé à l'ordre K, pour la perte L_p, $p \geq 1$, uniformément sur toute boule de Besov (à un facteur logarithmique près)*

$$\max\{T^{-\alpha(s,\pi,p)}, \Delta_T^{K+1}\},$$

où $\alpha(s,\pi,p)$ dépend de la régularité de la densité f.

Une borne inférieure. Le résultat précédent ne donne qu'une borne supérieure de la vitesse de convergence de l'estimateur, cependant un borne inférieure peut être facilement déduite de l'expérience limite microscopique. En effet, à la limite microscopique, le processus X est continûment observé : on dispose d'exactement N_T réalisations indépendantes de f pour l'estimation, où N_T est le processus de Poisson ou de renouvellement au temps T et est indépendant des sauts (ϵ_i). Alors, si on conditionne par N_T, on obtient une borne inférieure pour l'estimation de f en $N_T^{-\alpha(s,\pi,p)}$ (voir par exemple Donoho *et al.* [30] ou Härdle *et al.* [49]). Pour finir, on utilise la propriété de concentration de N_T autour de T si T est suffisamment grand (à un facteur multiplicatif près et avec quelques hypothèses sur la loi des interarrivées). Finalement on retrouve une borne inférieure de l'ordre de $T^{-\alpha(s,\pi,p)}$.

Vitesse de converge de l'estimateur. Puisque $\alpha(s,\pi,p) \leq 1/2$, à partir du Résultat 4, s'il existe K_0 tel que $T\Delta_T^{2K_0+2} \leq 1$, alors l'estimateur corrigé à l'ordre K_0 converge à la vitesse $T^{-\alpha(s,\pi,p)}$, qui est minimax. On a donc construit un estimateur qui permet de répondre positivement aux questions **i)** à **iii)** dans les régimes microscopiques : on a une unique procédure qui est adaptative et minimax. Si on considère l'estimateur corrigé à l'ordre 0, ce qui revient à faire une approximation microscopique (*i.e.* à ignorer le reste $r(\Delta)$), l'estimateur obtenu est minimax si Δ converge rapidement vers 0, si $T\Delta_T^2 = O(1)$. Alors la réponse à la question **iv)** est positive si Δ est plus rapide que $1/\sqrt{T}$ quand $T \to \infty$.

Illustration numérique. Ci-dessous on donne un résultat d'implémentation de l'estimateur corrigé à l'ordre K pour différentes valeurs de K, dans le cas de l'estimation de la loi composante d'un processus de Poisson composé. On choisit $T = 10000$ et $\Delta = 0,1$, ce qui correspond à un cas où $T\Delta^2$ est grand et $T\Delta^4 = 1$. Selon le Résultat 4 l'estimateur corrigé à l'ordre 1 devrait être minimax. On estime la loi composante qui est un mélange entre une loi normale $\mathcal{N}(0,1)$ et une loi de Laplace de paramètres 1 (position) et 0,1 (échelle). Pour évaluer la performance de notre estimateur on le compare à un oracle : l'estimateur que l'on aurait implémenté dans la situation idéalisée où tous les sauts du processus de Poisson sont observés. On obtient, pour la perte L_2, les résultats reproduits dans le tableau suivant

Estimateur	Oracle	$K=0$	$K=1$	$K=2$	$K=3$
Perte L_2 ($\times 10^{-2}$)	0,11	0,18	0,13	0,13	0,13
Écart type ($\times 10^{-5}$)	0,35	0,44	0,44	0,44	0,44

Ces résultats sont en accord avec la théorie, lorsque K augmente, jusqu'à $K = 1$, la perte L_2 diminue, mais augmenter K au delà n'améliore plus la perte de l'estimateur.

Quelques remarques sur la procédure. On donne ici quelques informations supplémentaires sur l'estimation en deux étapes détaillée ci-dessus. L'étape 1 s'inspire des méthodes dites de *decompounding* introduites dans Buchmann et Grübel [16], qui étudient l'estimation de lois de saut discrètes pour des processus de Poisson composés dans un régime intermédiaire ($\Delta = 1$). L'inverse de l'opérateur (1.4) admet une forme fermée dans le Chapitre 3 mais pas dans le Chapitre 4, puisque la forme de la loi des interarrivées n'est généralement pas connue pour un processus de renouvellement. Dans ce dernier cas, on utilise une méthode de point fixe pour approcher l'inverse de l'opérateur (1.4) : f est l'unique point fixe d'un opérateur contractant. Une approximation de l'inverse s'écrit de la façon suivante

$$\mathbf{P}_\Delta^{-1}[g] \approx \sum_{m=1}^{K+1} l_m(\Delta) g^{\star m},$$

où les coefficients $(l_m(\Delta))$ sont explicites et de l'ordre de Δ^{m-1} si T est grand. Le reste de l'approximation de $\mathbf{P}_\Delta^{-1}$ donnée par la formule précédente est de l'ordre de Δ^{K+1} si T est grand.

La difficulté majeure de l'étape 2 de la procédure, lorsqu'on estime les densités $\mathbf{P}_\Delta[f]^{\star m}$ où $m = 1, \ldots, K+1$, est due au fait que l'on travaille avec un nombre aléatoire de données. Pour résoudre cela, on utilise la concentration de ce nombre autour d'une valeur déterministe de l'ordre de T (à un facteur multiplicatif près).

Remarque 4. *Un processus de Poisson composé est un processus de renouvellement composé, les résultats du Chapitre 4 englobent ceux du Chapitre 3. Dans le Chapitre 4 la construction de l'estimateur est basée sur une méthode de point fixe alors que les formules sont explicites dans le Chapitre 3.*

Lien avec d'autres travaux. Le Chapitre 3 étudie dans les régimes microscopiques l'estimation de la loi composante d'un processus de Poisson composé en utilisant une méthode semblable à celle de Buchmann et Grübel [16, 17]. Buchmann et Grübel s'intéressent à l'estimation de lois discrètes dans un régime intermédiaire ($\Delta = 1$). Dans le cas de lois à densité, ils proposent un estimateur de la fonction de répartition, supposant l'intensité connue, mais n'étudient pas la vitesse de convergence de cette estimateur. Toujours dans le régime intermédiaire $\Delta = 1$ et en supposant l'intensité du processus de Poisson connue, van Es *et al.* [103] estiment la densité de la loi composante, leur estimateur se base sur la formule de Lévy Kintchine. Ils se concentrent aussi les incréments non nuls, mais ils traitent le fait que nombre de données est aléatoire de façon différente : ils introduisent un temps aléatoire qui leur permet de savoir exactement quand n (déterministe) incréments non nuls ont été observés. Ils ne donnent pas de vitesse de convergence de leur estimateur.

Dans le cas de l'estimation de processus de Lévy de saut pur dans les régimes microscopiques, des estimateurs de la densité de Lévy existent. Ils sont minimax si le pas d'échantillonnage vérifie la contrainte $T\Delta_T^2 = O(1)$ (*cf.* Bec et Lacour [9], Comte et Genon-Catalot [22, 25] et Figueroa-López [38]), cette contrainte est relâchée par notre estimateur dans le cas des processus de Poisson composés. Estimer la mesure de Lévy est bien plus général qu'estimer la densité de saut d'un processus de Poisson composé, mais très différent de l'estimation de la densité de saut d'un processus de renouvellement composé.

1.2.3 Perte d'identifiabilité aux échelles macroscopiques

Dans le dernier Chapitre 5, on s'intéresse aux régimes macroscopiques tels que $\Delta = \Delta_T \to \infty$ et $T/\Delta_T \to \infty$ lorsque $T \to \infty$. Un exemple typique d'un tel régime est Δ_T de l'ordre de T^γ pour γ in $(0,1)$. On considère la classe $\mathcal{F}_0$ des processus de Poisson composés. On suppose que le paramètre d'intérêt a au moins deux composantes pour ne pas se retrouver dans le même cadre que le Chapitre 2. On étudie les questions **i)**, **ii)** et **iv)**.

Dans un tel régimes, si la loi des sauts est centrée et a une variance finie, le théorème central limite suggère l'approximation gaussienne suivante

$$X_{i\Delta} - X_{(i-1)\Delta} \approx \sigma^2 (W_{i\Delta} - W_{(i-1)\Delta}),$$

où W est un mouvement Brownien standard et σ dépend de l'intensité du processus de Poisson et de la variance des sauts. Si cette approximation est vérifiée, lorsque Δ est grand on ne devrait être capable d'estimer seulement σ : l'intensité du pocessus de Poisson et les spécificités de la loi des sauts devraient être perdues. C'est ce que l'on appelle dans la suite perte d'identifiabilité.

Étude d'un modèle jouet. Attaquer le problème de la perte d'identifiabilité dans les régimes macroscopiques directement est délicat, c'est pourquoi on introduit et étudie dans un premier temps un modèle jouet. Le processus étudié dans ce modèle est la somme d'un processus de Poisson composé et d'un drift, il dépend d'un paramètre de dimension 2 (il n'appartient pas à $\mathcal{F}_0$). On identifie deux types de régimes macroscopiques :

- Un régime où le paramètre ne peux pas être estimé de façon consistante : γ appartient à $(1/2, 1)$. L'information de Fisher associée dégénère en une matrice de rang 1. Ce résultat apporte une réponse négative aux questions **i)** et **ii)**.
- Un régime où le paramètre reste identifiable (γ appartient à $(0, 1/2)$), apportant une réponse positive à la question **ii)**. Mais les estimateurs optimaux convergent à des vitesses en $\sqrt{T\Delta_T^{-2}}$, plus lentes que les vitesses paramétriques $\sqrt{\lfloor T\Delta^{-1}\rfloor}$ habituellement rencontrées.

À la limite $\gamma = 1/2$, le paramètre ne peut pas être estimé de façon consistante, entraînant une réponse négative à la question **ii)**. L'information de Fisher a une valeur propre bornée.

Perte d'identifiabilité pour la classe $\mathcal{F}_0$. De l'étude du modèle jouet précédent, on déduit une borne inférieure qui mène au résultat suivant.

Résultat 5. *Si $\Delta = \Delta_T$ est tel que $T/\Delta_T \to \infty$ et $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ lorsque $T \to \infty$ alors il n'est pas possible de construire un estimateur convergent pour les éléments de $\mathcal{F}_0$.*

Ce résultat répond négativement à la question **ii)** pour les régimes tels que γ appartient à $(1/2, 1)$. Ensuite on s'intéresse à un résultat d'équivalence asymptotique (*cf.* Le Cam et Yang [68] ou Nussbaum [82]). Avant de donner ce résultat, il est nécéssaire d'introduire quelques notations. Soit $K \geq 3$, on considère un processus de Poisson composé Y d'intensité λ et dont la loi composante est centrée et a ses K premiers moments finis, que l'on note $m_K = (0, m_{2,K}, \ldots, m_{K,K})$. Le paramètre d'intérêt est $\rho = (\lambda, m_K) \in \Sigma_K$ où Σ_K est un sous espace compact de $(0, \infty) \times \mathbb{R}^{K-1}$. On suppose que Y est observé sur $[0, T]$ aux temps $i\Delta$ pour $\Delta > 0$,

$$\boldsymbol{Y} = (Y_{i\Delta} - Y_{(i-1)\Delta}, i = 1, \ldots, \lfloor T\Delta^{-1} \rfloor),$$

et on définit la famille d'expériences

$$\mathcal{Y}_K^{\Delta} = \big(\mathbb{R}^{\lfloor T\Delta^{-1} \rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1} \rfloor}), \{\mathbb{P}_\rho^{T,\Delta}, \rho \in \Sigma_K\}\big),$$

où $\mathbb{P}_\rho^{T,\Delta}$ est la loi de $\boldsymbol{Y}$. Soit $a > 0$, on considère le processus de Poisson composé Z d'intensité $a\lambda$ et de loi composante centrée dont les K premiers moments sont finis et donnés par $g_a(m_K) := (0, \frac{m_{2,K}}{a}, \ldots, \frac{m_{K,K}}{a})$. Le paramètre d'intérêt est alors $h_a(\rho) = (a\lambda, g_a(m_K)) \in \Sigma_K$. On suppose que Z est aussi observé de façon discrète sur $[0, T]$ aux temps $i\Delta$ pour $\Delta > 0$,

$$\boldsymbol{Z} = (Z_{i\Delta} - Z_{(i-1)\Delta}, i = 1, \ldots, \lfloor T\Delta^{-1} \rfloor),$$

et on définit la famille d'expériences

$$\mathcal{Z}_K^{\Delta} = \big(\mathbb{R}^{\lfloor T\Delta^{-1} \rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1} \rfloor}), \{\mathbb{Q}_{h_a(\rho)}^{T,\Delta}, \rho \in \Sigma_K\}\big),$$

où $\mathbb{Q}_{h_a(\rho)}^{T,\Delta}$ caractérise la loi de $\boldsymbol{Z}$. On peut alors énoncer le résultat suivant :

Résultat 6.
1. *Si $K = 3$, $m_{3,3} = 0$ et $\Delta = \Delta_T$ vérifie $T/\Delta_T \to \infty$ et $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ lorsque $T \to \infty$, alors $\mathcal{Y}_3^{\Delta}$ est asymptotiquement équivalente à une expérience gaussienne qui consiste en $\lfloor T\Delta_T^{-1} \rfloor$ variables aléatoires gaussiennes centrées et de variance $\lambda m_{2,3}$.*
2. *Soit $K \geq 3$, si on a $T\Delta_T^{-(K+1)/2} = o((\log(T/\Delta_T))^{-1/4})$, alors les expériences $\mathcal{Y}_K^{\Delta_T}$ et $\mathcal{Z}_K^{\Delta_T}$ sont asymptotiquement équivalentes.*

D'après le Résultat 6.1. si γ appartient à $(1/2, 1)$ et si la loi composante à un moment d'ordre 3 nul, alors la réponse à la question **iv)** est positive. Les incréments du processus de Poisson composé ne sont pas distinguables d'incréments d'un processus gaussien. Le Résultat 6.2. assure que dans les régimes où γ est dans $(1/2, 1)$ mais où la loi composante à un moment d'ordre 3 non nul ($K = 3$) ou dans les régimes où γ est dans $(0, 1/2)$ ($K > 3$), l'expérience $\mathcal{Y}_K^\Delta$ n'est pas gaussienne à la limite, entraînant une réponse négative à la question **iv)**. Dans ces régimes, le Résultat 6.2. assure aussi que la réponse à la question **i)** est négative. En effet d'après ce résultat il est possible de construire deux processus de Poisson composés différents mais que l'on ne peut pas distinguer à partir de leurs incréments. Dans ce cas on devrait pourvoir déduire une borne inférieure assurant qu'il n'existe pas d'estimateur convergent pour les processus de Poisson composés dépendant de trop nombreux paramètres (relativement à la vitesse de convergence de Δ_T).

Remarque 5. *Puisque un processus de Poisson composé est à la fois un processus de renouvellement composé et un processus de Lévy, certains résultats du Chapitre 5 se généralisent immédiatement à ces familles, c'est par exemple le cas du Résultat 5.*

1.3 Rapide résumé et remarques supplémentaires

Dans ce travail nous avons tenté de motiver l'intérêt de l'estimation statistique à travers les échelles, en montrant pour différents scénarios les effets du choix du pas d'échantillonnage sur les procédures statistiques. Le Chapitre 2, dans un modèle jouet paramétrique, offre une vision assez complète des différents phénomènes à prendre en compte. Dans ce modèle jouet il apparaît une continuité des expériences statistiques et de l'information associée, lorsque l'on passe d'un régime microscopique à un régime intermédiaire jusqu'à un régime macroscopique. On observe aussi un phénomène un peu inattendu, l'estimateur classique de la variation quadratique empirique, s'il est efficace aux régimes microscopique et macroscopique présente une perte d'information aux échelles intermédiaires.

Continuité des régimes microscopiques aux régimes intermédiaires. Les Chapitres 3 et 4 s'intéressent à l'estimation de la densité des sauts pour des processus de renouvellement composés dans les régimes microscopiques. Ils proposent une procédure adaptative et minimax basée sur une méthode d'inversion par méthode de point fixe. Cette procédure donne aussi des indices sur ce que peut être une procédure d'estimation de la densité de saut qui soit minimax aux régimes intermédiaires. En effet la méthode d'estimation proposée reste valide pour des régimes intermédiaires tels que le pas d'échantillonnage limite Δ_∞ est suffisamment petit. Cependant nous n'avons pas étudié les vitesses de convergence de notre estimateur dans de tels régimes, mais devrait être possible d'observer à nouveau une continuité lorsque l'on passe d'un régime microscopique à un régime intermédiaire.

Les régimes intermédiaires. Pour les processus de Lévy (qui contiennent les processus de Poisson composés), estimer de façon minimax et adaptative la mesure de Lévy à partir des observations (1.2) est toujours possible aux échelles intermédiaires telles que $\Delta_T \to \Delta_\infty \in (0, \infty)$ quand $T \to \infty$. Une procédure d'estimation possible utilise la formule de Lévy Kintchine (voir par exemple Neumann et Reiß [83] ou Comte et Genon-Catalot [24]). Pour les processus de renouvellement composés, il existe un équivalent de la formule de Lévy Kintchine : la transformée de Fourier Laplace (voir par exemple Metzler et Klafter [76, 77] ou Vlahos *et al.* [106]), qui permet de caractériser et la loi des interarrivées et celle des sauts. Cependant on ne peut pas estimer cette quantité à partir de nos observations (1.2), en tout cas sans hypothèse supplémentaire. Si la loi des interarrivées est connue, alors on peut identifier la loi des sauts à partir de (1.2). En effet l'opérateur $\mathbf{P}_\Delta$ défini par (1.4) et donnant la loi des observations comme une transformation non linéaire de la loi des sauts, peut être inversé (appliquer le théorème d'inversion locale pour les séries formelles à la transformée de Fourier de $\mathbf{P}_\Delta$). Cependant, cela ne permet pas d'obtenir une expression fermée de l'inverse, adopter une procédure d'estimation analogue à celle proposée dans les Chapitres 3 et 4 semble impossible.

Des régimes intermédiaires aux régimes macroscopiques. Si pour les processus de Lévy l'estimation adaptative et minimax de la mesure de Lévy à partir de (1.2) est possible dans tous les régimes intermédiaires (voir ci-dessus), cela n'est plus vrai aux échelles macroscopiques. Le dernier Chapitre se concentre sur l'étude des régimes macroscopiques et parvient à identifier des cas où il n'existe pas d'estimateur convergent pour des processus de renouvellement composés ou des processus de Lévy. Mais la compréhension de la transition d'un régime intermédiaire à un régime macroscopique n'est pas aussi claire que le passage d'un régime microscopique à un régime intermédiaire. Déjà si l'estimation d'un processus de Lévy dans un régime intermédiaire est bien comprise, ce n'est pas le cas pour un processus de renouvellement composé. Ensuite, même pour le cas plus simple d'un processus de Poisson composé, il n'est pas clair qu'il soit possible d'estimer un paramètre de grande taille, même lorsque Δ va lentement à l'infini (par exemple logarithmiquement en T). Étudier comment le Résultat 6.2. est modifié si on laisse la taille du paramètre à estimer grandir pourrait apporter quelques éléments de réponse.

Quelques références appliquées. Dans la littérature plus appliquée, les processus de renouvellement composés sont souvent appelés marches aléatoires à temps continu (CTRW). Les CTRW ont été introduites par Montroll et Weiss [79], et sont très populaires en physique. C'est un modèle microscopique assez général qui permet à la limite macroscopique d'obtenir toutes sortes de diffusions (voir aussi Montroll et Scher [80]). Un processus de renouvellement composé est parfois appelé CTRW découplée puisque on suppose qu'il y a indépendance entre les sauts et le temps qui les sépare, ce n'est pas le cas pour les CTRW couplées.

Les CTRW aux échelle microscopiques. Les CTRW sont des processus à sauts relativement généraux pour modéliser des phénomènes observés à des échelles petites, comme lorsque l'on veut modéliser la pluie en un point donné comme dans Rodriguez-Irtube *et al.* [88]. Ces processus sont particulièrement adaptés lorsque la modélisation par des processus de Poisson est limité par la propriété d'oubli de mémoire de ces derniers. Ils sont utilisés en finance pour modéliser un prix (*cf.* Gerber et Shiu [42]) ou par les compagnies d'assurance (*cf.* Embrechts *et al.* [32] ou Scalas [94]).

Les CTRW aux échelle macroscopiques. Aux grandes échelles, les CTRW permettent d'obtenir différents comportements diffusifs (*cf.* Metzler et Klafter [76, 77]). Cette propriétés est largement utilisée en physique pour décrire le mouvement d'une particule (voir par exemple Anderson et Meerschaert [4] en hydrologie, Lawrence *et al.* [67] pour l'activité solaire, Berkowitz *et al.* [10] en géologie ou Buroni *et al.* [18]), ou encore en biologie pour modéliser, par exemple, comment se propagent certaines cellules cancéreuses ([35, 36]) ou des lipides (Jeon *et al.* [60]).

Extensions possibles.

Schéma d'échantillonnage non réguliers. Dans beaucoup de domaines appliqués, comme en biologie ou en sociologie, il peut être difficile d'obtenir des observations régulièrement espacées en raison de la difficulté de récolter les données. Il serait donc pertinent de s'intéresser à des schémas d'échantillonnages irréguliers ou aléatoires. L'estimation de processus de diffusions aléatoirement observés à été étudiée par Jacod [58] ou Ait-Sahalia et Mykland [1] et par Comte et Genon-Catalot [23] dans le cas de processus de Lévy.

Estimer la loi des interarrivées. Dans ce travail nous nous sommes surtout intéressé à l'estimation de la loi des sauts, mais il peut être parfois tout aussi important de connaître la loi du temps qui sépare deux événements. C'est par exemple le cas lorsque l'on étudie les tremblements de terre, connaître leur amplitude est aussi important que savoir à quel moment ils se produisent (*cf.* Helmstetter et Sornette [52], Alvarez [3] ou Garavaglia [39]). Ce problème a été considéré dans le cas de processus de renouvellement dans Vardi [102], Gill et Keiding [43] ou Guédon et Cocozza-Thivent [47]. Cependant leur cadre statistique est différent de celui défini ici, ils disposent de trajectoires indépendantes observées sur un intervalle de temps fixé.

Les CTRW couplées. L'hypothèse d'indépendance entre les interarrivées et les sauts peut devenir dans certains cas très restrictive. Il existe des phénomènes où pour pouvoir observer un saut conséquent, il faut avoir attendu un temps plus long que pour un petit saut : ce problème est abordé dans Masoliver *et al.* [71]. Ceci explique l'introduction de CTRW couplées en finance et en économie (*cf.* Scalas [92, 94], Scalas *et al.* [93], Meerschaert et Scalas [75], Schumer *et al.* [95] ou Jurlewicz *et al.* [62]).

Les processus de Hawkes. Les processus de Hawkes sont aussi des processus à sauts, ils ont d'abord été décrits par Hawkes [50, 51] (voir aussi Daley et Vere-Jones [28]). Ce sont des processus auto excités dont les intensités de chaques composantes sont liées. Ces processus sont déjà utilisés en finance dans la modélisation du carnet d'ordre (*cf.* Hewlett [53] et Errais *et al.* [33]) ou pour modéliser le bruit de microstructure (*cf.* Bacry *et al.* [5, 6]), mais aussi pour étudier les intéractions en biologie (*cf.* Reynaud-Bouret et Schbath [87]). Dans sa thèse, Al Dayri [29] utilise des processus de Hawkes pour modéliser les arrivées aléatoires des échanges en finance à partir de pas d'échantillonnages "fins" (échelles microscopiques) ou "grossiers" (échelles intermédiaires), ayant aussi une approche à différentes échelles.

Liste des travaux ayant contribués à la rédaction de la thèse :

- C. Duval et M. Hoffmann, *Statistical inference across time scales,* paru dans *Electronic Journal of Statistics*, 2011 (Chapitre 2).
- C. Duval, *Adaptive wavelet estimation of a compound Poisson process*, soumis (Chapitre 3).
- C. Duval, *Nonparametric estimation of a renewal reward process from discrete data*, soumis (Chapitre 4).
- C. Duval, *Identifiability loss at large scales of discretely observed jump processes*, travail en cours (Chapitre 5).

Abstract

We consider a compound Poisson process with symmetric Bernoulli jumps, observed at times $i\Delta$ for $i = 0, 1, \ldots$ over $[0, T]$, for different sizes of $\Delta = \Delta_T$ relative to T in the limit $T \to \infty$. We quantify the smooth statistical transition from a microscopic Poissonian regime (when $\Delta_T \to 0$) to a macroscopic Gaussian regime (when $\Delta_T \to \infty$). The classical quadratic variation estimator is efficient for estimating the intensity of the Poisson process in both microscopic and macroscopic scales but surprisingly, it shows a substantial loss of information in the intermediate scale $\Delta_T \to \Delta_\infty \in (0, \infty)$. This loss can be explicitly related to Δ_∞. We provide an estimator that is efficient simultaneously in microscopic, intermediate and macroscopic regimes. We discuss the implications of these findings beyond this idealised framework.

Keywords : Discretely observed random process, LAN property, Information loss.

Note

Chapter 2 is based on a paper published in *Electronic Journal of statistics*. It is a joint work with Marc Hoffmann. We are grateful to Sylvain Delattre for helpful discussions and comments. The suggestions of a referee helped to improve a former version of the paper.

Chapitre 2

Inference across time scales in a parametric case

2.1 Introduction

2.1.1 Motivation

Consider a 1-dimensional random process (X_t) defined by

$$X_t = X_0 + \sum_{i=1}^{N_t} \varepsilon_i, \quad t \geq 0, \tag{2.1}$$

where the $\varepsilon_i \in \{-1, 1\}$ are independent, identically distributed with

$$\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = \frac{1}{2},$$

and independent of the standard homogeneous Poisson process (N_t) with intensity $\vartheta \in \Theta = (0, \infty)$. Suppose we have discrete data over $[0, T]$ at times $i\Delta$. This means that we observe

$$\boldsymbol{X} = \big(X_0, X_\Delta, \dots, X_{\lfloor T\Delta^{-1} \rfloor \Delta}\big), \tag{2.2}$$

and we obtain a statistical experiment by taking $\mathbb{P}_\vartheta$ as the law of $\boldsymbol{X}$ defined by (2.2) when (X_t) is governed by (2.1). This toy model is central to several application fields, *e.g.* financial econometrics or traffic networks (see the discussion in Section 2.3 and the references therein). Moreover, it already contains several interesting properties that enlight a tentative concept of statistical inference across scales. This is the topic of the Chapter.

On the one hand, if we observe (X_t) *microscopically*, that is if $\Delta = \Delta_T \to 0$ as $T \to \infty$, then asymptotically, we can – essentially – locate the jumps of (N_t) that convey all the relevant information about the parameter ϑ.

In that case, $\boldsymbol{X}$ is "close" to the continuous path $(X_t, t \in [0,T])$.

On the other hand, if we observe (X_t) *macroscopically*, that is if $\Delta_T \to \infty$ under the constraint [1] $T/\Delta_T \to \infty$, we have a completely different picture : the diffusive approximation

$$X_{i\Delta_T} - X_{(i-1)\Delta_T} \approx \sqrt{\vartheta}\big(W_{i\Delta_T} - W_{(i-1)\Delta_T}\big), \tag{2.3}$$

becomes valid, where (W_t) is a standard Wiener process. We elaborate in Appendix the approximation (2.3). Inference on ϑ essentially transfers into a Gaussian variance estimation problem ;

in that case, the state space rather becomes $\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor+1}$. Finally if we observe (X_t) in the *intermediate scale* $0 < \liminf \Delta_T \leq \limsup \Delta_T < \infty$, we observe a process presenting too many jumps to be located accurately from the data, and too few to verify the Gaussian approximation (2.3). Therefore, depending on the scale parameter Δ_T, the state space may vary, and it has an impact on the underlying random scenarios $\mathbb{P}_\vartheta$, although the interpretation of the *parameter of interest* ϑ remains the same at all scales. What we have is rather a family of experiments

$$\mathcal{E}^{T,\Delta} = \{\mathbb{P}_\vartheta^{T,\Delta}, \vartheta \in \Theta\}, \tag{2.4}$$

where $\mathbb{P}_\vartheta^{T,\Delta}$ denotes the law of $\boldsymbol{X}$ given by (2.2) and these experiments $\mathcal{E}^{T,\Delta}$ may exhibit different behaviours at different scales Δ. Heuristically, we would like to state that in the *microscopic scale* $\Delta_T \to 0$, the measure $\mathbb{P}_\vartheta^{T,\Delta_T}$ conveys the same information about ϑ as the law of

$$(N_t, t \in [0,T]), \tag{2.5}$$

that is if the jump times of (X_t) were observed. On the other side, in the *macroscopic scale* $\Delta_T \to \infty$ with $T/\Delta_T \to \infty$, the measure $\mathbb{P}_\vartheta^{T,\Delta_T}$ shall convey the same information about ϑ as the law of

$$\big(0, \sqrt{\vartheta}W_{\Delta_T}, \ldots, \sqrt{\vartheta}W_{\lfloor T\Delta_T^{-1}\rfloor\Delta}\big), \tag{2.6}$$

that is if the data were drawn as a Brownian diffusion with variance ϑ.

The following questions naturally arise :

i) How does the model formulated in (2.4) interpolate – from a statistical inference perspective – from microscopic (when $\Delta = \Delta_T \to 0$) to macroscopic scales (when $\Delta = \Delta_T \to \infty$) ? In particular, how do intrinsic statistical information indices (such as the Fisher information) evolve as $\Delta = \Delta_T$ varies ?

ii) Is there any nontrivial phenomenon that occurs in the intermediate regime

$$0 < \liminf \Delta_T \leq \limsup \Delta_T < \infty?$$

1. This condition ensures that asymptotically infinitely many observations are recorded in the limit $T \to \infty$.

iii) Given **i)** and **ii)**, if a statistical procedure is optimal on a given scale Δ, how does it perform on another scale ? Is it possible to construct a single procedure that automatically adapts to each scale Δ, in the sense that it is efficient simultaneously over different time scales ?

2.1.2 Main results

In this Chapter, we systematically explore questions i), ii) and iii) in the simplified context of the experiments $\mathcal{E}^{T,\Delta}$ built upon the continuous time random walks model (2.1) for transparency. Some extensions to non-homogeneous compound Poisson processes are given, and the generalisation to a more general compound law is also discussed. As for i), we prove in Theorems 1, 2 and 3 that the LAN condition (Locally Asymptotic Normality [2]) holds for all scales Δ. This means that $\mathbb{P}_{\vartheta}^{T,\Delta}$ can be approximated – in appropriate sense – by the law of a Gaussian shift. We derive in particular the Fisher information of $\mathcal{E}^{T,\Delta}$ and observe that it smoothly depends on the scale Δ. We shall see that the answer to ii) is positive. More precisely, we first prove in Theorem 4 that the normalised quadratic variation estimator

$$\widehat{\vartheta}_T^{QV} = \frac{1}{T}\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor}\big(X_{i\Delta_T} - X_{(i-1)\Delta_T}\big)^2$$

is asymptotically efficient – it is asymptotically normal and its asymptotic variance is equivalent to the inverse of the Fisher information – in both microscopic and macroscopic regimes. In the microscopic regime, it stems from the fact that the approximation

$$\widehat{\vartheta}_T^{QV} \approx \frac{1}{T}\sum_{0\leq t\leq T}\big(X_t - X_{t^-}\big)^2 = \frac{N_T}{T}$$

becomes valid, as the jumps are ± 1, and the efficiency is then a consequence of N_T/T being the maximum likelihood estimator in the approximation experiment (2.5). In the macroscopic regime, thanks to the diffusive approximation (2.3) we have

$$\widehat{\vartheta}_T^{QV} \approx \frac{1}{T}\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor}\big(\sqrt{\vartheta}(W_{i\Delta_T} - W_{(i-1)\Delta_T})\big)^2,$$

which is precisely the maximum likelihood estimator in the macroscopic approximation experiment (2.6). Surprisingly, $\widehat{\vartheta}_T^{QV}$ fails to be efficient when

$$\Delta_T \to \Delta_\infty \in (0,\infty). \tag{2.7}$$

2. Recommended references are the textbooks [57] and [101], but we recall some definitions in Section 2.2.2 for sake of completeness.

More precisely, we show in Theorem 5 that, although rate optimal, $\widehat{\vartheta}_T^{QV}$ misses the optimal variance by a non-negligible factor, depending on Δ_∞, that can reach up to 23%. This phenomenon is due to the fact that in the intermediate regime (2.7), the process (X_t) is sampled at a rate which has the same order as the intensity of its jumps. On the one hand, $(X_{i\Delta_T} - X_{(i-1)\Delta_T})^2$ gives no accurate information whereas a jump has occurred or not during the period $[(i-1)\Delta_T, i\Delta_T]$, contrary to the case $\Delta_T \to 0$. On the other hand, there are not enough jumps to validate the approximation of $X_{i\Delta_T} - X_{(i-1)\Delta_T}$ by a Gaussian random variable, contrary to the case $\Delta_T \to \infty$. Finally, we construct in Theorem 6 a one-step correction of $\widehat{\vartheta}_T^{QV}$ that provides an estimator efficient in all scales, giving a positive answer to iii).

This Chapter is organised as follows. We first propose in Section 2.2.1 a canonical framework for different time scales by considering the family of experiments $\left(\mathcal{E}^{T,\Delta_T}\right)_{T>0}$. The way the scale parameter depends on T defines the terms *microscopic, intermediate* and *macroscopic* scales rigorously.

Specialising to model (2.1) for transparency, the results about the structure of the corresponding $\left(\mathcal{E}^{T,\Delta_T}\right)_{T>0}$ are stated in Section 2.2.2. We show in Theorems 1, 2 and 3 that the LAN (Local Asymptotic Normality) property holds simultaneously over all scales and provides an explicit expression for the Fisher information. The proof follows the classical route of [57] and boils down to obtaining accurate approximations of the distribution

$$f_{\Delta_T}(\vartheta, k) = \mathbb{P}_\vartheta^{T,\Delta_T}\big(X_{i\Delta_T} - X_{(i-1)\Delta_T} = k\big), \quad k \in \mathbb{Z},$$

in the limit $\Delta_T \to 0$ or ∞. Note that $f_{\Delta_T}(\vartheta, k)$ does not depend on i since (X_t) has stationary increments. However explicit, the intricate form of $f_{\Delta_T}(\vartheta, k)$ requires asymptotic expansions of modified Bessel functions of the first kind. In the macroscopic regime however, we were not able to obtain such expansions. We take another route instead, proving directly the asymptotic equivalence in the Le Cam sense, a stronger result at the expense of requiring a rate of convergence of Δ_T to ∞, presumably superfluous. We show in Theorems 4 and 5 of Section 2.2.3 that the quadratic variation estimator $\widehat{\vartheta}_T^{QV}$ is rate optimal and efficient in both microscopic and macroscopic regimes, but *not* in the intermediate scales (2.7). This negative result is however appended with the construction of an estimator based on a one-step correction of $\widehat{\vartheta}_T^{QV}$ that is efficient over all scales (Theorem 6). Moreover this estimator has the advantage of being computationally implementable, contrary to the theoretical optimal maximum likelihood estimator. Section 2.3 gives some extensions in the case of a non-homogeneous compound Poisson process (Theorem 7) and addresses the generalisation to more general compound laws. The comparison to related works on estimating Lévy processes from discrete data is also discussed. Section 2.4 is devoted to the proofs.

2.2 Statement of the results

2.2.1 Building up statistical experiments across time scales

Let $T > 0$ and $\Delta > 0$ be such that $\Delta \leq T$. On a rich enough probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we observe the process (X_t) defined in (2.1) at frequency Δ^{-1} over the period $[0, T]$. Thus we observe $\boldsymbol{X}$ defined in (2.2), and with no loss of generality [3], we take $X_0 = 0$. We obtain a family of statistical experiments

$$\mathcal{E}^{T,\Delta} := \big(\mathbb{Z}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{Z}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{P}_\vartheta^{T,\Delta}, \vartheta \in \Theta\}\big),$$

where $\mathbb{P}_\vartheta^{T,\Delta}$ denotes the law of $\boldsymbol{X}$ when (X_t) has the form (2.1), and $\Theta \subseteq (0, \infty)$ is a parameter set with non empty interior. The experiment $\mathcal{E}^{T,\Delta}$ is dominated by the counting measure μ_T on $\mathbb{Z}^{\lfloor \Delta^{-1}T\rfloor}$.

Abusing notation slightly, we may [4] (and will) identify $\boldsymbol{X}$ with the canonical observation in $\mathcal{E}^{T,\Delta}$.

under $\mathbb{P}_\vartheta^{T,\Delta}$, we obtain the following expression for the likelihood

$$\frac{d\mathbb{P}_\vartheta^{T,\Delta}}{d\mu_T}(\boldsymbol{X}) = \prod_{i=1}^{\lfloor T\Delta^{-1}\rfloor} f_\Delta(\vartheta, X_{i\Delta} - X_{(i-1)\Delta}),$$

where we have set, for $k \in \mathbb{Z}$,

$$f_\Delta(\vartheta, k) := \mathbb{P}_\vartheta^{T,\Delta}\big(X_{i\Delta} - X_{(i-1)\Delta} = k\big) = \mathbb{P}_\vartheta^{T,\Delta}\big(X_\Delta = k\big). \tag{2.8}$$

We shall repeatedly use the terms *microscopic, intermediate* and *macroscopic* scale (or regime). In order to define these terms precisely, we let $\Delta = \Delta_T$ depend on T with $0 < \Delta_T \leq T$, and we adopt the following terminology.

Definition 1. *The sub-family of experiments* $(\mathcal{E}^{T,\Delta_T})_{T>0}$ *is said to be*

1. *On a microscopic scale (or regime) if* $\Delta_T \to 0$ *as* $T \to \infty$.
2. *On an intermediate scale (or regime) if* $\Delta_T \to \Delta_\infty$ *as* $T \to \infty$, *for some* $\Delta_\infty \in (0, \infty)$.
3. *On a macroscopic scale (or regime) if* $\Delta_T \to \infty$ *and* $T/\Delta_T \to \infty$ *as* $T \to \infty$.

2.2.2 The regularity of $(\mathcal{E}^{T,\Delta_T})_{T>0}$ across time scales

Let us recall [5] that the family of experiments $(\mathcal{E}^{T,\Delta_T})_{T>0}$ satisfies the Local Asymptotic Normality (LAN) property at point $\vartheta \in \Theta$ with normalisation $I_{T,\Delta_T}(\vartheta) > 0$ if, for every

3. By assuming $X_0 = 0$, the first data point does not give information about the parameter ϑ. If only asymptotic properties of the statistical model are studied, which is always the case here, it has no effect.
4. By taking for instance $\Omega = \mathbb{Z}^{\lfloor T\Delta_T^{-1}\rfloor}$.
5. See for instance the textbooks [57] or [101].

$v \in \mathbb{R}$ such that $\vartheta + vI_{T,\Delta_T}(\vartheta)^{1/2} \in \Theta$, the following decomposition holds

$$\frac{d\mathbb{P}^{T,\Delta_T}_{\vartheta + vI_{T,\Delta_T}(\vartheta)^{-1/2}}}{d\mathbb{P}^{T,\Delta_T}_{\vartheta}}(\boldsymbol{X}) = \exp\big(v\,\xi_T - \frac{1}{2}v^2 + r_T\big), \tag{2.9}$$

where

$$\xi_T \to \mathcal{N}(0,1) \quad \text{in distribution under } \mathbb{P}^{T,\Delta_T}_{\vartheta} \quad \text{as } T \to \infty \tag{2.10}$$

and

$$r_T \to 0 \quad \text{in probability under } \mathbb{P}^{T,\Delta_T}_{\vartheta} \quad \text{as } T \to \infty. \tag{2.11}$$

If (2.9), (2.10) and (2.11) hold, we informally say that $(\mathcal{E}^{T,\Delta_T})_{T>0}$ is *regular with information* $I_{T,\Delta_T}(\vartheta)$. This means that locally around ϑ, the law of $\boldsymbol{X}$ can be approximated by the law of a Gaussian shift experiment, where one observes a single random variable

$$\boldsymbol{Y} = \vartheta + I_{T,\Delta_T}(\vartheta)^{-1/2}\xi_T,$$

with ξ_T being approximately distributed as a standard Gaussian random variable under $\mathbb{P}^{T,\Delta_T}_{\vartheta}$ as $T \to \infty$. In particular, $I_{T,\Delta_T}(\vartheta)$ is the Fisher information of the Gaussian shift experiment : the optimal rate of convergence for recovering ϑ up to constants from $\boldsymbol{X}$ is the same as the one obtained from $\boldsymbol{Y}$ and is given by $I_{T,\Delta_T}(\vartheta)^{-1/2}$ provided $I_{T,\Delta_T}(\vartheta) \to \infty$ as $T \to \infty$. Note also that if the convergence of the remainder term $r_T = r_T(v)$ in (2.11) holds locally uniformly in v, *i.e.* the supremum of $r_T(v)$, for v in any compact subset of $(0,\infty)$, goes to 0 in probability, then $I_{T,\Delta_T}(\vartheta)$ can be replaced by any function $J_{T,\Delta_T}(\vartheta)$ such that

$$J_{T,\Delta_T}(\vartheta) \sim I_{T,\Delta_T}(\vartheta) \quad \text{as } T \to \infty$$

without affecting the LAN property. Hereafter, the symbol $\sim$ means asymptotic equivalence up to constants. Our first result states the LAN property for the experiment $\big(\mathcal{E}^{T,\Delta}\big)_{T>0}$ on every scale $\Delta \in (0,\infty)$.

Theorem 1 (The intermediate regime)**.** *Assume $\Delta_T \to \Delta_\infty \in (0,\infty)$ as $T \to \infty$. Then the family $(\mathcal{E}^{T,\Delta_T})_{T>0}$ is regular and we have*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,\Delta_\infty}(\vartheta) = T\Delta_\infty\Big(\mathbb{E}^{T,\Delta_\infty}_{\vartheta}\big[\big(h_{\Delta_\infty}(\vartheta, X_{\Delta_\infty}) + (\vartheta\Delta_\infty)^{-1}|X_{\Delta_\infty}| - 1\big)^2\big]\Big),$$

with

$$h_{\Delta_\infty}(\vartheta, k) = \frac{\mathcal{I}_{|k|+1}(\vartheta\Delta_\infty)}{\mathcal{I}_{|k|}(\vartheta\Delta_\infty)},$$

where, for $x \in \mathbb{R}$ and $\nu \in \mathbb{N}$,

$$\mathcal{I}_\nu(x) = \sum_{m\geq 0} \frac{(x/2)^{2m+\nu}}{m!(\nu+m)!}$$

denotes the modified Bessel function of the first kind.

Remark 1. *By taking $\Delta_T = \Delta_\infty \in (0,\infty)$ constant, we include the case of a fixed Δ, therefore the same regularity result holds for $\big(\mathcal{E}^{T,\Delta}\big)_{T>0}$. An inspection of the proof of Theorem 1 reveals that the mapping $\Delta \rightsquigarrow I_{T,\Delta}(\vartheta)$ is continuous over $(0,\infty)$.*

Our next result shows that formally we can let $\Delta_\infty \to 0$ in the expression of $I_{T,\Delta_\infty}(\vartheta)$ given by Theorem 1 in the microscopic case. Moreover we obtain a simplified expression for the information rate.

Theorem 2 (The microscopic case)**.** *Assume $\Delta_T \to 0$ as $T \to \infty$. Then the family $(\mathcal{E}^{T,\Delta_T})_{T>0}$ is regular and we have*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,0}(\vartheta) := \lim_{\Delta\to 0} I_{T,\Delta}(\vartheta) = \frac{T}{\vartheta}.$$

The macroscopic case is a bit more involved. In that case, we cannot formally let $\Delta_\infty \to \infty$ in the expression of $I_{T,\Delta_\infty}(\vartheta)$ given by Theorem 1. However we have the following simplification.

Theorem 3 (The macroscopic case)**.** *Assume $\Delta_T \to \infty$, $T/\Delta_T \to \infty$ as $T \to \infty$ and $T/\Delta_T^{1+\frac{1}{4}} = o\big((\log(T/\Delta_T))^{-\frac{1}{4}}\big)$. Then the family $(\mathcal{E}^{T,\Delta_T})_{T>0}$ is regular and we have*

$$I_{T,\Delta_T}(\vartheta) \sim I_{T,\infty}(\vartheta) := \frac{T\Delta_T^{-1}}{2\vartheta^2}.$$

The condition $T/\Delta_T^{1+\frac{1}{4}} = o\big((\log(T/\Delta_T))^{-\frac{1}{4}}\big)$ is technical but quite stringent ; it is satisfied for example if $\Delta_T = T^\beta$ with $\frac{4}{5} < \beta < 1$ and stems from our method of proof, see Section 2.4.3. It is presumably superfluous, but we do not know how to relax it.

2.2.3 The distortion of information across time scales

On each scale $\Delta > 0$, let us introduce the empirical quadratic variation estimator

$$\widehat{\vartheta}^{QV}_{T,\Delta} = \frac{1}{T}\sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor}(X_{i\Delta} - X_{(i-1)\Delta})^2$$

that mimics the behaviour of the maximum likelihood estimator in both macroscopic and microscopic regimes (see Section 2.1). More precisely, we have the following asymptotic normality result.

Theorem 4. *Let $\Delta = \Delta_T > 0$ be such that $T/\Delta_T \to \infty$ as $T \to \infty$. We have*

$$\widehat{\vartheta}^{QV}_{T,\Delta_T} = \vartheta + \big(I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}\big)^{1/2}\xi_T,$$

where $\xi_T \to \mathcal{N}(0,1)$ in distribution under $\mathbb{P}^{T,\Delta}_\vartheta$, and $I_{T,0}(\vartheta)$ and $I_{T,\infty}(\vartheta)$ are the information of the microscopic and macroscopic experiments given in Theorems 2 and 3 respectively.

On a microscopic scale $\Delta_T \to 0$, we have

$$\frac{I_{T,\infty}(\vartheta)^{-1}}{I_{T,0}(\vartheta)^{-1}} \to 0 \quad \text{as} \quad T \to \infty.$$

On a macroscopic scale $\Delta_T \to \infty$ with $T/\Delta_T \to \infty$, we have on the contrary

$$\frac{I_{T,0}(\vartheta)^{-1}}{I_{T,\infty}(\vartheta)^{-1}} \to 0 \quad \text{as} \quad T \to \infty.$$

As a consequence, we readily see that $\widehat{\vartheta}^{QV}_{T,\Delta_T}$ is asymptotically normal and that its asymptotic variance is equivalent to $I_{T,0}(\vartheta)^{-1}$ on a microscopic scale and to $I_{T,\infty}(\vartheta)^{-1}$ on a macroscopic scale. At a heuristical level, this phenomenon can be explained directly by the form of the empirical quadratic variation estimator, as we already did in Section 2.1. At intermediate scales however, this is no longer true.

Theorem 5 (Loss of efficiency in the intermediate regime)**.** *Assume that*

$$0 < \liminf \Delta_T \leq \limsup \Delta_T \leq \frac{1}{4\vartheta}. \tag{2.12}$$

Then

$$\liminf_{T\to\infty} \frac{I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}}{I_{T,\Delta_T}(\vartheta)^{-1}} > 1,$$

where $I_{T,\Delta_T}(\vartheta)$ is defined in Theorem 1.

Remark 2. *For technical reasons, we are unable to prove that Theorem 5 remains valid beyond the restriction* $\limsup \Delta_T \leq 1/(4\vartheta)$*. Numerical simulations suggest however that Theorem 5 is valid whenever* $\limsup \Delta_T < \infty$*, see Figure 2.2.3.*

Let us denote by

$$\mathcal{R}_T\big(\widehat{\vartheta}^{QV}_{T,\Delta_T}, \vartheta\big) = \mathbb{E}^{T,\Delta_T}_{\vartheta}\big[\big(\widehat{\vartheta}^{QV}_{T,\Delta_T} - \vartheta\big)^2\big]$$

the squared error loss of the quadratic variation estimator. By Theorems 1, 2 and 3, the family $\big(\mathcal{E}^{T,\Delta_T}\big)_{T>0}$ is regular in all regimes and we may apply the classical minimax lower bound of Hajek, see for instance Theorem 12.1 in [57] : we have, for any $\vartheta_0 \in \Theta$ and $\delta > 0$ such that $[\vartheta_0 - \delta, \vartheta_0 + \delta] \subset \Theta$

$$\liminf_{T\to\infty} \sup_{|\vartheta-\vartheta_0|\leq\delta} I_{T,\Delta_T}(\vartheta)\mathcal{R}_T\big(\widehat{\vartheta}^{QV}_{T,\Delta_T}, \vartheta\big) \geq 1. \tag{2.13}$$

On the one hand, Theorem 4 suggests[6] that the lower bound (2.13) can be achieved in microscopic and macroscopic regimes. On the other hand, Theorem 5 shows that inequality

6. This is actually true as the uniform integrability of $\widehat{\vartheta}^{QV}_{T,\Delta_T}$ under $\mathbb{P}^{T,\Delta}_{\vartheta}$, locally uniformly in ϑ, can easily be obtained. We leave the details to the reader.

(2.13) is strict in the intermediate case, whenever the restriction (2.12) is satisfied, thus revealing a loss of efficiency in this sense. Define

$$\varphi(\vartheta,\Delta) = \mathbb{E}_{\vartheta}^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\big],$$

where $h_\Delta(\vartheta, k)$ is defined in Theorem 1. An inspection of the proof of Theorem 5 shows that $\varphi(\vartheta,\Delta) = \psi(\vartheta\Delta)$, for some univariate function ψ, and that

$$\frac{I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}}{I_{T,\Delta}(\vartheta)^{-1}} = \psi(\vartheta\Delta)\big(2(\vartheta\Delta)^2 + \vartheta\Delta\big).$$

The maximal loss of information is obtained for

$$\Delta^\star(\vartheta) \sim \vartheta^{-1}\mathrm{argmax}_{x>0}\psi(x)\big(2x^2 + x\big)$$

as $T \to \infty$. Numerical simulations show that the maximum loss of efficiency is close to 23%.

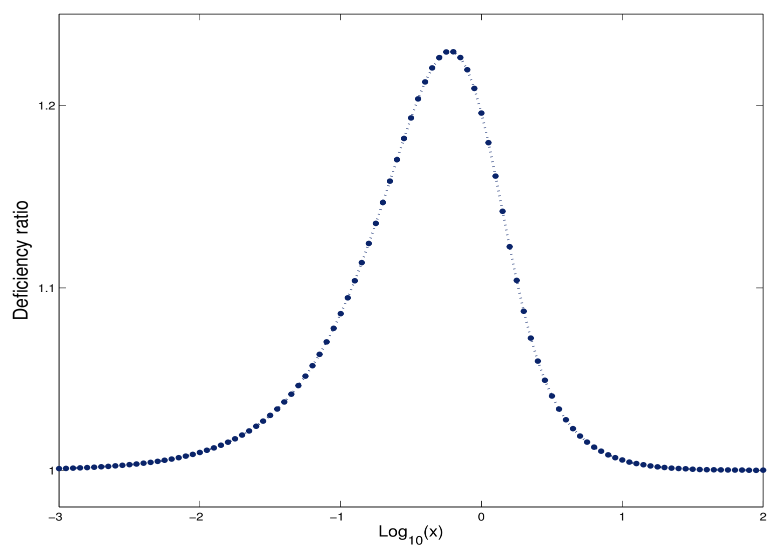

FIGURE 2.1 – **Deficiency ratio through scales** (x-axis $x = \vartheta\Delta$, logarithmic scale). Ratio between the information $I_{T,\Delta_T}(\vartheta)$ and the inverse of the variance of $\widehat{\vartheta}^{QV}_{T,\Delta_T}$. The maximum is 1.2297 and is attained at $x = 0.600$.

Since $\big(\mathcal{E}^{T,\Delta}\big)_{T>0}$ is regular for every $\Delta > 0$, an asymptotically normal estimator with asymptotic variance equivalent to $I_{T,\Delta}(\vartheta)^{-1}$ is given by the maximum likelihood estimator. However due to the absence of a closed-form for the likelihood ratio that involves the intricate function $f_\Delta(\vartheta, k)$ defined in (2.8) (see also Section 2.4.1), it seems easier to start from $\widehat{\vartheta}^{QV}_{T,\Delta}$ which is already rate-optimal by Theorem 4 and correct it by a classical one-step iteration based on the Newton-Rhapson method, see for instance the textbook [101] pp. 71–75. To that end, define

$$\widehat{\vartheta}_{T,\Delta}^{OS} = \widehat{\vartheta}_{T}^{QV} - \frac{\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \partial_\vartheta \log f_\Delta\big(\widehat{\vartheta}_{T,\Delta}^{QV}, X_{i\Delta} - X_{(i-1)\Delta}\big)}{\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \partial_\vartheta^2 \log f_\Delta\big(\widehat{\vartheta}_{T,\Delta}^{QV}, X_{i\Delta} - X_{(i-1)\Delta}\big)}. \tag{2.14}$$

Theorem 6. *In all three regimes (microscopic, intermediate and macroscopic), we have*

$$I_{T,\Delta_T}^{1/2}\big(\widehat{\vartheta}_{T,\Delta_T}^{OS} - \vartheta\big) \longrightarrow \mathcal{N}(0,1) \quad \textit{in } \mathbb{P}_\vartheta^{T,\Delta_T}\textit{-distribution as } T \to \infty.$$

Démonstration. In essence the regularity of $f_\Delta\big(\widehat{\vartheta}_{T,\Delta}^{QV}, X_{i\Delta} - X_{(i-1)\Delta}\big)$ enables to apply Theorem 5.45 of Van der Vaart [101]. □

Theorem 6 expresses the fact that $\widehat{\vartheta}_{T,\Delta_T}^{OS}$ automatically adapts to I_{T,Δ_T} and is therefore optimal across scales.

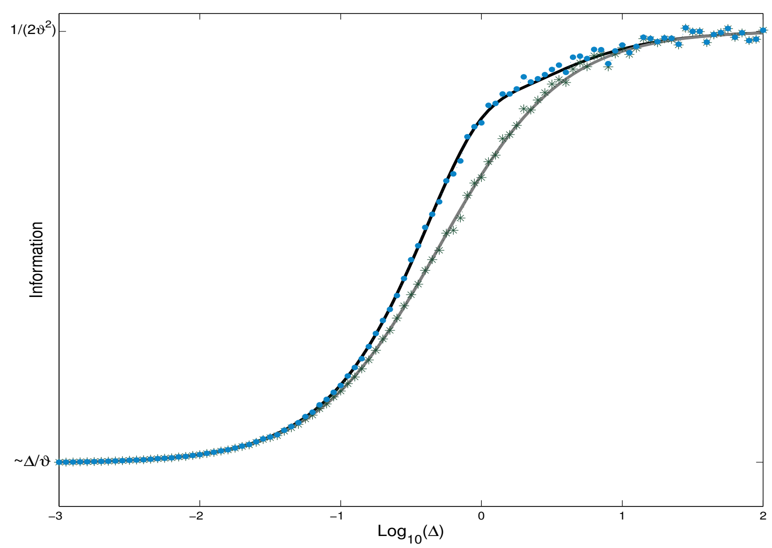

FIGURE 2.2 – **Information deficiency through scales ($\vartheta = 1$, x-axis Δ, logarithmic scale).** Information I_{T,Δ_T} (solid black). Inverse of the variance of $\widehat{\vartheta}_{T,\Delta_T}^{QV}$, theoretical (solid grey) and empirical (star green). The inverse of the empirical variance of $\widehat{\vartheta}_{T,\Delta_T}^{OS}$ (dotted blue) is close to the optimal I_{T,Δ_T} on all scales. The empirical variances were computed using 10^4 Monte-Carlo simulations from $T/\Delta = 10^5$ data, *i.e.* with the same number of data through scales.

2.3 Discussion

The compound Poisson process (X_t) with Bernoulli symmetric jumps defined in (2.1) is the simplest model of a continuous time symmetric random walk on a lattice that diffuses to a Brownian motion on a macroscopic scale. The intensity ϑ of the Poisson arrivals on a microscopic scale is transferred into the variance ϑ of the Brownian motion on a macroscopic scale :

$$\Big(\frac{1}{\sqrt{T}}X_{tT}, t \in [0,1]\Big) \longrightarrow \big(\sqrt{\vartheta}W_t, t \in [0,1]\big) \tag{2.15}$$

in distribution as $T \to \infty$, where (W_t) is a standard Brownian motion (see Appendix). The statistical inference program we have developed across time scales on the toy model given by (X_t) can be useful in several applied fields. For instance, in financial econometrics, (X_t) may be viewed as a toy model for a price process (last traded price, mid-price or bets bid/ask price) observed at the level of the order book, see *e.g.* [8] or [71]. The parameter ϑ can be interpreted as a trading intensity on microscopic scales that transfers into a macroscopic volatility in the diffusion regimes. Our results convey the message that if a practitioner samples (X_t) at high frequency at the same rate as price changes, which is customary in practice, then the realised volatility estimator $\widehat{\vartheta}^{QV}_{T,\Delta_T}$ is not efficient, and a modified estimator like $\widehat{\vartheta}^{OS}_{T,\Delta_T}$ should be used instead. However, this framework is a bit too simple and needs to be generalised in order to be more realistic in practice. Two directions can be explored in a relatively straightforward manner :

i) The extension to a non-homogeneous intensity Poisson process.

ii) The extension to an arbitrary compound law on a discrete lattice.

Extension to the non-homogeneous case

Theorems 1, 2 and 3 extend to the non-homogeneous case, when one allows the intensity of the jumps to depend on time. In this setting, the counting process (N_t) defined in (2.1) is defined on $[0,T]$ and has intensity

$$\Lambda_T\big(t,\vartheta\big) = \int_0^t \lambda(\vartheta, \tfrac{s}{T})ds, \quad \text{for } \ t \in [0,T]$$

where

$$\lambda : \vartheta \times [0,1] \to (0,\infty)$$

is the nonvanishing (integrable) intensity function, so that the process

$$\big(N_t - \Lambda_T(t,\vartheta), t \in [0,T]\big)$$

is a martingale. The homogeneous case is recovered by setting $\lambda(\vartheta,t) = \vartheta$ for every $t \in [0,1]$. In this context, the macroscopic approximation (2.15) becomes

$$\Big(\frac{1}{\sqrt{T}}X_{tT}, t \in [0,1]\Big) \longrightarrow \Big(\int_0^t \sqrt{\lambda(\vartheta,s)}\, dW_s, t \in [0,1]\Big)$$

in distribution as $T \to \infty$. We state – without proof – an extension of Theorems 1, 2 and 3 for the associated family of experiments $\big(\mathcal{E}^{T,\Delta_T}\big)_{T>0}$ across scales.

Assumption 1. *We have that* $\vartheta \rightsquigarrow \lambda(\vartheta, t)$ *is continuously differentiable for almost all* $t \in [0,1]$ *and moreover* $\sup_{\vartheta\in\Theta, t\in[0,1]} \lambda(\vartheta, t) < \infty$.

Theorem 7. *We have Theorems 1, 2 and 3 with the following generalisation*

1. *In the microscopic case* $\Delta_T \to 0$,

$$I_{T,0}(\vartheta) = T\int_0^1 \big(\partial_\vartheta \log \lambda(\vartheta, s)\big)^2 \lambda(\vartheta, s) ds.$$

2. *In the intermediate regime* $\Delta_T \to \Delta_\infty \in (0, \infty)$,

$$I_T(\vartheta) = T\Delta_\infty \int_0^1 \big(\partial_\vartheta \log \lambda(\vartheta, s)\big)^2 \lambda(\vartheta, s)^2 H(\vartheta, s) ds,$$

with

$$H(\vartheta, s) = \mathbb{E}_\vartheta^{T,\Delta_\infty}\Big[\Big(h_{\Delta_\infty}(\vartheta, s, X_{\Delta_\infty}) + \big(\lambda(\vartheta, s)\Delta_\infty\big)^{-1}|X_{\Delta_\infty}| - 1\Big)^2\Big],$$

and

$$h_{\Delta_\infty}(\vartheta, s, k) = \frac{\mathcal{I}_{|k|+1}(\lambda(\vartheta, s)\Delta_\infty)}{\mathcal{I}_{|k|}(\lambda(\vartheta, s)\Delta_\infty)}.$$

3. *In the macroscopic case* $\Delta_T \to \infty$ *with* $T/\Delta_T \to \infty$ *and* $T/\Delta_T^{1+\frac{1}{4}} = o((\log(T/\Delta_T))^{-\frac{1}{4}})$,

$$I_T(\vartheta) = \frac{T\Delta_T^{-1}}{2}\int_0^1 \big(\partial_\vartheta \log \lambda(\vartheta, s)\big)^2 ds.$$

The proof of Theorem 7 relies on the approximation

$$\int_{(i-1)\Delta_T}^{i\Delta_T} \lambda(\vartheta, \tfrac{s}{T}) ds = \Delta_T \lambda\Big(\vartheta, \frac{i-1}{T\Delta_T^{-1}}\Big) + \Delta_T r_T,$$

for $i = 1, \dots, \lfloor T\Delta_T^{-1}\rfloor$, where $r_T \to 0$ as $T \to \infty$ in all three regimes. Assumption 1 ensures that the convergence of the remainder is uniform in i and ϑ. This reduction enables us to transfer the problem of proving Theorems 1, 2 and 3 when substituting independent identically distributed random variables by independent non-equally distributed ones. This is not essentially more difficult, and the regularity of λ enables us to piece together the local information given by each increment $X_{i\Delta_T} - X_{(i-1)\Delta_T}$ in order to obtain the formulae of Theorem 7.

An analogous program as in Section 2.2.3 for the distortion of information could presumably be carried over, with appropriate modifications. For instance, one can show that

$$\widehat{\vartheta}_{T,\Delta_T}^{QV} = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \big(X_{i\Delta_T} - X_{(i-1)\Delta_T}\big)^2 \longrightarrow \int_0^1 \lambda(\vartheta, s) ds \quad \text{as } T \to \infty$$

in $\mathbb{P}_\vartheta^{T,\Delta_T}$-probability, in all three regimes. Then, in order to estimate ϑ efficiently, one should rather consider a contrast estimator that maximises

$$\widetilde{\vartheta} \rightsquigarrow U_{T,\Delta_T}(\widetilde{\vartheta}) = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} g_{\Delta_T}\big(\lambda(\widetilde{\vartheta}, i\Delta_T), X_{i\Delta_T} - X_{(i-1)\Delta_T}\big)$$

for a suitable function g_{T,Δ_T}, and make further assumptions on existence of a unique maximum for the limit – whenever it exists – of U_{T,Δ_T} under $\mathbb{P}_\vartheta^{T,\Delta_T}$ as $T \to \infty$. We do not pursue this here.

Extension to more general compound laws

The situation is a bit more delicate when one tries to generalise Theorems 1, 2 and 3 to an arbitrary compound law $(\zeta(\vartheta, k), k \in \mathbb{Z})$, for every $\vartheta \in \Theta$, with

$$0 \leq \zeta(\vartheta, k) \leq 1, \quad \text{for} \quad k \in \mathbb{Z} \quad \text{and} \quad \sum_{k\in\mathbb{Z}} \zeta(\vartheta, k) = 1,$$

(and $\zeta(\vartheta, 0) = 0$ for obvious identifiability conditions). We then observe a process (X_t) of the form (2.1), except that the jumps (ε_i) are now distributed according to

$$\mathbb{P}(\varepsilon_i = k) = \zeta(\vartheta, k), \quad k \in \mathbb{Z}.$$

In order to keep up with the preceding case, we normalise the compound law, imposing

$$\sum_{k\in\mathbb{Z}} k\,\zeta(\vartheta, k) = 0 \quad \text{and} \quad \sum_{k\in\mathbb{Z}} k^2 \zeta(k, \vartheta) = 1. \tag{2.16}$$

First, in the microscopic case, we approximately observe over the period $[0, T]$ a random number of jumps, namely N_T which is of order ϑT. Second, conditionally on N_T, the size of the jumps form a sequence of independent and identically distributed random variables with law $\zeta(\vartheta, k)$. On the other side, in the macroscopic limit, the effect of the size of the jumps is only tracked through their second moment, which is normalised to 1 by (2.16). Therefore it gives no additional information about ϑ. The situation is rather different from the case of symmetric Bernoulli jumps : here, the extraneous information about ϑ lies in the effect of the jumps, which are recovered in the microscopic regime and lost in the macroscopic one. There is however one way to reconcile with our initial setting, assuming that the compound law $\zeta(k)$ does not depend on ϑ and is known for simplicity. Then, for $k \in \mathbb{Z}$, we have

$$f_\Delta(\vartheta, k) = \mathbb{P}_\vartheta^{T,\Delta}\big(X_{i\Delta} - X_{(i-1)\Delta} = k\big) = \sum_{m\in\mathbb{Z}} \zeta^{\star m}(k) \frac{e^{-\vartheta\Delta}}{m!}(\vartheta\Delta)^m,$$

where $\zeta^{\star m}(k)$ is the probability that a random walk with law $\zeta(k)$ started at 0 reaches k in m steps exactly. Therefore

$$f_\Delta(\vartheta, k) = e^{-\vartheta\Delta}\mathcal{G}_k(\vartheta\Delta), \quad \text{with} \quad \mathcal{G}_k(x) = \sum_{m\in\mathbb{Z}} \zeta^{\star m}(k)\frac{x^m}{m!}. \tag{2.17}$$

In the symmetric Bernoulli case, we have $\mathcal{G}_k(x) = \mathcal{I}_{|k|}(x)$, where $\mathcal{I}_\nu(x)$ is the modified Bessel function of the first kind. Anticipating the proof of Theorems 1, 2 and 3, analogous results could presumably be obtained for an arbitrary compound law $\zeta(k)$ satisfying (2.16), provided accurate asymptotic expansions of $\mathcal{G}_k(x)$ are available in the viscinity of 0 and ∞. The same subsequent results about the distortion of information that are developed in Section 2.2.3 would presumably follow, with the same estimators $\widehat{\vartheta}^{QV}_{T,\Delta_T}$ and $\widehat{\vartheta}^{QS}_{T,\Delta_T}$, and the appropriate changes for $f_\Delta(\vartheta, k)$ in (2.14).

Relation to other work

Concerning the estimation of the law of the jumps, say ζ, we have an inverse problem. One tries to recover ζ from the observations of a compound Poisson process, the link between ζ and the law of the process being given by (2.17). In the setting of positive compound laws, Buchmann and Grübel [16, 17] succeed to invert that relation and give an estimator of ζ in the discrete and continuous case. That method which consists in inverting (2.17) is called decompounding. It was generalised by Bøgsted and Pitts [13] to renewal reward processes when the law of the holding times is known, inrestriction to the case of having positive jumps only.

The compound Poisson process is a pure jump Lévy process that can be studied accordingly. Using the Lévy-Khintchine formula, it is possible to estimate nonparametrically its Lévy measure which is given by the product $\vartheta \times \zeta$ in that case. This strategy is exploited by van Es *et al.* [103] for a known intensity. This estimation procedure does not restrict to compound Poisson processes and it includes the case of pure jump Lévy processes in general. Nonparametric estimation of the Lévy measure from high frequency data (that corresponds to our microscopic case $\Delta_T \to 0$) is thoroughly studied in Comte and Genon-Catalot [22] as well as in the intermediate regime (with $\Delta_T = \Delta_\infty$ fixed) in [24]. In that latter case, we also have the results of Neumann and Reiß [83].

2.4 Proofs

2.4.1 Preparation

Some estimates for $f_\Delta(\vartheta, k)$

We have, for $k \in \mathbb{Z}$:

$$f_\Delta(\vartheta, k) = \mathbb{P}_\vartheta^{T,\Delta}\big(X_\Delta = k\big) = \sum_{m\geq 0} \phi_m(k) \frac{e^{-\vartheta\Delta}}{m!}(\vartheta\Delta)^m$$

where $\phi_m(k)$ is the probability that a symmetric random walk in $\mathbb{Z}$ started from 0 has value k after m steps exactly :

$$\phi_m(k) = \begin{cases} 0 & \text{if } |k| > m \text{ or } |k| - m \text{ is odd} \\ 2^{-m}\begin{pmatrix} m \\ \frac{1}{2}(m+|k|) \end{pmatrix} & \text{otherwise.} \end{cases}$$

Let us introduce the modified Bessel function of the first kind [7]

$$\mathcal{I}_\nu(x) = \sum_{m\geq 0} \frac{(x/2)^{2m+\nu}}{m!\Gamma(\nu+m+1)},$$

for every $x, \nu \in \mathbb{R}$, and where

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$$

denotes the Gamma function. Straightforward computations show that

$$f_\Delta(\vartheta, k) = \exp(-\vartheta\Delta)\, \mathcal{I}_{|k|}\big(\vartheta\Delta\big). \tag{2.18}$$

See for instance [91], p. 21 Example 4.7. We gather some technically useful properties of the function $\mathcal{I}_\nu(x)$ that we will repeatedly use in the sequel.

Lemma 1. *1. For every $x \in \mathbb{R}\setminus\{0\}$ and $\nu \in \mathbb{R}$, we have*

$$\partial_x \mathcal{I}_\nu(x) = \mathcal{I}_{\nu+1}(x) + \frac{\nu}{x}\mathcal{I}_\nu(x). \tag{2.19}$$

2. For every $\mu > \nu > -\frac{1}{2}$ and $x > 0$, we have

$$\mathcal{I}_\mu(x) < \mathcal{I}_\nu(x). \tag{2.20}$$

Démonstration. Property 1 can be found in the textbook of Watson [109] and readily follows from the fact that $x \rightsquigarrow \mathcal{I}_\nu(x)$ is analytical with an infinite radius of convergence. Property 2 is less obvious and follows from Nasell [81]. □

7. The function $x \rightsquigarrow \mathcal{I}_\nu(x)$ can also be defined as the solution to the differential equation

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2+\nu^2)y = 0.$$

The Fisher information of $\mathcal{E}^{T,\Delta}$

For $i = 1, \dots, \lfloor T\Delta^{-1} \rfloor$, let $\mathcal{E}_i^{T,\Delta}$ denote the experiment generated by the observation of the incerement $X_{i\Delta} - X_{(i-1)\Delta}$. Since (X_t) has independent stationary increments, we have, for $k \in \mathbb{Z}$

$$\mathbb{P}_\vartheta^{T,\Delta}\big(X_{i\Delta} - X_{(i-1)\Delta} = k\big) = \mathbb{P}_\vartheta^{T,\Delta}\big(X_\Delta = k\big) = f_\Delta(\vartheta, k).$$

Using that $X_0 = 0$, it follows that

$$\mathcal{E}^{T,\Delta} = \bigotimes_{i=1}^{\lfloor T\Delta^{-1} \rfloor} \mathcal{E}_i^{T,\Delta} \tag{2.21}$$

as a product of independent observations given by the increments $X_{i\Delta} - X_{(i-1)\Delta}$, each experiment $\mathcal{E}_i^{T,\Delta}$ being dominated by the counting measure on $\mathbb{Z}$ with density $f_\Delta(\vartheta, k)$ given by (2.18) that does not depend on i. Moreover, $\mathcal{E}_i^{T,\Delta}$ has (possibly infinite) Fisher information given by

$$\mathfrak{I}_\Delta(\vartheta) = \mathbb{E}_\vartheta^{T,\Delta_T}\left[\big(\partial_\vartheta \log f_\Delta(\vartheta, X_\Delta)\big)^2\right] = \sum_{k \in \mathbb{Z}} \frac{\big(\partial_\vartheta f_\Delta(\vartheta, k)\big)^2}{f_\Delta(\vartheta, k)} \leq +\infty$$

which does not depend on i. We study the regularity of $\mathcal{E}^{T,\Delta}$ in the classical sense of Ibragimov and Hasminskii (see [57] p. 65).

Definition 2. *The experiment $\mathcal{E}_i^{T,\Delta}$ is regular (in the sense of Ibragimov and Hasminskii) if*

i) *The mapping $\vartheta \rightsquigarrow f_\Delta(\vartheta, k)$ is continuous on Θ for every $k \in \mathbb{Z}$.*

ii) *The Fisher information is finite : $\mathfrak{I}_\Delta(\vartheta) < +\infty$ for every $\vartheta \in \Theta$.*

iii) *The mapping $\vartheta \rightsquigarrow \partial_\vartheta\big(f_\Delta(\vartheta, \cdot)^{1/2}\big)$ is continuous in $\ell^2(\mathbb{Z})$.*

Lemma 2. *The experiments $\mathcal{E}_i^{T,\Delta}$ are regular.*

Démonstration. For every $k \in \mathbb{Z}$, $f_\Delta(\vartheta, k) = \exp(-\vartheta\Delta)\mathcal{I}_{|k|}(\vartheta\Delta)$, therefore i) is readily satisfied since $\vartheta \in \Theta \subset (0, \infty)$ and $\Delta > 0$. We also have $f_\Delta(\vartheta, k) > 0$ for every $k \in \mathbb{Z}$, then $\mathfrak{I}_\Delta(\vartheta)$ is well defined, but possibly infinite. In order to prove ii), we write

$$\begin{aligned} \partial_\vartheta \log f_\Delta(\vartheta, X_\Delta) &= \partial_\vartheta \log\left(e^{-\vartheta\Delta}\mathcal{I}_{|X_\Delta|}(\vartheta\Delta)\right) = -\Delta + \frac{\partial_\vartheta \mathcal{I}_{|X_\Delta|}(\vartheta\Delta)}{\mathcal{I}_{|X_\Delta|}(\vartheta\Delta)} \\ &= \Delta\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big), \end{aligned} \tag{2.22}$$

where we have set, for every $k \in \mathbb{Z}$,

$$h_\Delta(\vartheta, k) = \frac{\mathcal{I}_{|k|+1}(\vartheta\Delta)}{\mathcal{I}_{|k|}(\vartheta\Delta)}$$

and used Property (2.19). It follows that

$$\mathfrak{I}_\Delta(\vartheta) = \Delta^2\, \mathbb{E}_\vartheta^{T,\Delta}\left[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\right]. \tag{2.23}$$

Moreover the function $|X_\Delta| \rightsquigarrow \mathcal{I}_{|X_\Delta|}(\vartheta\Delta)$ is decreasing (see (2.20)), thus

$$0 \le h_\Delta(\vartheta, X_\Delta) \le 1 \tag{2.24}$$

and since X_Δ has all moments under $\mathbb{P}_\vartheta^{T,\Delta}$, we obtain ii). We proceed similarly for iii). First, for any $\vartheta \in \Theta$ and ε such that $\vartheta + \varepsilon \in \Theta$, we have

$$\partial_\vartheta\big(f_\Delta(\vartheta+\varepsilon,k)^{1/2}\big) - \partial_\vartheta\big(f_\Delta(\vartheta,k)^{1/2}\big) = \varepsilon\, \partial_\vartheta^2\big(f_\Delta(\vartheta_\varepsilon,k)^{1/2}\big)$$

for some $\vartheta_\varepsilon \in [\vartheta, \vartheta+\varepsilon]$. Second, we write

$$\partial_\vartheta\big(f_\Delta(\vartheta_\varepsilon,k)^{1/2}\big) = f_\Delta(\vartheta_\varepsilon,k)^{1/2}\tfrac{1}{2}\partial_\vartheta \log f_\Delta(\vartheta_\varepsilon,k),$$

and, differentiating a second time, we obtain that $\partial_\vartheta^2\big(f_\Delta(\vartheta_\varepsilon,k)^{1/2}\big)$ equals

$$f_\Delta(\vartheta_\varepsilon,k)^{1/2}\Big(\big(\tfrac{1}{2}\partial_\vartheta \log f_\Delta(\vartheta_\varepsilon,k)\big)^2 + \tfrac{1}{2}\partial_\vartheta^2 \log f_\Delta(\vartheta_\varepsilon,k)\Big).$$

Therefore, taking square and summing in k, we derive

$$\begin{aligned}&\sum_{k\in\mathbb{Z}}\Big(\partial_\vartheta\big(f_\Delta(\vartheta+\varepsilon,k)^{1/2}\big) - \partial_\vartheta\big(f_\Delta(\vartheta,k)^{1/2}\big)\Big)^2\\ &= \varepsilon^2\, \mathbb{E}_{\vartheta_\varepsilon}^{T,\Delta}\Big[\Big(\big(\tfrac{1}{2}\partial_\vartheta \log f_\Delta(\vartheta_\varepsilon,X_\Delta)\big)^2 + \tfrac{1}{2}\partial_\vartheta^2 \log f_\Delta(\vartheta_\varepsilon,X_\Delta)\Big)^2\Big].\end{aligned} \tag{2.25}$$

From ii), we have

$$\partial_\vartheta \log f_\Delta(\vartheta, X_\Delta) = \Delta\big(h_\Delta(\vartheta,X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big), \tag{2.26}$$

and this last quantity has moments of all orders under $\mathbb{P}_\vartheta^{T,\Delta}$, locally uniformly in ϑ. Likewise, using (2.26) and (2.19), it is easily seen that

$$\begin{aligned}\partial_\vartheta^2 \log f_\Delta(\vartheta, X_\Delta) &= \Delta^2 \frac{\mathcal{I}_{|X_\Delta|+2}(\vartheta\Delta)}{\mathcal{I}_{|X_\Delta|}(\vartheta\Delta)} + \Delta\vartheta^{-1}h_\Delta(\vartheta,X_\Delta)\\ &\quad - \Delta^2 h_\Delta(\vartheta,X_\Delta)^2 - \vartheta^{-2}|X_\Delta|.\end{aligned}$$

Thus $\partial_\vartheta^2 \log f_\Delta(\vartheta, X_\Delta)$ has moments of all orders under $\mathbb{P}_\vartheta^{T,\Delta_T}$ locally uniformly in ϑ, thanks to (2.20) and (2.24). The same property carries over to the term within the expectation in (2.25) and we thus obtain iii) by letting $\varepsilon \to 0$. □

By the factorisation (2.21), we infer that $\mathcal{E}^{T,\Delta}$ has Fisher information

$$\mathfrak{I}_{T,\Delta}(\vartheta) = \lfloor T\Delta_T^{-1}\rfloor \mathfrak{I}_\Delta(\vartheta) = \lfloor T\Delta^{-1}\rfloor \sum_{k\in\mathbb{Z}} \frac{\big(\partial_\vartheta f_\Delta(\vartheta,k)\big)^2}{f_\Delta(\vartheta,k)}$$

which is finite thanks to ii) of Lemma 2.

Lemma 3. *For every $\vartheta \in \Theta$, we have*

$$\mathfrak{I}_{T,\Delta}(\vartheta) = \lfloor T\Delta^{-1} \rfloor \Delta^2 \big(\mathbb{E}_\vartheta^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\big]\big), \tag{2.27}$$

Moreover in the microscopic and intermediate regimes, we have

$$\frac{\mathfrak{I}_{T,\Delta_T}(\vartheta)}{I_{T,\Delta_T}(\vartheta)} \to 1 \quad as \;\; T \to \infty. \tag{2.28}$$

Démonstration. In the course of the proof of Lemma 2, we have seen by (2.23) that

$$\mathfrak{I}_\Delta(\vartheta) = \Delta^2\, \mathbb{E}_\vartheta^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\big].$$

It follows that

$$\begin{aligned}
\mathfrak{I}_{T,\Delta}(\vartheta) &= \lfloor T\Delta^{-1} \rfloor \Delta^2\, \mathbb{E}_\vartheta^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\big] \\
&= \lfloor T\Delta^{-1} \rfloor \Delta^2 \Big(\mathbb{E}_\vartheta^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta|\big)^2\big] \\
&+ 1 - 2\mathbb{E}_\vartheta^{T,\Delta}\big[h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta|\big]\Big).
\end{aligned}$$

Since $\mathcal{E}_i^{T,\Delta}$ is regular by Lemma 2, we have $\mathbb{E}_\vartheta^{T,\Delta}\big[\partial_\vartheta \log f_\Delta(\vartheta, X_\Delta)\big] = 0$. Combining this with the equality

$$\partial_\vartheta \log f_\Delta(\vartheta, X_\Delta) = \Delta\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)$$

that we obtained in (2.22), we derive

$$\mathbb{E}_\vartheta^{T,\Delta}\big[h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta|\big] = 1,$$

and (2.27) follows. Expanding (2.27) further, we obtain the useful representation

$$\begin{aligned}
\mathfrak{I}_{T,\Delta}(\vartheta) = \lfloor T\Delta^{-1} \rfloor &\Big(\Delta^2\, \mathbb{E}_\vartheta^{T,\Delta}\big[h_\Delta(\vartheta, X_\Delta)^2\big] \\
&+ \frac{2\Delta}{\vartheta}\mathbb{E}_\vartheta^{T,\Delta}\big[|X_\Delta| h_\Delta(\vartheta, X_\Delta)\big] + \frac{\Delta}{\vartheta} - \Delta^2\Big).
\end{aligned} \tag{2.29}$$

Let us now assume that $\Delta = \Delta_T \to 0$. We will need the following asymptotic expansion of the function $\mathcal{I}_\nu(x)$ near 0.

Lemma 4. *We have, for $\nu \in \mathbb{N}$,*

$$\mathcal{I}_\nu(x) = \frac{1}{2^\nu \nu!} x^\nu \Big(1 + x r_\nu(x)\Big), \tag{2.30}$$

where $x \rightsquigarrow r_\nu(x)$ is continuous and satisfies $\sup_{\nu \geq 0} r_\nu(x) \to 0$ *when* $x \to 0$.

Proof of Lemma 4. We have an expression of $\mathcal{I}_\nu(x)$ as a power series, thus its Taylor expansion in a neighborhood of 0 is given by

$$\mathcal{I}_\nu(x) = \Big(\frac{x}{2}\Big)^\nu \frac{1}{\nu!}\Big(1 + x r_\nu(x)\Big),$$

where

$$r_\nu(x) = \frac{x}{4}\sum_{m\geq 0} \frac{\nu!}{(m+1)!(m+1+\nu)!}\Big(\frac{x}{2}\Big)^{2m} \leq x \sum_{m\geq 0}\frac{1}{m!}\Big(\frac{x}{2}\Big)^{2m} = xe^{x^2/4}.$$

Then $x \rightsquigarrow r_\nu(x)$ is continuous and satisfies $\sup_{\nu\geq 0} r_\nu(x) \to 0$ when $x \to 0$. □

By Lemma 4, a simple Taylor expansion shows that

$$h_{\Delta_T}(\vartheta, X_\Delta) = \frac{\mathcal{I}_{|X_\Delta|+1}(\vartheta\Delta_T)}{\mathcal{I}_{|X_{\Delta_T}|}(\vartheta\Delta_T)} = \frac{2\Delta_T}{\vartheta}\frac{1}{|X_\Delta|+1} + \Delta_T r'_T(\vartheta, X_\Delta)$$

where $|r'_T(\vartheta, X_\Delta)| \leq c(\vartheta)$, for some deterministic locally bounded $c(\vartheta)$. Plugging this last expression in (2.29), we obtain

$$\mathfrak{I}_T(\vartheta) = \frac{T}{\vartheta} + T\Delta_T r''_T(\vartheta),$$

with r''_T having the same property as r_T, whence (2.28) in the microscopic case. In the intermediate case, since $\Delta \rightsquigarrow \mathbb{E}_\vartheta^{T,\Delta}\big[\big(h_\Delta(\vartheta, X_\Delta) + (\vartheta\Delta)^{-1}|X_\Delta| - 1\big)^2\big]$ is continuous on $(0,\infty)$, we readily obtain the result using that $\lfloor T\Delta_T^{-1}\rfloor\Delta_T^2$ is equivalent to $T\Delta_T$ as $T \to \infty$. The proof of Lemma 3 is complete. □

2.4.2 Proof of Theorems 1 and 2

A technically convenient consequence of Lemma 3 in the microscopic and macroscopic cases is that it suffices to prove Theorems 1 and 2 with $\mathfrak{I}_{T,\Delta_T}(\vartheta)$ instead of $I_{T,\Delta_T}(\vartheta)$, provided the convergence (2.11) is valid locally uniformly. As $\mathcal{E}^{T,\Delta_T}$ is the product of $\mathcal{E}_i^{T,\Delta_T}$ generated by the $X_{i\Delta_T} - X_{(i-1)\Delta_T}$ that form independent and identically distributed random variables under $\mathbb{P}_\vartheta^{T,\Delta_T}$ with distribution depending on T, we are in the framework of Theorem 3.1' p. 128 in Ibragimov and Hasminskii [57]. It turns out that the LAN property is a consequence of the following two conditions :

i) For every $\vartheta_0 \in \Theta$ and h such that $\vartheta_0 + h \in \Theta$, we have

$$\frac{\lfloor T\Delta_T^{-1}\rfloor}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^2} \sup_{|\vartheta-\vartheta_0|\leq \frac{h}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^{1/2}}} \sum_{k\in\mathbb{Z}} \big|\partial_\vartheta^2\big(f_{\Delta_T}(\vartheta,k)^{1/2}\big)\big|^2 \to 0$$

as $T \to \infty$.

ii) For every $h > 0$ and $\vartheta \in \Theta$, we have

$$\frac{\lfloor T\Delta_T^{-1}\rfloor}{\mathfrak{I}_{T,\Delta_T}(\vartheta)}\mathbb{E}_\vartheta^{T,\Delta_T}\big[\big(\partial_\vartheta \log f_{\Delta_T}(\vartheta, X_{\Delta_T})\big)^2\mathbf{1}_{\{|\partial_\vartheta \log f_{\Delta_T}(\vartheta,X_\Delta)|\geq h\mathfrak{I}_{T,\Delta_T}(\vartheta)^{1/2}\}}\big] \to 0$$

as $T \to \infty$.

Remark 3. *Actually, as stated in Theorem 3.1 (and its asymptotic parameter dependent version Theorem 3.1' since the density $f_{\Delta_T}(\vartheta, k)$ of our observations depends on T) pp. 124–128 of the book by Ibragimov and Hasminskii [57], the LAN property at ϑ_0 is implied by the following two conditions*[8] *:*

For all h such that[9] *$\vartheta + h \in \Theta$, we have*

$$\sup_{u\leq h}\frac{\lfloor T\Delta_T^{-1}\rfloor}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)}\sum_{k\in\mathbb{Z}}\big|\partial_\vartheta\big(f_{\Delta_T}(\vartheta_0+u\,\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^{-1/2},k)\big)^{1/2}-\partial_\vartheta\big(f_{\Delta_T}(\vartheta_0,k)\big)^{1/2}\big|^2 \to 0 \qquad (3.1)$$

as $T \to \infty$.

For every $h > 0$ and $\vartheta \in \Theta$, we have

$$\frac{\lfloor T\Delta_T^{-1}\rfloor}{\mathfrak{I}_{T,\Delta_T}(\vartheta)}\mathbb{E}_\vartheta^{T,\Delta_T}\big[\big(\partial_\vartheta \log f_{\Delta_T}(\vartheta, X_{\Delta_T})\big)^2\mathbf{1}_{\{|\partial_\vartheta \log f_{\Delta_T}(\vartheta,X_\Delta)|\geq h\mathfrak{I}_{T,\Delta_T}(\vartheta)^{1/2}\}}\big] \to 0 \qquad (3.2)$$

as $T \to \infty$. Moreover, as explained in Remark 3.2 *pp.* 128 − 129 *in [57], Condition* (3.1) *is actually implied by the sufficient condition*

$$\frac{1}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^2}\sup_{|\vartheta-\vartheta_0|\leq\frac{h}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^{1/2}}}\sum_{j=1}^{\lfloor T\Delta_T^{-1}\rfloor}\sum_{k\in\mathbb{Z}}\big|\partial_\vartheta^2\big(f_{\Delta_T,j}(\vartheta,k)\big)^{1/2}\big|^2\to 0 \qquad (3.1')$$

as $T \to \infty$ for independent nonhomogeneous observations with density $f_{\Delta_T,j}(\vartheta, k)$. In our case we have $f_{\Delta_T,j}(\vartheta, k) = f_{\Delta_T}(\vartheta, k)$ for all j hence Condition $(3.1')$ *of [57] in Remark 3.2 is exactly our condition i). Condition* (3.2) *is exactly our condition ii) above.*

In the same way as for the proof of ii) in Lemma 2, we have

$$\begin{aligned}\partial_\vartheta^2\big(f_{\Delta_T}(\vartheta,k)\big)^{1/2} &= \tfrac{1}{2}f_{\Delta_T}(\vartheta,k)^{1/2}\Big(\tfrac{1}{2}\big(-\Delta_T+|k|\vartheta^{-1}+\Delta_T h_{\Delta_T}(\vartheta,k)\big)^2\\ &\quad -|k|\vartheta^{-2}+\Delta_T^2\big(\frac{\mathcal{I}_{|k|+2}(\vartheta\Delta_T)}{\mathcal{I}_{|k|}(\vartheta\Delta_T)}-h_{\Delta_T}(\vartheta,k)^2\big)+\frac{\Delta}{\vartheta}h_{\Delta_T}(\vartheta,k)\Big)\\ &= \tfrac{1}{2}f_{\Delta_T}(\vartheta,k)^{1/2}\mathcal{H}_{\Delta_T}(\vartheta,k),\quad \text{say.}\end{aligned}$$

8. with the labelling of their book for the conditions but with our notation for the statistical model and the fact that the dominating measure is the counting measure on $\mathbb{Z}$

9. remember that $\Theta \subset (0, \infty)$

Therefore, taking squares and summing in k, i) is proved if we show that

$$\frac{\lfloor T\Delta_T^{-1}\rfloor}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^2}\sup_{|\vartheta-\vartheta_0|\leq\frac{h}{\mathfrak{I}_{T,\Delta_T}(\vartheta_0)^{1/2}}}\mathbb{E}_{\vartheta}^{T,\Delta_T}\big[\mathcal{H}_{\Delta_T}(\vartheta,X_{\Delta_T})^2\big]\to 0 \tag{2.31}$$

as $T\to\infty$. Using (2.20), we have

$$0\leq h_{\Delta_T}(\vartheta,k)\leq 1\quad\text{and}\quad 0\leq\frac{\mathcal{I}_{|k|+2}(\vartheta\Delta_T)}{\mathcal{I}_{|k|}(\vartheta\Delta_T)}\leq 1,$$

hence $\mathcal{H}_{\Delta}(\vartheta,X_{\Delta_T})^2$ is less than

$$c(\vartheta)\big(\Delta_T^4+\Delta_T(1+\Delta_T^2)|X_{\Delta_T}|+(1+\Delta_T^2)X_{\Delta_T}^2+(1+\Delta_T)\big|X_{\Delta_T}\big|^3+X_{\Delta_T}^4\big)$$

for a locally bounded $c(\vartheta)$, which in turn is less than

$$c'(\vartheta,\Delta_T)\big(\Delta_T+\Delta_T^4+X_{\Delta_T}^2+X_{\Delta_T}^4\big), \tag{2.32}$$

for some $c'(\vartheta,\Delta_T)$ locally bounded on $\Theta\times[0,\infty)$. Since (X_t) is a compound Poisson process under $\mathbb{P}_{\vartheta}^{T,\Delta_T}$ with intensity ϑ and jumps in $\{-1,+1\}$ with equal probability, the characteristic function of X_{Δ_T} is explicitly given by

$$\mathbb{E}_{\vartheta}^{T,\Delta_T}\big[e^{iuX_{\Delta_T}}\big]=\exp\big(-\vartheta\Delta_T(1-\cos u)\big),\quad u\in\mathbb{R},$$

from which we obtain

$$\mathbb{E}_{\vartheta}^{T,\Delta_T}[X_{\Delta}^4]=\vartheta\Delta_T(1+3\vartheta\Delta_T). \tag{2.33}$$

Integrating (2.32), we derive

$$\mathbb{E}_{\vartheta}^{T,\Delta_T}\big[\mathcal{H}_{\Delta_T}(\vartheta,X_{\Delta_T})^2\big]\leq c''(\vartheta,\Delta_T)\Delta_T,$$

where c'' has the same property as c'. Since $\mathfrak{I}_{T,\Delta_T}(\vartheta_0)$ is of order T as $T\to\infty$ in both microscopic and intermediate scales, we obtain (2.31) and i) follows.

It remains to prove ii). From the explicit representation

$$\partial_\vartheta\log f_{\Delta_T}(\vartheta,k)=\Delta_T\big(-1+h_{\Delta_T}(\vartheta,k)\big)+|k|\vartheta^{-1} \tag{2.34}$$

and the fact that $0\leq h_{\Delta_T}(\vartheta,k)\leq 1$, we have

$$\big|\partial_\vartheta\log f_{\Delta_T}(\vartheta,k)\big|\leq\Delta_T+|k|\vartheta^{-1},$$

from which we readily obtain

$$\mathbb{E}_{\vartheta}^{T,\Delta_T}\big[\big(\partial_\vartheta\log f_{\Delta_T}(\vartheta,X_{\Delta_T})\big)^2\big]\leq c'''(\vartheta,\Delta_T),$$

where c''' has the same property as c'. Since $\mathfrak{I}_{T,\Delta_T}(\vartheta)\to\infty$ as $T\to\infty$, we obtain ii) in both microscopic and intermediate scales. The proof of Theorems 1 and 2 is complete.

2.4.3 Proof of Theorem 3

The strategy of the proof is quite different from that of Theorems 1 and 2, for we were not able to obtain asymptotic expansions for $\mathcal{I}_\nu(x)$ in a viscinity of $x = \infty$ with appropriate bounds on the stochastic remainders.

Consider instead the experiment $\mathcal{Q}^{T,\Delta_T} = \{\mathbb{Q}_\vartheta^{T,\Delta_T}, \vartheta \in \Theta\}$ generated by the observation of $\lfloor T\Delta_T^{-1} \rfloor$ independent centred Gaussian random variables with variance $\vartheta\Delta_T$, for $\Delta_T \to \infty$ satisfying the rate restriction

$$T/\Delta_T^{1+\frac{1}{4}} = o((\log(T/\Delta_T))^{-\frac{1}{4}}) \tag{2.35}$$

We plan to prove that under the restriction (2.35), the experiments $\mathcal{E}^{T,\Delta_T}$ and $\mathcal{Q}^{T,\Delta_T}$ are asymptotically equivalent as $T \to \infty$. Theorem 3 then follows from the Le Cam theory, see for instance [68]. To that end, we map each increment $X_{i\Delta_T} - X_{(i-1)\Delta_T}$ in $\mathcal{E}^{T,\Delta_T}$ with

$$Y_i^{\Delta_T} = X_{i\Delta_T} - X_{(i-1)\Delta_T} + U_i$$

where the U_i are independent random variables uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. Let us denote by $\widetilde{\mathcal{E}}^{T,\Delta_T} = \{\widetilde{\mathbb{P}}_\vartheta^{T,\Delta_T}, \vartheta \in \Theta\}$ the experiment generated by the $Y_i^{\Delta_T}$. Since the increments $X_{i\Delta} - X_{(i-1)\Delta_T}$ take values in $\mathbb{Z}$, we have a one-to-one correspondence between $Y_i^{\Delta_T}$ and the increment $X_{i\Delta_T} - X_{(i-1)\Delta_T}$ and therefore the experiments $\mathcal{E}^{T,\Delta_T}$ and $\widetilde{\mathcal{E}}^{T,\Delta_T}$ are equivalent. Moreover, $\widetilde{\mathcal{E}}^{T,\Delta_T}$ and $\mathcal{Q}^{T,\Delta_T}$ live on the same state space $\mathbb{R}^{\lfloor T\Delta_T^{-1} \rfloor}$ and have smooth densities with respect to the Lebesgue measure. The proof of Theorem 3 is therefore implied by the following bound

$$\|\widetilde{\mathbb{P}}_\vartheta^{T,\Delta_T} - \mathbb{Q}_\vartheta^{T,\Delta_T}\|_{TV} \to 0 \;\text{ as }\; T \to \infty,$$

locally uniformly in ϑ and where $\|\cdot\|_{TV}$ denotes the variational norm. This bound is implied in turn by the bound

$$\|\mathcal{L}(Y_1^{\Delta_T}) - \mathcal{N}(0, \vartheta\Delta_T)\|_{TV} = o\big((T/\Delta_T)^{-1}\big) \tag{2.36}$$

locally uniformly in ϑ, since each experiment is the $\lfloor T\Delta_T^{-1} \rfloor$-fold product independent and identically distributed random variables[10]. Let us further denote by p_{ϑ,Δ_T} and q_{ϑ,Δ_T} the densities of $Y_1^{\Delta_T}$ and the Gaussian law $\mathcal{N}(0, \vartheta\Delta_T)$ respectively. We have

$$\begin{aligned}\|\mathcal{L}(Y_1^{\Delta_T}) - \mathcal{N}(0, \vartheta\Delta_T)\|_{TV} &= \int_{\mathbb{R}} |p_{\vartheta,\Delta_T}(x) - q_{\vartheta,\Delta_T}(x)| dx \\ &\leq I + II + III,\end{aligned}$$

10. For instance, by using the bound

$$\|\mathbb{P}^{\otimes n} - \mathbb{Q}^{\otimes n}\|_{TV} \leq \sqrt{2}\big(1 - \big(1 - \tfrac{1}{2}\|\mathbb{P} - \mathbb{Q}\|_{TV}\big)^n\big)^{1/2}.$$

where, applying successively the triangle inequality and Cauchy-Schwarz, for any $\eta > 0$,

$$I = \sqrt{2\eta}\Big(\int_{\mathbb{R}} \big(p_{\vartheta,\Delta_T}(x) - q_{\vartheta,\Delta_T}(x)\big)^2 dx\Big)^{1/2},$$

$$II = \mathbb{P}_\vartheta^{\Delta_T}\big(|X_{\Delta_T} + U_1| \geq \eta\big),$$

$$III = \int_{|x|\geq\eta} q_{\vartheta,\Delta_T}(x)dx.$$

Set $\eta = \eta_T = \kappa\sqrt{\Delta_T \log(T/\Delta_T)}$. We claim that for $\kappa^2 > 2\vartheta$, the terms I, II and III are $o\big((T/\Delta_T)^{-1}\big)$ hence (2.36) and the result, for an appropriate choice of κ so that the convergence can hold locally uniformly in ϑ. Since $q_{\vartheta,\Delta_T}(x)$ is the density of the Gaussian law $\mathcal{N}(0, \vartheta\Delta_T)$, we readily obtain

$$III \leq 2\exp\big(-\frac{\eta_T^2}{2\vartheta\Delta_T}\big) = (T/\Delta_T)^{-\kappa^2/(2\vartheta)} = o\big((T/\Delta_T)^{-1}\big)$$

using $\kappa^2 > 2\vartheta$. For the term II, we observe that since $|U_1| \leq 1/2$, we have

$$\mathbb{P}_\vartheta^{\Delta_T}\big(|X_{\Delta_T} + U_1| \geq \eta_T\big) \leq \mathbb{P}_\vartheta^{\Delta_T}\big(\Big|\sum_{i=1}^{N_{\Delta_T}} \varepsilon_i\Big| \geq \eta_T - \tfrac{1}{2}\big),$$

where the $\varepsilon_i \in \{-1, 1\}$ are independent and symmetric. By Hoeffding inequality, this term is further bounded by

$$2\mathbb{E}_\vartheta^{T,\Delta_T}\big[\exp\big(-\tfrac{(\eta_T-1/2)^2}{2N_{\Delta_T}}\big)\big] \leq 2\Big(\exp\big(-\tfrac{(\eta_T-1/2)^2}{2\kappa'\Delta_T}\big) + \mathbb{P}_\vartheta^{T,\Delta_T}\big(N_{\Delta_T} \geq \kappa'\Delta_T\big)\Big)$$

for every $\kappa' > 0$. If $\kappa' < \kappa^2/2$, one readily checks that

$$\exp\big(-\tfrac{(\eta_T-1/2)^2}{2\kappa'\Delta_T}\big) = o\big((T/\Delta_T)^{-1}\big).$$

Moreover, if $\kappa' > \vartheta$, we have, by Chernov inequality,

$$\mathbb{P}_\vartheta^{T,\Delta_T}\big(N_{\Delta_T} - \vartheta\Delta_T \geq (\kappa' - \vartheta)\Delta_T\big) \leq \exp\Big(-\Delta_T(\kappa'\log(\kappa'/\vartheta) - (\kappa' - \vartheta))\Big)$$

and this term is also $o\big((T/\Delta_T)^{-1}\big)$. Thus II and III have the right order and it remains to bound the main term I. If we ddenote by $\widehat{p}_{\vartheta,\Delta_T}$ and $\widehat{q}_{\vartheta,\Delta_T}$ the Fourier transforms of p_{ϑ,Δ_T} and q_{ϑ,Δ_T}, by Plancherel equality we obtain the following explicit expression :

$$\int_{\mathbb{R}} \big(p_{\vartheta,\Delta_T}(x) - q_{\vartheta,\Delta_T}(x)\big)^2 dx = (2\pi)^{-1}\int_{\mathbb{R}} (\widehat{p}_{\vartheta,\Delta_T}(\xi) - \widehat{q}_{\vartheta,\Delta_T}(\xi))^2 d\xi$$

$$= (2\pi)^{-1}\int_{\mathbb{R}} \Big(e^{-\vartheta\Delta_T(1-\cos\xi)}\frac{\sin\frac{\xi}{2}}{\frac{\xi}{2}} - e^{-\frac{1}{2}\vartheta\Delta_T\xi^2}\Big)^2 d\xi$$

$$= (2\pi)^{-1}\int_{\mathbb{R}} \Big(e^{-\vartheta\Delta_T\big(1-\cos(\frac{\xi}{\sqrt{\Delta_T}})\big)}\frac{\sin(\frac{\xi}{2\sqrt{\Delta_T}})}{\frac{\xi}{2\sqrt{\Delta_T}}} - e^{-\frac{1}{2}\vartheta\xi^2}\Big)^2 \frac{d\xi}{\sqrt{\Delta_T}}$$

$$\leq IV + V + VI,$$

with

$$IV = (2\pi)^{-1}\int_{|\xi|\leq\rho\sqrt{\Delta_T}}\Big(e^{-\vartheta\Delta_T\big(1-\cos(\frac{\xi}{\sqrt{\Delta_T}})\big)}\frac{\sin(\frac{\xi}{2\sqrt{\Delta_T}})}{\frac{\xi}{2\sqrt{\Delta_T}}}-e^{-\frac{1}{2}\vartheta\xi^2}\Big)^2\frac{d\xi}{\sqrt{\Delta_T}},$$

$$V = (2\pi)^{-1}\int_{|\xi|\geq\rho\sqrt{\Delta_T}}e^{-2\vartheta\Delta_T\big(1-\cos(\frac{\xi}{\sqrt{\Delta_T}})\big)}\Big(\frac{\sin(\frac{\xi}{2\sqrt{\Delta_T}})}{\frac{\xi}{2\sqrt{\Delta_T}}}\Big)^2\frac{d\xi}{\sqrt{\Delta_T}},$$

$$VI = (2\pi)^{-1}\int_{|\xi|\geq\rho\sqrt{\Delta_T}}e^{-\vartheta\xi^2}\frac{d\xi}{\sqrt{\Delta_T}},$$

for any $\rho\geq 0$. By a first order expansion, we have that IV is less than

$$\int_{|\xi|\leq\rho\sqrt{\Delta_T}}e^{-\vartheta\xi^2}\Big(\Big(\frac{\xi^4\alpha(\frac{\xi}{\sqrt{\Delta_T}})}{\Delta_T}+\frac{\xi^6\alpha(\frac{\xi}{\sqrt{\Delta_T}})}{\Delta_T^2}\Big)e^{\frac{\xi^4\alpha\big(\frac{\xi}{\sqrt{\Delta_T}}\big)}{\Delta_T}}+\frac{\xi^2\alpha(\frac{\xi}{\sqrt{\Delta_T}})}{\Delta_T}\Big)^2\frac{d\xi}{\sqrt{\Delta_T}}$$

for some bounded function $\xi\rightsquigarrow\alpha(\xi)$. Set $\alpha^\star=\sup_x|\alpha(x)|$. We thus obtain that IV is less than a constant times

$$\int_{|\xi|\leq\rho\sqrt{\Delta_T}}\frac{\xi^8}{\Delta_T^2}(\alpha^\star)^2e^{-(\vartheta-2\rho\alpha^\star)\xi^2}\frac{d\xi}{\sqrt{\Delta_T}}.$$

If we pick ρ such that $\vartheta>2\rho\alpha^\star$, the term IV is of order $\Delta_T^{-5/2}$. For the term

$$V=\int_{|\xi|>\rho}e^{-2\vartheta\Delta_T(1-\cos\xi)}\Big(\frac{\sin(\frac{\xi}{2})}{\frac{\xi}{2}}\Big)^2d\xi,$$

noting that $(\sin x)^2=\big(1-\cos(2x)\big)/2$, we bound the 2π-periodic, even and continuous function $\xi\rightsquigarrow e^{-2\vartheta\Delta_T(1-\cos\xi)}(1-\cos\xi)$ by its supremum $(4e\vartheta\Delta_T)^{-1}$. The integrability of ξ^{-2} away from 0 enables to conclude that V is of order Δ_T^{-1}. Finally, by Gaussian approximation, we readily obtain that VI is of order $\Delta^{-1/2}e^{-\rho^2\vartheta\Delta_T}$.

In conclusion, we have that $\int_{\mathbb{R}}\big(p_{\vartheta,\Delta_T}(x)-q_{\vartheta,\Delta_T}(x)\big)^2dx$ is dominated by the term V and is thus of order Δ_T^{-1}. It follows that I is of order $\eta_T^{1/2}\Delta_T^{-1/2}$ and the choice $\eta_T=\kappa\sqrt{\Delta_T\log(T/\Delta_T)}$ implies $I=o\big((T/\Delta_T)^{-1}\big)$ thanks to the restriction condition $T/\Delta_T^{1+\frac{1}{4}}=o((\log(T/\Delta_T))^{-\frac{1}{4}})$. The proof of Theorem 3 is complete.

2.4.4 Proof of Theorem 4

Set

$$\xi_{i,T}=\frac{\big(X_{i\Delta_T}-X_{(i-1)\Delta_T}\big)^2-\vartheta\Delta_T}{\big(\lfloor T\Delta_T^{-1}\rfloor\vartheta\Delta_T(1+2\ \vartheta\Delta_T)\big)^{1/2}}.$$

Under $\mathbb{P}_\vartheta^{T,\Delta_T}$, the variables $\xi_{i,T}$ are independent, identically distributed, and we have

$$\mathbb{E}_\vartheta^{T,\Delta_T}\big[\xi_{i,T}\big] = 0 \;\text{ and }\; \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \mathrm{Var}\big[\xi_{i,T}\big] = 1$$

by (2.33). Moreover, for every $\delta > 0$,

$$\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \mathbb{E}_\vartheta^{T,\Delta_T}\Big[\big|\xi_{i,T}\big|^2 \mathbf{1}_{\{|\xi_{i,T}|\geq\delta\}}\Big] \to 0 \;\text{ as }\; T\to\infty,$$

therefore, by the central limit theorem $U_T = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \xi_{i,T} \to \mathcal{N}(0,1)$ in distribution under $\mathbb{P}_\vartheta^{T,\Delta_T}$ as $T\to\infty$ in all three regimes (microscopic, intermediate and macroscopic). We thus obtain the following representation

$$\widehat{\vartheta}_{T,\Delta_T}^{QV} = T^{-1}\Delta_T\lfloor T\Delta_T^{-1}\rfloor\vartheta + T^{-1}\big(\lfloor T\Delta_T^{-1}\rfloor\Delta_T\vartheta(1+2\vartheta\Delta_T)\big)^{1/2}U_T$$

and the result follows from $T^{-1}\Delta_T\lfloor T\Delta_T^{-1}\rfloor \sim 1$ and

$$T^{-2}\lfloor T\Delta_T^{-1}\rfloor\Delta_T\vartheta(1+2\vartheta\Delta_T) \sim I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}$$

as $T\to\infty$ in all three regimes.

2.4.5 Proof of Theorem 5

By (2.28) of Lemma 3, it suffices to prove

$$\liminf_{T\to\infty} \mathfrak{I}_{T,\Delta_T}(\vartheta)\big(I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}\big) > 1. \tag{2.37}$$

Up to taking a subsequence, we may assume that $\Delta_T \to \Delta \in (0, 1/(4\vartheta)]$ as $T\to\infty$. Using (2.29) and Theorem 4, for every $\vartheta\in\Theta$, we have

$$\begin{aligned}
&\mathfrak{I}_{T,\Delta_T}(\vartheta)\big(I_{T,0}(\vartheta)^{-1} + I_{T,\infty}(\vartheta)^{-1}\big)\\
&\sim \Big(\vartheta\Delta\mathbb{E}_\vartheta^{T,\Delta}\big[h_\Delta(\vartheta,X_\Delta)^2\big] + 2\mathbb{E}_\vartheta^{T,\Delta}\big[|X_\Delta|h_\Delta(\vartheta,X_\Delta)\big] + 1 - \vartheta\Delta\Big)(2\vartheta\Delta+1)\\
&=: \mathcal{M}(\vartheta\Delta) \;\text{ as }\; T\to\infty,
\end{aligned}$$

where $\mathcal{M}$ is a univariate function since

$\mathbb{P}_\vartheta^{T,\Delta}$ has density $f_\Delta(\vartheta,k) = e^{-\vartheta\Delta}\mathcal{I}_{|k|}(\vartheta\Delta)$ with respect to the counting measure on $\mathbb{Z}$. Therefore, Theorem 5 is equivalent to proving that

$$\mathcal{M}(x) > 1 \;\text{ for every }\; x\in\big(0,\tfrac{1}{4}\big]. \tag{2.38}$$

Using (2.19) of Lemma 1 we have

$$\begin{aligned}&\partial_x \mathcal{M}(\vartheta\Delta)\\ =&1-4\vartheta\Delta+(1+4\vartheta\Delta)\mathbb{E}_{\vartheta}^{T,\Delta}\big[h_\Delta(\vartheta,X_\Delta)^2\big]+4\mathbb{E}_{\vartheta}^{T,\Delta}\big[|X_\Delta|h_\Delta(\vartheta,X_\Delta)\big]\\ &+2(1+2\vartheta\Delta)\mathbb{E}_{\vartheta}^{T,\Delta}\big[\big(|X_\Delta|+\vartheta\Delta h_\Delta(\vartheta,X_\Delta)\big)\partial_\vartheta h_x(\vartheta,X_\Delta)\big]\end{aligned}$$

where

$$\partial_\vartheta h_\Delta(\vartheta,k)=\frac{\mathcal{I}_{|k|+2}(\vartheta\Delta)}{\mathcal{I}_{|k|}(\vartheta\Delta)}+\frac{1}{\vartheta\Delta}h_\Delta(\vartheta,k)-h_\Delta(\vartheta,k)^2$$

is positive (see Theorem 1 of Baricz [7]) and $h(\vartheta,k)$ is in $[0,1]$ according to (2.20) of Lemma 1. We derive

$$\partial_x\mathcal{M}(x)\geq 1-4x\geq 0 \ \text{ for } \ x\in\big(0,\tfrac14\big],$$

hence (2.38). Since $\mathcal{M}(x)\to 1$ as $x\to 0$, the conclusion follows.

Appendix

Proposition 1. *Suppose that ε_i are independent and identically distributed, with $\mathbb{P}(\varepsilon_i=-1)=\mathbb{P}(\varepsilon_i=1)=\frac12$, and independent of the homogeneous Poisson process (N_t) with intensity $\vartheta\in\Theta=(0,\infty)$. Let*

$$X_\Delta=\sum_{i=1}^{N_\Delta}\varepsilon_i.$$

Then

$$\Delta^{-1/2}X_\Delta\longrightarrow\mathcal{N}(0,\vartheta)$$

in distribution, as $\Delta\to\infty$.

Démonstration. By Skorokhod representation for weak convergence (see for instance Billingsley [11]) we have

$$\Big(\Delta^{-1/2}\sum_{i=1}^{\lfloor\Delta\cdot\rfloor}\varepsilon_i,\frac{N_\Delta}{\Delta}\Big)\longrightarrow(W,\vartheta)$$

in distribution as $\Delta\to\infty$, on $\mathbb{D}\big([0,\infty)\big)\times[0,\infty)$, where $\mathbb{D}\big([0,\infty)\big)$ is the Skorokhod space of càdlàg functions on $[0,\infty)$ and W is a standard Brownian motion. The statement is now trivial. □

Abstract

We study the nonparametric estimation of the jump density of a renewal reward process from the discrete observation of one trajectory over $[0, T]$. We consider the microscopic regime when the sampling rate $\Delta = \Delta_T \to 0$ as $T \to \infty$. We propose an adaptive wavelet threshold density estimator and study its performance for the L_p loss, $p \geq 1$, over Besov spaces. We achieve minimax rates of convergence for sampling rates Δ_T that vanish with T at polynomial rate. The estimating procedure is based on the inversion of the compounding operator in the same spirit as Buchmann and Grübel (2003). The inverse is explicit in the case of a compound Poisson (see Chapter 3), but has no closed form expression for renewal reward processes in general (see Chapter 4). In that latter case the inverse is approached with a fixed point technique.

Keywords : Compound Poisson process, Renewal reward process, Continuous time random walk, Discretely observed random process, Decompounding, Wavelet density estimation.

Note

Chapters 3 and 4 are based on submitted papers.

Chapitre 3

Estimation of a compound Poisson process at microscopic scales

3.1 Introduction

3.1.1 Statistical setting

Let R be a standard homogeneous Poisson process with intensity ϑ in $(0,\infty)$, we define the compound Poisson process X as

$$X_t = \sum_{i=1}^{R_t} \xi_i, \qquad t \geq 0$$

where the $\big(\xi_i\big)$ are independent and identically distributed random variables and independent of the Poisson process R.

Assume that we have discrete observations of the process X over $[0,T]$ at times $i\Delta$ for some $\Delta > 0$

$$\big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big). \tag{3.1}$$

We focus on the *microscopic regime*, namely

$$\Delta = \Delta_T \to 0 \qquad \text{as } \; T \to \infty$$

and work under the following assumption.

Assumption 1. *The law of the ξ_i has density f which is absolutely continuous with respect to the Lebesgue measure.*

We denote by $\mathcal{F}(\mathbb{R})$ the space of densities with respect to the Lebesgue measure supported by $\mathbb{R}$. We investigate the nonparametric estimation of the density f on a compact interval $\mathcal{D}$ included in $\mathbb{R}$ from the observations (3.1). To that end we use wavelet threshold density estimators and study their rate of convergence uniformly over Besov balls for the following loss function

$$\big(\mathbb{E}\big[\|\widehat{f}-f\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p}, \tag{3.2}$$

where $\widehat{f}$ is an estimator of f, $p \geq 1$ and

$$\|f\|_{L_p(\mathcal{D})} = \Big(\int_{\mathcal{D}} |f(x)|^p dx\Big)^{1/p}.$$

We also denote by $\|f\|_{L_p(\mathbb{R})}$ the usual L_p norm for $p \geq 1$

$$\|f\|_{L_p(\mathbb{R})} = \Big(\int_{\mathbb{R}} |f(x)|^p dx\Big)^{1/p}.$$

We do not assume the intensity ϑ to be known : it is a nuisance parameter.

By Assumption 1, on the event $\{X_{i\Delta} - X_{(i-1)\Delta} = 0\}$ no jump occurred between $(i-1)\Delta$ and $i\Delta$ and the increment $X_{i\Delta} - X_{(i-1)\Delta}$ gives no information on f. In the microscopic regime many increments are zero, therefore to estimate f we focus on the nonzero increments and denote by N_T their number over $[0, T]$. In that statistical context different difficulties arise. First the sample size N_T is random. Second on the event $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the increment $X_{i\Delta} - X_{(i-1)\Delta}$ is not necessarily a realisation of the density f. Indeed even if Δ is small there is always a positive probability that more than one jump occurred between $(i-1)\Delta$ and $i\Delta$. Conditional on $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the law of $X_{i\Delta} - X_{(i-1)\Delta}$ has density given by (see Proposition 1 in Section 3.2 below)

$$\mathbf{P}_\Delta[f](x) = \sum_{m=1}^{\infty} \mathbb{P}\big(R_\Delta = m \big| R_\Delta \neq 0\big) f^{\star m}(x), \qquad \text{for } x \in \mathbb{R}, \tag{3.3}$$

where $\star$ is the convolution product and $f^{\star m} = f \star \ldots \star f$, m times.

Adaptive estimators of the density f in that statistical context already exist. Under the condition $T\Delta_T \leq 1$ (or $T\Delta_T^2 \leq 1$ if f is smooth enough), they attain minimax rates of convergence over Sobolev spaces for the L_2 loss (see Bec and Lacour [9], Comte and Genon-Catalot [22, 25] and Figueroa-López [38]). In this Chapter we try to answer the following questions.

i) Is it possible to construct an estimator of f when Δ_T decays slowly to 0, for instance when Δ_T vanishes polynomially slowly with T.

ii) Is it possible to construct adaptive wavelet estimators that attain, over Besov spaces for the L_p loss defined in (3.2), the classical minimax rates of convergence of the experiment where we observe T independent realisations of f.

Without loss of generality, assuming T is an integer if we observe T independent realisations of a density f of regularity s measured with the L_π norm, $\pi > 0$, it is possible to achieve the minimax rates of convergence for the L_p loss –up to constants and logarithmic factors– which is of the form

$$T^{-\alpha(s,\pi,p)}$$

where $\alpha(s,\pi,p) \leq 1/2$ (see for instance Donoho *et al.* [30] and (3.16) hereafter). When the process X is continuously observed over $[0,T]$, we have R_T independent and identically distributed realisations of f. Moreover for T large enough, R_T is of the order of T. That is why we want compare the performance of estimators of f in the regime $\Delta_T \to 0$ with the classical minimax rate we would have if X were continuously observed.

3.1.2 Our Results

We build our estimator of f using equation (3.3) and proceed in two steps. The first step is the computation of the inverse of the operator $f \to \mathbf{P}_\Delta[f]$. The inverse takes the form

$$\mathbf{P}_\Delta^{-1}[\nu] = \sum_{m\geq 1}^{\infty} a_m(\vartheta, \Delta_T)\nu^{\star m}, \qquad \nu \in \mathcal{F}(\mathbb{R})$$

where the $\big(a_m(\vartheta,\Delta_T)\big)$ are explicit (see Proposition 1 below). They depend on the intensity ϑ and can be estimated. We take advantage of

$$f \approx \mathbf{L}_{\Delta,K}\big[\mathbf{P}_\Delta[f]\big], \tag{3.4}$$

where $\mathbf{L}_{\Delta,K}$ is the Taylor expansion of order K in Δ of $\mathbf{P}_\Delta^{-1}$. It depends only on $\big(\mathbf{P}_\Delta[f]^{\star m}, m = 1,\ldots,K+1\big)$. That step can be referred as decompounding as introduced in Buchmann and Grübel [16].

The second step consists in estimating the densities $\mathbf{P}_\Delta[f]^{\star m}$, for $m = 1,\ldots,K+1$. For that we use the N_T nonzero increments which are independent and with density $\mathbf{P}_\Delta[f]$. The difficulty here is that N_T is random. In Theorem 1 we show that conditional on N_T wavelet threshold estimators of $\mathbf{P}_\Delta[f]^{\star m}$ attain a rate of convergence –up to logarithmic factors– in $N_T^{-\alpha(s,\pi,p)}$. For T large enough we prove (see Proposition 2 in Section 3.5) that N_T concentrates around a deterministic value of the order of T, giving an unconditional rate of convergence in $T^{-\alpha(s,\pi,p)}$. We inject those estimators into $\mathbf{L}_{\Delta,K}$, defined in (3.4), and obtain an estimator of f that we call *estimator corrected at order K.*

The study of the rate of convergence of the estimator corrected at order K requires to control two distinct error terms. A deterministic one due the first step which is the error made when approximating f by $\mathbf{L}_{\Delta,K}\big[\mathbf{P}_\Delta[f]\big]$ in (3.4). And a statistical one due to the replacement of the $\mathbf{P}_\Delta[f]^{\star m}$ by estimators in the second step. The deterministic error decreases when K increases, then the idea is to choose K sufficiently large for the deterministic error term to be negligible in front of the statistical one. We give in Theorem 1 an upper bound for the rate of convergence of the estimator corrected at order K which is in –up to constants and logarithmic factors–

$$\max\{T^{-\alpha(s,\pi,p)}, \Delta_T^{K+1}\}.$$

It decreases with K and if there exists K_0 such that

$$T\Delta_T^{2K_0+2} \leq 1, \tag{3.5}$$

since $\alpha(s,\pi,p) \leq 1/2$ the estimator corrected at order K_0 attains the minimax rates of convergence. It follows that for every Δ_T polynomially decreasing with T, it is possible to exhibit K_0 such that (3.5) is valid and the estimator corrected at order K_0 provides a positive answer to **i)** and **ii)**. If no K enables to verify condition (3.5), Theorem 1 provides an upper bound for the rate of convergence of the estimator corrected at order K, in that case the estimator still provides a positive answer to **i)**.

In the case of a compound Poisson processes, the results of the present Chapter generalise to some extent those of Bec and Lacour [9], Comte and Genon-Catalot [22, 25] and Figueroa-López [38]. This is discussed in further details in Section 3.4. In Section 3.2 we give the main results of the Chapter. We properly define wavelet functions and Besov spaces used for the estimation before having a complete construction of the estimator corrected at order K. Then we give an upper bound for its rate of convergence for the L_p loss defined in (3.2), $p \geq 1$, uniformly over Besov balls. A numerical example illustrates the behavior of the estimator corrected at order K in Section 3.3. Finally Section 3.5 is dedicated to the proofs.

The model of this Chapter is central in many application fields *e.g.* statistical physics (see Moharir [78]), biology (see Huelsenbeck *et al.* [56]), financial series or mathematical insurance (see Scalas [94]). It is well adapted to study phenomena where random independent events occur at random times. For instance, in insurance failure theory these events can model the claims that insurance companies have to pay to the subscribers. The insurer's surplus at a given time t can be modeled by the following process

$$K(t) = K_0 + kt - X_t,$$

where K_0 is the capital of the company at time 0, the second term is a deterministic trend corresponding to the average income received from the subscribers and X is a compound Poisson process modeling the insurance claims occurring at random times with random amount of money at stake. It is the Cramér-Lundberg model ; see Embrechts *et al.* [32] or Scalas [94]. Compound Poisson processes can also model the changes of an asset price in finance ; see Masoliver *et al.* [71].

3.2 Main results

3.2.1 Besov spaces and wavelet thresholding

To estimate the densities $\big(\mathbf{P}_\Delta[f]^{\star m}, m = 1, \ldots, K+1\big)$ we use wavelet threshold density estimators and study their performance uniformly over Besov balls. In this paragraph we reproduce some classical results on Besov spaces, wavelet bases and wavelet threshold estimators (see Cohen [19], Donoho *et al.* [30] or Kerkyacharian and Picard [63]) that we use in the next sections.

Wavelets and Besov spaces

We describe the smoothness of a function with Besov spaces on $\mathcal{D}$. We recall here some well documented results on Besov spaces and their connection to wavelet bases (see Cohen [19], Donoho *et al.* [30] or Kerkyacharian and Picard [63]). Let $\big(\psi_\lambda\big)_\lambda$ be a regular wavelet basis adapted to the domain $\mathcal{D}$. The multi-index λ concatenates the spatial index and the resolution level $j = |\lambda|$. Set $\Lambda_j := \{\lambda, |\lambda| = j\}$ and $\Lambda = \cup_{j\geq -1}\Lambda_j$, for f in $L_p(\mathbb{R})$ we have

$$f = \sum_{j\geq -1}\sum_{\lambda\in\Lambda_j}\langle f, \psi_\lambda\rangle\psi_\lambda, \tag{3.6}$$

where $j = -1$ incorporates the low frequency part of the decomposition and $\langle ., \rangle$ denotes the usual L_2 inner product. We define Besov spaces in term of wavelet coefficients, for $s > 0$ and $\pi \in (0, \infty]$ a function f belongs to the Besov space $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ if the norm

$$\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} := \sup_{j\geq -1} 2^{j(s+1/2-1/\pi)}\Big(\sum_{\lambda\in\Lambda_j}|\langle f, \psi_\lambda\rangle|^\pi\Big)^{1/\pi} \tag{3.7}$$

is finite, with usual modifications if $\pi = \infty$.

We need additional properties on the wavelet basis $\big(\psi_\lambda\big)_\lambda$, which are listed in the following assumption.

Assumption 2. *For $p \geq 1$,*

- *We have for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1}2^{|\lambda|(p/2-1)} \leq \|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{|\lambda|(p/2-1)}.$$

- *For some $\mathfrak{C} > 0$, $\sigma > 0$ and for all $s \leq \sigma$, $J \geq 0$, we have*

$$\Big\|f - \sum_{j\leq J}\sum_{\lambda\in\Lambda_j}\langle f, \psi_\lambda\rangle\psi_\lambda\Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{-Js}\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}. \tag{3.8}$$

– *If $p \geq 1$, for some $\mathfrak{C} \geq 1$ and for any sequence of coefficients $\left(u_\lambda\right)_{\lambda\in\Lambda}$,*

$$\mathfrak{C}^{-1}\Big\|\sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda\Big\|_{L_p(\mathcal{D})} \leq \Big\|\Big(\sum_{\lambda\in\Lambda}|u_\lambda\psi_\lambda|^2\Big)^{1/2}\Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C}\Big\|\sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda\Big\|_{L_p(\mathcal{D})}. \tag{3.9}$$

– *For any subset $\Lambda_0 \subset \Lambda$ and for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1}\sum_{\lambda\in\Lambda_0}\|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \int_{\mathcal{D}}\Big(\sum_{\lambda\in\Lambda_0}|\psi_\lambda(x)|^2\Big)^{p/2} \leq \mathfrak{C}\sum_{\lambda\in\Lambda_0}\|\psi_\lambda\|^p_{L_p(\mathcal{D})}. \tag{3.10}$$

Property (3.8) ensures that definition (3.7) of Besov spaces matches the definition in terms of linear approximation. Property (3.9) ensures that $\left(\psi_\lambda\right)_\lambda$ is an unconditional basis of L_p and (3.10) is a super-concentration inequality (see Kerkyacharian and Picard [63] p. 304 and p. 306).

Wavelet threshold estimator

Let (ϕ, ψ) be a pair of a scaling function and a mother wavelet that generate a basis $\left(\psi_\lambda\right)_\lambda$ satisfying Assumption 2 for some $\sigma > 0$. We rewrite (3.6)

$$f = \sum_{k\in\Lambda_0}\alpha_{0k}\phi_{0k} + \sum_{j\geq 1}\sum_{k\in\Lambda_j}\beta_{jk}\psi_{jk},$$

where $\phi_{0k}(\bullet) = \phi(\bullet - k)$ and $\psi_{jk}(\bullet) = 2^{j/2}\psi(2^j\bullet - k)$ and

$$\alpha_{0k} = \int \phi_{0k}(x)f(x)dx$$
$$\beta_{jk} = \int \psi_{jk}(x)f(x)dx.$$

For every $j \geq 0$, the set Λ_j has cardinality 2^j and incorporates boundary terms that we choose not to distinguish in the notation for simplicity. An estimator of a function f is obtained when replacing the (α_{0k}) and (β_{jk}) by estimated values. In the sequel we uses (γ_{jk}) to design either (α_{0k}) or (β_{jk}) and (g_{jk}) for the wavelet functions (ϕ_{0k}) or (ψ_{jk}).

We consider classical hard threshold estimators of the form

$$\widehat{f}(\bullet) = \sum_{k\in\Lambda_0}\widehat{\alpha_{0k}}\phi_{0k}(\bullet) + \sum_{j=1}^{J}\sum_{k\in\Lambda_j}\widehat{\beta_{jk}}\mathbb{1}_{\left\{|\widehat{\beta_{jk}}|\geq\eta\right\}}\psi_{jk}(\bullet),$$

where $\widehat{\alpha_{0k}}$ and $\widehat{\beta_{jk}}$ are estimators of α_{0k} and β_{jk}, J and η are respectively the resolution level and the threshold, possibly depending on the data. Thus to construct $\widehat{f}$ we have to specify estimators $(\widehat{\gamma_{jk}})$ of the (γ_{jk}) and the coefficients J and η.

3.2.2 Construction of the estimator

Assume that we have $\lfloor T\Delta^{-1}\rfloor$ discrete data at times $i\Delta$ for some $\Delta > 0$ of the process X

$$\big(X_\Delta, \dots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big).$$

Introduce the increments

$$\mathbf{D}^\Delta X_i = X_{i\Delta} - X_{(i-1)\Delta}, \quad \text{for } i = 1, \dots, \lfloor T\Delta^{-1}\rfloor,$$

where $X_0 = 0$. They are independent and identically distributed since X is a compound Poisson process. Define

$$\begin{aligned} S_1 &= \inf\big\{j, \mathbf{D}^\Delta X_j \neq 0\big\} \wedge T \\ S_i &= \inf\big\{j > S_{i-1}, \mathbf{D}^\Delta X_j \neq 0\big\} \wedge T \quad \text{for } i \geq 1, \end{aligned}$$

where S_i is the random index of the ith jump and

$$N_T = \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} \mathbb{1}_{\{\mathbf{D}^\Delta X_i \neq 0\}}$$

the random number of nonzero increments observed over $[0, T]$. By Assumption 1, on the event $\{\mathbf{D}^\Delta X_i = 0\}$, no jump occurred between $(i-1)\Delta$ and $i\Delta$. In the microscopic regime when $\Delta = \Delta_T \to 0$ as T goes to infinity infinitely many increments are null and convey no information on f, hence for the estimation of f we focus on the nonzero ones

$$\big(\mathbf{D}^\Delta X_{S_1}, \dots, \mathbf{D}^\Delta X_{S_{N_T}}\big).$$

Proposition 1. *The distribution of the increment $\mathbf{D}^\Delta X_{S_1}$ has density with respect to the Lebesgue measure given by*

$$\mathbf{P}_\Delta[f] = \sum_{m=1}^{\infty} p_m(\Delta) f^{\star m},$$

where

$$p_m(\Delta) = \mathbb{P}\big(R_\Delta = m | R_\Delta \neq 0\big) = \frac{1}{e^{\vartheta\Delta} - 1} \frac{\big(\vartheta\Delta\big)^m}{m!}.$$

Let Δ_0 be such that

$$\sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta_0\big)^{m-2}}{m!} \leq 1.$$

For $\Delta \leq \Delta_0$, we have that

$$1 - \vartheta\Delta \leq p_1(\Delta) \leq 1.$$

It is straightforward to verify that the nonlinear operator $\mathbf{P}_\Delta$ is a mapping from $\mathcal{F}(\mathbb{R})$ to itself. The observations $\left(\mathbf{D}^\Delta X_{S_i}\right)$ are realisations of the density $\mathbf{P}_\Delta[f]$ and by Proposition 1 the weight $p_1(\Delta) \to 1$ in the limit $\Delta = \Delta_T \to 0$. It follows that for Δ_T small enough most of the $\left(\mathbf{D}^\Delta X_{S_i}\right)$ have distribution f. Then a naive method to estimate f is to apply classical density estimators to the $\left(\mathbf{D}^\Delta X_{S_i}\right)$. That estimator requires a convergence condition on Δ_T to achieve minimax rate of convergence (see Theorem 1). However we wish to construct an estimator that attains minimax rates of convergence with weaker conditions on Δ_T.

We adopt the estimating strategy of section 3.1.2 and construct an approximation of f.

Lemma 1. *The inverse $\mathbf{P}_\Delta^{-1}$ of $\mathbf{P}_\Delta$, such that for all densities f in $\mathcal{F}(\mathbb{R})$ if $\mathbf{P}_\Delta[f] = \nu$ we have $\mathbf{P}_\Delta^{-1}[\nu] = f$, is given by*

$$\mathbf{P}_\Delta^{-1}[\nu] = \frac{1}{\vartheta\Delta}\sum_{m=1}^{\infty}\frac{(-1)^{m+1}}{m}(e^{\vartheta\Delta}-1)^m\nu^{\star m}.$$

The proof of Lemma 1 is given in the Appendix. To build the estimator corrected at order K we use that $\mathbf{P}_\Delta^{-1}$ is a power series whose coefficients are equivalent to increasing powers of Δ. Then $\mathbf{L}_{\Delta,K}$ the Taylor expansion of order K in Δ of $\mathbf{P}_\Delta^{-1}$ is obtained by keeping the first $K+1$ terms of the inverse

$$\mathbf{L}_{\Delta,K}[\nu] = \frac{1}{\vartheta\Delta}\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}(e^{\vartheta\Delta}-1)^m\nu^{\star m}, \quad \nu \in \mathcal{F}(\mathbb{R}). \tag{3.11}$$

Next we construct wavelet threshold density estimators of the first $K+1$ convolution powers of $\mathbf{P}_\Delta[f]$ that will be plugged in (3.11). Define

$$\widehat{\gamma}_{jk}^{(m)} = \frac{1}{N_{T,m}}\sum_{i=1}^{N_{T,m}} g_{jk}\left(\mathbf{D}_m^\Delta X_{S_i}\right) \quad m \geq 1, \tag{3.12}$$

where $N_{T,m} = \lfloor N_T/m \rfloor$ with $N_{T,m} \geq 1$ for large enough T and

$$\mathbf{D}_m^\Delta X_{S_i} = \mathbf{D}^\Delta X_{S_i} + \mathbf{D}^\Delta X_{S_{N_{T,m}+i}} + \cdots + \mathbf{D}^\Delta X_{S_{(m-1)N_{T,m}+i}}.$$

The $\left(\mathbf{D}^\Delta X_{S_i}\right)$ are independent and identically distributed with density $\mathbf{P}_\Delta[f]$, thus the $\left(\mathbf{D}_m^\Delta X_{S_i}\right)$ are independent and identically distributed with density $\mathbf{P}_\Delta[f]^{\star m}$. Let $\eta > 0$ and $J \in \mathbb{N}\setminus\{0\}$, define $\widehat{P_{\Delta,m}}$ the estimator of $\mathbf{P}_\Delta[f]^{\star m}$ over $\mathcal{D}$

$$\widehat{P_{\Delta,m}}(x) = \sum_k \widehat{\alpha}_{0k}^{(m)}\phi_{0k}(x) + \sum_{j=0}^{J}\sum_k \widehat{\beta}_{jk}^{(m)}\mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{(m)}|\geq\eta\right\}}\psi_{jk}(x), \quad x \in \mathcal{D}. \tag{3.13}$$

Definition 1. *We define* $\widetilde{f}_{T,\Delta}^K$ *the estimator corrected at order* K *for* K *in* $\mathbb{N}$ *and* x *in* $\mathcal{D}$ *as*

$$\widetilde{f}_{T,\Delta}^K(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\left(e^{\widehat{\vartheta}_T \Delta} - 1\right)^m}{\widehat{\vartheta}_T \Delta} \widehat{P_{\Delta,m}}(x), \tag{3.14}$$

where

$$\widehat{\vartheta}_T = -\frac{1}{\Delta} \log(1 - \widehat{p}_T) \tag{3.15}$$

and

$$\widehat{p}_T = \frac{N_T}{\lfloor T\Delta^{-1} \rfloor}$$

is the empirical estimator of $p(\Delta) = \mathbb{P}(R_\Delta \neq 0) = 1 - e^{-\vartheta\Delta}$.

Lemma 1 justifies the form of the estimator corrected at order K.

3.2.3 Convergence rates

We estimate densities f which verify a smoothness property in term of Besov balls

$$\mathcal{F}(s, \pi, \mathfrak{M}) = \left\{ f \in \mathcal{F}(\mathbb{R}), \|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \mathfrak{M} \right\},$$

where $\mathfrak{M}$ is a positive constant. We are interested in estimating f on the compact interval $\mathcal{D}$, that is why we only impose that its restriction to $\mathcal{D}$ belongs to a Besov ball.

Theorem 1. *We work under Assumptions 1 and 2. Let* $\sigma > s > 1/\pi$, $p \geq 1 \wedge \pi$ *and* $\widehat{P_{\Delta_T,m}}$ *be the threshold wavelet estimator of* $\mathbf{P}_{\Delta_T}[f]^{\star m}$ *on* $\mathcal{D}$ *constructed from* (ϕ, ψ) *and defined in* (3.13)*. Take* J *such that*

$$2^J N_T^{-1} \log\left(N_T^{1/2}\right) \leq 1,$$

and

$$\eta = \kappa N_T^{-1/2} \sqrt{\log\left(N_T^{1/2}\right)},$$

for some $\kappa > 0$*. Let*

$$\alpha(s, p, \pi) = \min\left\{ \frac{s}{2s+1}, \frac{s + 1/p - 1/\pi}{2\big(s + 1/2 - 1/\pi\big)} \right\}. \tag{3.16}$$

1) The estimator $\widehat{P_{\Delta_T,m}}$ *verifies for large enough* T *and sufficiently large* $\kappa > 0$

$$\sup_{\mathbf{P}_{\Delta_T}[f]^{\star m} \in \mathcal{F}(s,\pi,\mathfrak{M})} \left(\mathbb{E}\left[\left\| \widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m} \right\|_{L_p(\mathcal{D})}^p \middle| N_T \right] \right)^{1/p} \leq \mathfrak{C} N_T^{-\alpha(s,p,\pi)},$$

up to logarithmic factors in T *and where* $\mathfrak{C}$ *depends on* $s, \pi, p, \mathfrak{M}, \phi, \psi$.

2) The estimator corrected at order K $\widetilde{f}^K_{T,\Delta_T}$ defined in (3.14) *verifies for T large enough and any positive constants $\underline{\mathfrak{T}}$ and $\overline{\mathfrak{T}}$*

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\big\|\widetilde{f}^K_{T,\Delta_T}-f\big\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\leq\mathfrak{C}\max\big\{T^{-\alpha(s,p,\pi)},\Delta_T^{K+1}\big\},$$

up to logarithmic factors in T and where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi,\underline{\mathfrak{T}},\overline{\mathfrak{T}},K$.

The proof of Theorem 1 is postponed to Section 3.5. From a practical point of view when one computes the estimator $\widetilde{f}^K_{T,\Delta_T}$ from (3.1) the sample size is N_T, which is why in Theorem 1 we give the resolution level J and the threshold η as functions of N_T instead of replacing N_T by its deterministic counterpart. Explicit bound for κ is given in Lemma 5 hereafter.

In practice the values T and Δ_T are imposed or chosen by the practitioner. Theorem 1 ensures that the estimator corrected at order K attains the minimax rate $T^{-\alpha(s,p,\pi)}$ for the smallest K such that

$$\Delta_T=O\big(T^{-\frac{\alpha(s,p,\pi)}{K+1}}\big).$$

Since $\alpha(s,p,\pi)\leq 1/2$ it is sufficient to choose K such that

$$T\Delta_T^{2K+2}=O(1).$$

If Δ_T decays as a power of T *i.e.* if there exists $\delta>0$ such that for some $\mathfrak{C}>0$

$$\Delta_T\leq\mathfrak{C}T^{-\delta},$$

it is always possible to find a correction level K satisfying the previous constraint. The case $K=0$ corresponds to the uncorrected estimator ; it is the naive estimator one would compute making the approximation $f\approx\mathbf{P}_\Delta[f]$. In that case we get a rate of convergence in

$$\max\{T^{-\alpha(s,p,\pi)},\Delta_T\},$$

which attains the minimax rate if $T^{\alpha(s,p,\pi)}\Delta_T\leq 1$. Since $\alpha(s,\pi,p)\leq 1/2$, it follows that the condition $T^{\alpha(s,p,\pi)}\Delta_T\leq 1$ already improves on the condition $T\Delta_T^2\leq 1$ of Bec and Lacour [9], Comte and Genon-Catalot [22, 25] or Figueroa-López [38] (see Section 3.4 for comparison with other works).

3.3 A numerical example

We illustrate the behaviour of the estimator corrected at order K when K increases and compare its performance with an oracle : the wavelet estimator we would compute in the idealised framework where all the jumps are observed

$$\widehat{f}^{Oracle}(x)=\sum_k\widehat{\alpha}^{Oracle}_{0k}\phi_{0k}(x)+\sum_{j=0}^J\sum_k\widehat{\beta}^{Oracle}_{jk}\mathbb{1}_{\big\{|\widehat{\beta}^{Oracle}_{jk}|\geq\eta\big\}}\psi_{jk}(x),$$

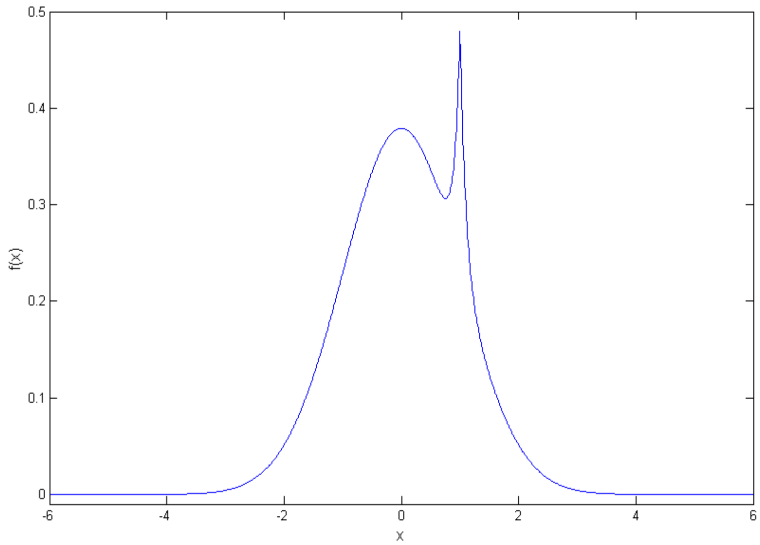

FIGURE 3.1 – Density f : $f(x) = 0.95 f_1(x) + 0.05 f_2(x)$ $x \in [-6, 6]$.

where

$$\widehat{\alpha}_{0k}^{Oracle} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i) \quad \text{and} \quad \widehat{\beta}_{jk}^{Oracle} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i),$$

R_T being the value of the Poisson process R at time T and (ξ_i) the jumps. The parameters J and η as well as the wavelet bases (ϕ, ψ) are the same as those used to compute the estimator corrected at order K. We consider a compound Poisson process of intensity $\vartheta = 1$ on $[0, T]$ and of compound law

$$f(x) = (1-a) f_1(x) + a f_2(x)$$

where f_1 is the density of a Gaussian $\mathcal{N}(0, 1)$ and f_2 of a Laplace with location parameter 1 and scale parameter 0.1, we take $a = 0.05$.
We estimate the mixture f (see Figure 3.1) on $\mathcal{D} = [-6, 6]$ with the estimator corrected at order K for different values of K and study the results with the L_2 error. We also compare them with the oracle $\widehat{f}^{Oracle}$. Wavelet estimators are based on the evaluation of the first wavelet coefficients, to perform those we use Symlets 4 wavelet functions and a resolution level $J = 10$. Moreover we transform the data in an equispaced signal on a grid of length 2^L with $L = 8$, it is the binning procedure (see Härdle *et al.* [49] Chap. 12). The threshold is chosen as in Theorem 1. The estimators we obtain take the form of a vector giving the estimated values of the density f on the uniform grid $[-6, 6]$ with mesh 0.01. We use the wavelet toolbox of Matlab.

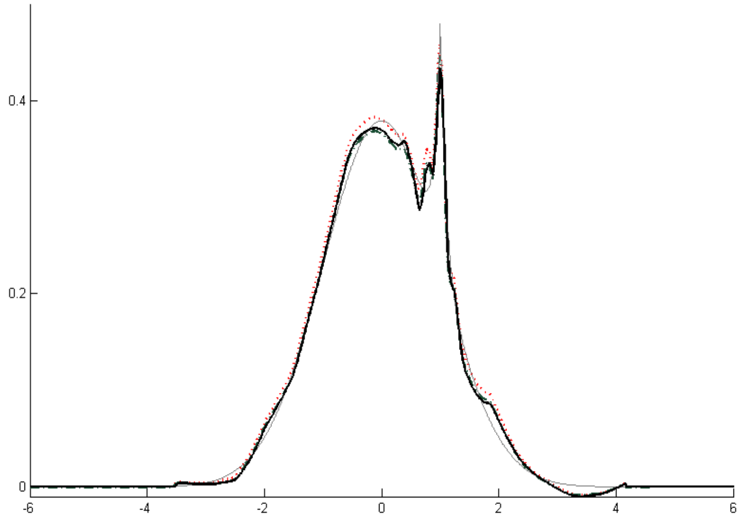

FIGURE 3.2 – Estimators of the density f (plain grey) for $T = 10000$ and $\Delta = 0.1$: the uncorrected (dotted red), the 1-corrected (dashed green) and the oracle (plain dark).

Figure 3.2 represents the corrected estimator for $K = 0$ and $K = 1$ and the oracle. All the estimators are evaluated on the same trajectory. They manage to reproduce the shape of the density f. As expected the oracle looks better than the other two and the uncorrected ($K = 0$) seems to make larger errors than the 1-corrected in estimating f. Figure 3.3 represents for every values in $[-6, 6]$ the absolute distance between those estimators –evaluated on the same trajectory– and the true density f. Therefore it enables to determine in which area an estimator fails to estimate f and to get an idea of the error made. The graphic was obtained after $M = 1000$ Monte-Carlo simulations of each estimator and averaging the results. The uncorrected estimator is not as good as the estimator corrected at order 1. The oracle and the estimator corrected at order 1 seem to have similar performances. Each of the estimators makes larger errors around 1 which is where the density f is peaked.

Evaluation of the L_2 errors enables to confirm the former graphical observation. We approximate the L_2 errors by Monte Carlo. For that we compute $M = 1000$ times each estimator (for $T = 10000$ and $\Delta = 0.1$) and approximate the L_2 loss by

$$\frac{1}{M}\sum_{i=1}^{M}\Big(\sum_{p=0}^{1200}\big(\widehat{f}(-6+0.01p)-f(-6+0.01p)\big)^2\times 0.01\Big),$$

where $\widehat{f}$ is one of the estimators. For each Monte Carlo iteration the corrected and oracle estimators are evaluated on the same trajectory. The results are reproduced in the following table.

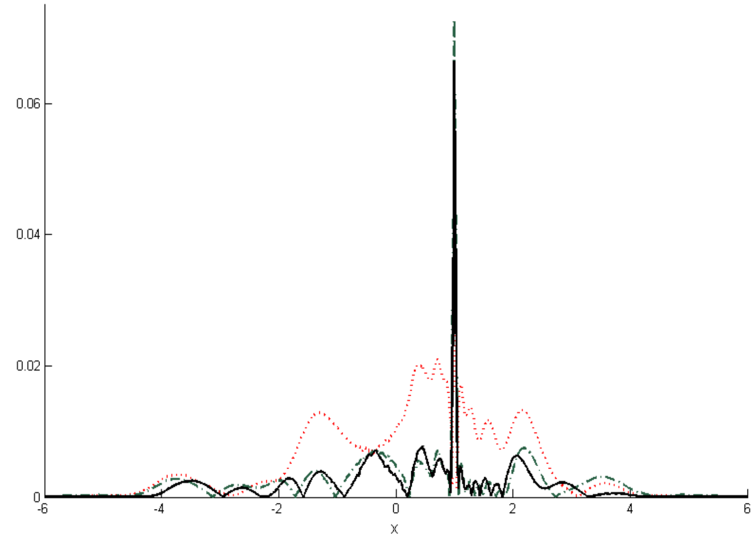

FIGURE 3.3 – Mean absolute error between the estimators and the true density (M=1000, $T = 10000$ and $\Delta = 0.1$) : the uncorrected (dotted red), the 1-corrected (dashed green) and the oracle (plain dark).

Estimator	Oracle	$K = 0$	$K = 1$	$K = 2$	$K = 3$
L_2 error ($\times 10^{-2}$)	0.11	0.18	0.13	0.13	0.13
Standard deviation ($\times 10^{-5}$)	0.35	0.44	0.44	0.44	0.44

This confirms that there is an actual gain in considering the estimator corrected at order 1 instead of the uncorrected one. In the following table we estimate the $\big(p_m(\Delta)\big)$ defined in Proposition 1.

Estimated quantity	$\widehat{p_1}$	$\widehat{p_2}$	$\widehat{p_3}$
Estimation	0.95	0.05	0.00
Standard deviation	0.0022	0.0022	0.0004

It turns out that without the correction we estimate the density f on a data set where 5% of the observations are realisations of a law which is not f. This explains why it is relevant to take them into account when estimating f. Considering more than 1 corrections is unnecessary as the L_2 losses get stable afterwards. The L_2 loss of the oracle is strictly lower than the loss of the estimator corrected at order K, even for large K. That difference is explained by the fact that to estimate the mth convolution power we do not use N_T data points but $N_{T,m} = \lfloor N_T/m \rfloor$. Therefore we do not loose in terms of rate of convergence, but we surely deteriorate the constants in comparison with the oracle. Numerical results are consistent with the theoretical results of Theorem 1 where we proved a rate of convergence for the estimator corrected at order K in

$$\max\left\{T^{-\alpha(s,p,\pi)}, \Delta_T^{K+1}\right\}.$$

Since $\alpha(s,p,\pi) \leq 1/2$, the rate decreases with K and becomes stable once $\Delta_T^{2K+2}T \leq \mathfrak{C}$. In the numerical example we took $T = 10000$ and $\Delta = 0.1$ thus $T\Delta^4 = 1$ which explains why in the example we did not observe improvements when correcting with K greater than 2.

3.4 Discussion

3.4.1 Relation to other works

A compound Poisson process is a pure jump Lévy process and can be studied accordingly using Lévy-Kintchine formula. Estimating the jump density f is then equivalent to estimating the Lévy measure since for compound Poisson process it is the product $\vartheta f(x)dx$. A possible estimation strategy in that case is to provide an estimator of the Fourier transform of the density. That strategy is quite different from the one introduced in this Chapter but is usually adopted when estimating the compound law of a compound Poisson process (see Figueroa-López [38], Comte and Genon-Catalot [22, 25] or Bec and Lacour [9]).

The nonparametric estimation of the Lévy measure from the discrete observation of a pure jump Lévy process from high frequency data (which corresponds to our microscopic regime $\Delta_T \to 0$) has been studied in great detail by Comte and Genon-Catalot [22, 25] and Figueroa-López [38]. In [38] the nonparametric estimation of the Lévy density is made via a sieve estimator. They show that it attains minimax rates of convergence for the L_2 loss uniformly over a class of Besov functions for a sampling size Δ_T such that –with our notation– $T\Delta_T \leq 1$. Comte and Genon-Catalot [22, 25] construct an adaptive nonparametric estimator of the Lévy measure, which attains minimax rates of convergence on Sobolev spaces for the L_2 loss for a sampling size Δ_T such that $T\Delta_T \leq 1$ (or $T\Delta_T^2 \leq 1$ under smoother assumptions). Bec and Lacour [9] obtained similar results when $T\Delta_T^2 \leq 1$. The statistical setting of [25] is more general since they estimate the Lévy measure from observations of a Lévy process with a Brownian component.

Our result is limited to the Poisson case contrary to Bec and Lacour [9], Comte and Genon-Catalot [22] and Figueroa-López [38] who worked on the larger class of pure jump Lévy processes. However in the case of a Poisson process we generalise them since we provide an adaptive density estimator which attains minimax rates of convergence, for the L_p loss, $p \geq 1$, uniformly over Besov balls for regime where Δ_T is polynomially slow. If Δ_T decays even slower, for instance logarithmically in T, we still have an upper bound for the rate of convergence of our estimator.

3.4.2 Possible extensions

In this Chapter we give an adaptive minimax procedure for the estimation of the compound density of a compound Poisson process in the microscopic regime. The same estimation problem in an intermediate regime, namely when the process is observed at a sampling rate $\Delta > 0$ fixed, has been studied in van Es *et al.* [103] and in the more general setting of Lévy processes by Comte and Genon-Catalot [24] and Reiß [83]. van Es *et al.* [103] provide a consistent kernel density estimator of the compound density of a compound Poisson process of known intensity. They also focus on the nonzero increments for the estimation, but sidestep the problem of the random number of data N_T by assuming that they have a sample of a given size.

The estimator corrected at order K presented here should extend to intermediate regime where $\Delta_T \to \Delta_\infty < 1$ and the rate of convergence given in Theorem 1 should generalise in

$$\max\left\{T^{-\alpha(s,p,\pi)}, \Delta_\infty^{K+1}\right\}.$$

An improvement of the results would be the estimation of the compound density of renewal reward processes, or Continuous Time Random Walk, where it is no longer imposed that the elapsed time between jumps is exponentially distributed. Then the Lévy property is lost, the increments of the renewal process are no longer independent nor identically distributed. An estimation strategy based on the Lévy-Kintchine formula is not possible. Such processes enable to model random phenomena where the elapse time between events is not memoryless ; they have many applications for instance in finance (see Meerschaert *et al.* [75]), in biology (see Fedotov *et al.* [35]) or for modelling earthquakes (see Helmstetter *et al.* [52]).

3.5 Proof of Theorem 1

In the sequel $\mathfrak{C}$ denotes a generic constant which may vary from line to line. Its dependencies may be indicated in the index.

3.5.1 Proof of part 1) of Theorem 1

Preliminary lemmas

To prove part 1) of Theorem 1 we apply the general results of Kerkyacharian and Picard [63]. For that we establish some technical lemmas.

Lemma 2. *If f belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$ then for $m \geq 1$, $\mathbf{P}_\Delta[f]^{\star m}$ also belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$.*

Proof of Lemma 2. It is straightforward to derive $\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{L_1(\mathbb{R})} = 1$. The remainder of the proof is a consequence of the following result : Let $f \in \mathcal{B}^s_{\pi\infty}$ and $g \in L_1$ we have

$$\|f \star g\|_{s\pi\infty} \leq \|f\|_{s\pi\infty}\|g\|_{L_1(\mathbb{R})}. \qquad (\diamondsuit)$$

To prove the ($\diamondsuit$) we use the following norm which is equivalent to the Besov norm (see [49])

$$\|\nu\|_{s\pi\infty} = \|\nu\|_{L_\pi(\mathbb{R})} + \|\nu^{(n)}\|_{L_\pi(\mathbb{R})} + \left\| \frac{w_\pi^2(\nu^{(n)}, t)}{t^a} \right\|_\infty \tag{3.17}$$

where $s = n + a$, $n \in \mathbb{N}$ and $a \in (0, 1]$, and w is the modulus of continuity

$$w_\pi^2(\nu, t) = \sup_{|h| \leq t} \left\| \mathbf{D}^h \mathbf{D}^h[\nu] \right\|_{L_\pi(\mathbb{R})},$$

where $\mathbf{D}^h[\nu](x) = \nu(x - h) - \nu(x)$. The result is a consequence of Young's inequality and elementary properties of the convolution product. We use the definition (3.17) of the norm and treat each term separately. First Young's inequality gives

$$\|f_1 \star f_2\|_{L_\pi(\mathbb{R})} \leq \|f_1\|_{L_\pi(\mathbb{R})} \|f_2\|_{L_1(\mathbb{R})}. \tag{3.18}$$

Then the differentiation property of the convolution product leads for $n \geq 1$ to

$$\left\| \frac{d^n}{dx^n}(f_1 \star f_2) \right\|_{L_\pi(\mathbb{R})} = \left\| \Big(\frac{d^n}{dx^n} f_1 \Big) \star f_2 \right\|_{L_\pi(\mathbb{R})} \leq \left\| \frac{d^n}{dx^n} f_1 \right\|_{L_\pi(\mathbb{R})} \|f_2\|_{L_1(\mathbb{R})}. \tag{3.19}$$

Finally translation invariance of the convolution product enables to get

$$\begin{aligned} \left\| \mathbf{D}^h \mathbf{D}^h[(f_1 \star f_2)^{(n)}] \right\|_{L_\pi(\mathbb{R})} &= \left\| (\mathbf{D}^h \mathbf{D}^h[f_1^{(n)}]) \star f_2 \right\|_{L_\pi(\mathbb{R})} \\ &\leq \left\| \mathbf{D}^h \mathbf{D}^h[f_1^{(n)}] \right\|_{L_\pi(\mathbb{R})} \|f_2\|_{L_1(\mathbb{R})}. \end{aligned} \tag{3.20}$$

Inequality ($\diamondsuit$) is then obtained by bounding (3.17) with (3.18), (3.19) and (3.20) lead to the result. To complete the proof of Lemma 2, we apply $m - 1$ times ($\diamondsuit$) which leads to

$$\forall m \in \mathbb{N} \setminus \{0\}, \qquad \left\| \mathbf{P}_\Delta[f]^{\star m} \right\|_{s\pi\infty} \leq \left\| \mathbf{P}_\Delta[f] \right\|_{s\pi\infty}.$$

The triangle inequality gives $\|\mathbf{P}_\Delta[f]^{\star m}\|_{s\pi\infty} \leq \|f\|_{s\pi\infty} \leq \mathfrak{M}$ which concludes the proof. □

Lemma 3. *Let $n \geq 1$ and $\Delta > 0$, then $(\mathbf{D}^\Delta X_{S_1}, \ldots, \mathbf{D}^\Delta X_{S_n})$ are independent and identically distributed and independent of N_T.*

Proof of Lemma 3. The idea of proof is the same as for the proof of the reject algorithm. The result is a consequence of

$$(S_i, i = 1, \ldots, n) \text{ and } (\mathbf{D}^\Delta X_{S_i}, i = 1 \ldots, n) \text{ are independent.} \tag{3.21}$$

Indeed if (3.21) holds with the fact that X is a compound Poisson process, we derive that $(\mathbf{D}^\Delta X_{S_1}, \ldots, \mathbf{D}^\Delta X_{S_n})$ are independent and identically distributed. And the independence between $(\mathbf{D}^\Delta X_{S_1}, \ldots, \mathbf{D}^\Delta X_{S_n})$ and N_T is deduced from (3.21) and

$$N_T = \sum_{i=1}^{\infty} \mathbb{1}_{\{S_i < \lfloor T\Delta^{-1} \rfloor\}}.$$

Then to prove the Lemma, we have to prove (3.21). Since the (S_i) are stopping times for the filtration $\mathcal{F}_n = \sigma\big(\mathbf{D}^{\Delta} X_i, i = 1, \dots, n\big)$, by the strong Markov property it is enough to prove the independence of S_i and $\mathbf{D}^{\Delta} X_{S_i}$ for all i. Notice that $S_i = S_{i-1} + A$ where A have geometric distribution with parameter $\mathbb{P}(\mathbf{D}^{\Delta} X_1 \neq 0)$ and S_{i-1} is independent of A and S_i by the strong Markov property. We have for any measurable function h

$$\begin{aligned}\mathbb{E}\big[h(\mathbf{D}^{\Delta} X_{S_i})\mathbb{1}_{\{S_i=n\}}\big] &= \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_n)\mathbb{1}_{\{\mathbf{D}^{\Delta} X_{S_{i-1}+1}=0,\dots,\mathbf{D}^{\Delta} X_{n-1}=0,\mathbf{D}^{\Delta} X_n \neq 0\}}\big] \\ &= \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1)\mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big]\mathbb{P}(\mathbf{D}^{\Delta} X_1)^{n-1}, \end{aligned} \tag{3.22}$$

using that the increments are independent and identically distributed. Moreover we have

$$\begin{aligned}\mathbb{E}\big[h(\mathbf{D}^{\Delta} X_{S_i})\big] &= \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{D}^{\Delta} X_1 = 0)^{k-1} \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1)\mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big] \\ &= \frac{\mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1)\mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big]}{\mathbb{P}(\mathbf{D}^{\Delta} X_1 = 0)}. \end{aligned} \tag{3.23}$$

Then from (3.22) and (3.23) we derive (3.21), which concludes the proof. □

Lemma 4. *Let $2^j \leq N_T$ then for all $m \in \mathbb{N} \setminus \{0\}$ and for $p \geq 1$ we have*

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p \big| N_T\big] \leq \mathfrak{C}_{p,m,\|g\|_{L_p(\mathbb{R})},\mathfrak{M},\vartheta} N_T^{-p/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (3.12) *and*

$$\gamma_{jk}^{(m)} = \int g_{jk}(y)\mathbf{P}_{\Delta}[f]^{\star m}(y)dy. \tag{3.24}$$

Proof of Lemma 4. The proof is obtained with Rosenthal's inequality : let $p \geq 1$ and let $(Y_1, \dots, Y_n)$ be independent random variables such that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\big[|Y_i|^p\big] < \infty$. Then there exists $\mathfrak{C}_p$ such that

$$\mathbb{E}\Big[\Big|\sum_{i=1}^{n} Y_i\Big|^p\Big] \leq \mathfrak{C}_p \Big\{\sum_{i=1}^{n} \mathbb{E}\big[|Y_i|^p\big] + \Big(\sum_{i=1}^{n} \mathbb{E}\big[|Y_i|^2\big]\Big)^{p/2}\Big\}. \tag{3.25}$$

The $\big(\mathbf{D}_m^{\Delta_T} X_{S_i}\big)$ are independent and identically distributed with common density $\mathbf{P}_{\Delta_T}[f]^{\star m}$ and $\mathbb{E}[\widehat{\gamma}_{jk}^{(m)}] = \gamma_{jk}^{(m)}$. Then $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is a sum of $N_{T,m} = \lfloor N_T/m \rfloor$ centered, independent and identically distributed random variables. It follows that

$$\begin{aligned}\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] &\leq 2^p 2^{jp/2} \int |g(2^j y - k)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}(y)dy \\ &= 2^p 2^{j(p/2-1)} \int |g(z)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}\Big(\frac{z+k}{2^j}\Big)dz \\ &\leq 2^p 2^{j(p/2-1)} \|g\|_{L_p(\mathbb{R})}^p \big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_{\infty}, \end{aligned}$$

where we made the substitution $z = 2^j x - k$. To control $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty$ we use the Sobolev embeddings (see [19, 30, 49])

$$\mathcal{B}^s_{\pi\infty} \hookrightarrow \mathcal{B}^{s'}_{p\infty} \quad \text{and} \quad \mathcal{B}^{s'}_{\pi\infty} \hookrightarrow \mathcal{B}^{s}_{\infty\infty}, \tag{3.26}$$

where $p > \pi$, $s\pi > 1$ and $s' = s - 1/\pi + 1/p$, it follows that

$$\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \leq \mathfrak{C}_{s,\pi}\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{\mathcal{B}^s_{\pi\infty(\mathcal{D})}}.$$

We deduce from Lemma 2 that $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \leq \mathfrak{C}_{s,\pi}\mathfrak{M}$. We get

$$\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] \leq 2^p 2^{j(p/2-1)}\|g\|^p_{L_p(\mathbb{R})}\mathfrak{M}$$

and $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{Delta_T} X_{S_i})\big|^2\big] \leq \mathfrak{M}$ since $\|g\|_2^2 = 1$.
Lemma 3 ensures that for all $n \geq 1$ the increments $\big(\mathbf{D}^{\Delta_T} X_{S_1}, \ldots, \mathbf{D}^{\Delta_T} X_{S_n}\big)$ are independent of N_T and then $N_{T,m}$. Indeed the $\big(\mathbf{D}^{\Delta_T} X_i, i = 1, \ldots, \lfloor T\Delta_T^{-1}\rfloor\big)$ are independent and identically distributed and the $\big(\mathbf{D}^{\Delta_T} X_{S_i}\big)$ are constructed with $S_i = \inf\big\{j > S_{i-1}, \mathbf{D}^{\Delta_T} X_j \neq 0\big\}$. Therefore we can apply Rosenthal's inequality conditional on N_T to $\widehat{\gamma}^{(m)}_{jk} - \gamma^{(m)}_{jk}$ and derive for $p \geq 1$

$$\mathbb{E}\big[\big|\widehat{\gamma}^{(m)}_{jk} - \gamma^{(m)}_{jk}\big|^p\big|N_T\big] \leq \mathfrak{C}_p\Big\{2^p\Big(\frac{2^j}{N_{T,m}}\Big)^{p/2-1}\|g\|^p_{L_p(\mathbb{R})}\mathfrak{M} + \mathfrak{M}^{p/2}\Big\}N_{T,m}^{-p/2}.$$

This concludes the proof. □

Lemma 5. *Choose j and c such that*

$$2^j N_T^{-1}\log\big(N_T^{1/2}\big) \leq 1 \text{ and } c^2 \geq \frac{16m}{3}\Big(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\Big).$$

For all $m \in \mathbb{N}\setminus\{0\}$ and $r \geq 1$ let $\kappa_r = cr$. We have

$$\mathbb{P}\Big(|\widehat{\gamma}^{(m)}_{jk} - \gamma^{(m)}_{jk}| \geq \frac{\kappa_r}{2}N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}\Big|N_T\Big) \leq N_T^{-r/2},$$

where $\widehat{\gamma}^{(m)}_{jk}$ is defined in (3.12) and $\gamma^{(m)}_{jk}$ in (3.24).

Proof of Lemma 5. The proof is obtained with Bernstein's inequality. Consider $Y_1, \ldots, Y_n$ independent random variables such that $|Y_i| \leq \mathfrak{A}$, $\mathbb{E}[Y_i] = 0$ and $b_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2]$. Then for any $\lambda > 0$,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n Y_i\Big| > \lambda\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2(b_n^2 + \frac{\lambda\mathfrak{A}}{3})}\Big). \tag{3.27}$$

For all $m \geq 1$, $\widehat{\gamma}^{(m)}_{jk} - \gamma^{(m)}_{jk}$ is a sum of $N_{T,m} = \lfloor N_T/m\rfloor$ centered independent and identically distributed random variables bounded by $2^{j/2}\|g\|_\infty$ and $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^2\big] \leq \mathfrak{M}$. Lemma 3 ensures that for all $n \geq 1$ the increments $\left(\mathbf{D}^{\Delta_T} X_{S_1}, \ldots, \mathbf{D}^{\Delta_T} X_{S_n}\right)$ are independent of N_T, we apply Bernstein's inequality conditional on N_T. We have

$$\begin{aligned}\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\Big| N_T\Big) \\ \leq & 2\exp\left(-\frac{\kappa_r^2 N_T^{-1}\log\left(N_T^{1/2}\right)N_{T,m}^2}{8\Big(N_{T,m}\mathfrak{M} + \frac{\kappa_r N_{T,m} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}2^{j/2}\|g\|_\infty}{6}\Big)}\right) \\ = & 2\exp\left(-\frac{c^2 r N_T^{-1} N_{T,m}}{8\Big(\mathfrak{M} + \frac{\kappa_r N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}2^{j/2}\|g\|_\infty}{6}\Big)} r\log\left(N_T^{1/2}\right)\right).\end{aligned}$$

Using that

$$mN_T^{-1}N_{T,m} = \frac{m}{N_T}\left\lfloor\frac{N_T}{m}\right\rfloor \geq \frac{3}{2},$$

for T large enough and $2^{j/2}N_T^{-1}\sqrt{\log\left(N_T^{1/2}\right)} \leq 1$ it follows that

$$\begin{aligned}\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\Big| N_T\Big) \\ \leq 2\exp\left(-\frac{3c^2 r}{16m\left(\mathfrak{M} + \frac{\kappa_r\|g\|_\infty}{6}\right)} r\log\left(N_T^{1/2}\right)\right).\end{aligned}$$

With $c^2 \geq \frac{16m}{3}\left(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\right)$ we get

$$\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\Big| N_T\Big) \leq N_T^{-r/2}.$$

The proof is complete. □

Completion of proof of part 1) of Theorem 1

Part 1) of Theorem 1 is a consequence of Lemma 2, 4, 5 and of the general theory of wavelet threshold estimators of [63]. It suffices to have conditions (5.1) and (5.2) of Theorem 5.1 of [63], which are satisfied –Lemma 4 and 5– with $c(T) = N_T^{-1/2}$ and $\Lambda_n = c(T)^{-1}$ (with the notations of [63]). We can now apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [63] to obtain the result.

3.5.2 Proof of part 2) of Theorem 1

Preliminary result

The result of part 1) of Theorem 1 was given conditional on N_T. To prove part 2) we replace N_T by its deterministic counterpart. We introduce the following result.

Proposition 2. *For all $r > 0$, there exist $1 \leq \mathfrak{C}_\vartheta < \infty$, where $\vartheta \to \mathfrak{C}_\vartheta$ is continuous, such that*

$$1/\mathfrak{C}_\vartheta T^{-r} \leq \mathbb{E}\big[N_T^{-r}\big] \leq \mathfrak{C}_\vartheta T^{-r}.$$

Proof of Proposition 2. We have

$$N_T = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}},$$

where

$$\mathbb{E}\big[\mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}}\big] = p(\Delta_T) = 1 - \exp(-\vartheta\Delta_T).$$

Introduce $Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}} - p(\Delta_T)$, the Y_i are centered independent and identically distributed random variables bounded by 2 and $\mathbb{E}[Y_i^2] \leq p(\Delta_T)$, it follows from Bernstein's inequality (3.27) that for $\lambda > 0$

$$\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| > \lambda\Big) \leq \exp\bigg(-\frac{\lfloor T\Delta_T^{-1}\rfloor\lambda^2}{2\big(p(\Delta_T) + \frac{2\lambda}{3}\big)}\bigg). \tag{3.28}$$

We choose $\lambda = p(\Delta_T)/2$, on the set $\big\{\big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\big| \leq \lambda\big\}$ we have

$$\lfloor T\Delta_T^{-1}\rfloor\frac{p(\Delta_T)}{2} \leq N_T \leq \lfloor T\Delta_T^{-1}\rfloor\frac{3p(\Delta_T)}{2}.$$

Moreover for Δ_T small enough we have that

$$\frac{\vartheta}{2} \leq p(\Delta_T) = 1 - \exp(-\vartheta\Delta_T) \leq \vartheta\Delta_T.$$

We have for all $\lambda > 0$

$$\mathbb{E}\big[N_T^{-r}\big] = \mathbb{E}\Big[N_T^{-r}\mathbb{1}_{\big\{|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)| > \lambda\big\}}\Big] + \mathbb{E}\Big[N_T^{-r}\mathbb{1}_{\big\{|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)| \leq \lambda\big\}}\Big].$$

Since for $r > 0$ the function $x \to x^{-r}$ is decreasing and $N_T \geq 1$ we have using (3.28) the upper bound

$$\begin{aligned}
\mathbb{E}\big[N_T^{-r}\big] &\leq \mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| > \frac{p(\Delta_T)}{2}\Big) + \Big(\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \\
&\leq \exp\bigg(-\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)^2}{8\big(p(\Delta_T) + \frac{p(\Delta_T)}{3}\big)}\bigg) + \Big(\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \\
&\leq \exp\Big(-\frac{3\vartheta}{64}T\Big) + \Big(\frac{T\vartheta}{4}\Big)^{-r}.
\end{aligned}$$

For the lower bound we have

$$\mathbb{E}\big[N_T^{-r}\big] \geq \Big(\frac{3\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \geq \Big(\frac{3T\vartheta}{2}\Big)^{-r}.$$

Then there exists $1 \leq \mathfrak{C}_\vartheta < \infty$ with $\vartheta \to \mathfrak{C}_\vartheta$ continuous such that

$$1/\mathfrak{C}_\vartheta T^{-r} \leq \mathbb{E}\big[N_T^{-r}\big] \leq \mathfrak{C}_\vartheta T^{-r}.$$

The proof is now complete. □

Completion of proof of part 2) of Theorem 1

To prove Theorem 1 we define the quantity for K in $\mathbb{N}$ and x in $\mathcal{D}$

$$\widehat{f}_{T,\Delta}^K(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\big(e^{\vartheta\Delta}-1\big)^m}{\vartheta\Delta} \widehat{P_{\Delta,m}}(x).$$

It is the estimator of f one would compute if ϑ were known. We decompose the L_p error as follows

$$\begin{aligned}\big(\mathbb{E}\big[\|\widetilde{f}_{T,\Delta_T}^K - f\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \leq & \big(\mathbb{E}\big[\|\widetilde{f}_{T,\Delta_T}^K - \widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \\ & + \big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K - f\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p},\end{aligned}$$

and control each term separately.

First we control $\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K - f\|_{L_p(\mathcal{D})}^p\big]$, using the triangle inequality we get

$$\begin{aligned}&\Big(\mathbb{E}\Big[\Big\|\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{\vartheta\Delta_T} \widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}^{-1}\big[\mathbf{P}_{\Delta_T}[f]\big]\Big\|_{L_p(\mathcal{D})}^p\Big]\Big)^{1/p} \\ &\leq \sum_{m=1}^{K+1} \frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{m\vartheta\Delta_T} \big(\mathbb{E}\big[\big\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \qquad (3.29) \\ &\quad + \sum_{m=K+2}^{\infty} \frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{m\vartheta\Delta_T} \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathbb{R})}. \qquad (3.30)\end{aligned}$$

To bound (3.29) we use part 1) of Theorem 1 in which the supremum is taken over the class $\{\mathbf{P}_{\Delta_T}[f]^{\star m} \in \mathcal{F}(s,\pi,\mathfrak{M})\}$. With the inclusion

$$\{\mathbf{P}_{\Delta_T}[f]^{\star m}, f \in \mathcal{F}(s,\pi,\mathfrak{M})\} \subset \mathcal{F}(s,\pi,\mathfrak{M})$$

and Proposition 2 applied with $r = \alpha(s,p,\pi)p > 0$, we deduce the upper bound for $m \geq 1$

$$\begin{aligned}\mathbb{E}\big[\big\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}^{-1}\big[\mathbf{P}_{\Delta_T}[f]\big\|_{L_p(\mathcal{D})}^p\big] &\leq \mathfrak{C}\mathbb{E}\big[N_T^{-\alpha(s,p,\pi)p}\big] \\ &\leq \mathfrak{C}T^{-\alpha(s,p,\pi)p}, \qquad (3.31)\end{aligned}$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K,\vartheta)$. To bound (3.30) Young's inequality and $\left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_1(\mathbb{R})}=1$ enable to get

$$\left\|\mathbf{P}_{\Delta_T}[f]^{\star m}\right\|_{L_p(\mathbb{R})}\leq\left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_p(\mathbb{R})}\quad\text{for } m\geq 1.$$

The triangle inequality leads to $\left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_p(\mathbb{R})}\leq\|f\|_{L_p(\mathbb{R})}$ and we use the Sobolev embeddings (3.26) to get $\|f\|_{L_p(\mathbb{R})}\leq\mathfrak{C}_{s,\pi,p}\mathfrak{M}$. We derive the upper bound

$$\begin{aligned}\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{\vartheta\Delta_T}&\left\|\mathbf{P}_{\Delta_T}[f]^{\star m}\right\|_{L_p(\mathbb{R})}\\&\leq\|f\|_{L_p(\mathbb{R})}\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{\vartheta\Delta_T}\\&\leq\mathfrak{C}_{K,\vartheta,\mathfrak{M}}{\Delta_T}^{K+1}.\end{aligned}\tag{3.32}$$

Thus from (3.31) and (3.32) we obtain

$$\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\left(\mathbb{E}\left[\left\|\widehat{f}_{T,\Delta_T}^K-f\right\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p}\leq\mathfrak{C}\max\left\{T^{-\alpha(s,p,\pi)},\Delta_T^{K+1}\right\},$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K,\vartheta)$. Since $\vartheta\to\mathfrak{C}$ is continuous we get for $p\geq 1$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\ \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\left(\mathbb{E}\left[\left\|\widehat{f}_{T,\Delta_T}^K-f\right\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p}\leq\mathfrak{C}_{K,\mathfrak{M}}\max\left\{T^{-\alpha(s,p,\pi)},\Delta_T^{K+1}\right\},$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K)$

We now control $\mathbb{E}\left[\|\widetilde{f}_{T,\Delta_T}^K-\widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\right]$ and use (3.15) to derive

$$\widetilde{f}_{T,\Delta_T}^K=\sum_{m=1}^{K+1}\frac{(-1)^m}{m}\frac{\left((1-\widehat{p}_T)^{-1}-1\right)^m}{\log(1-\widehat{p}_T)}\widehat{P_{\Delta_T,m}},$$

where $\widehat{P_{\Delta_T,m}}$ does not depend on ϑ (see (3.12)). Define

$$G_m(x)=\frac{((1-x)^{-1}-1)^m}{\log(1-x)}.$$

The triangle inequality leads to

$$\begin{aligned}&\left(\mathbb{E}\left[\|\widehat{f}_{T,\Delta_T}^K\big(\widehat{\vartheta}\big)-\widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p}\\&\qquad\leq\sum_{m=1}^{K+1}\left(\mathbb{E}\left[\|\big(G_m(\widehat{p}_T)-G_m(p(\Delta_T))\big)\widehat{P_{\Delta_T,m}}\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p},\end{aligned}$$

where $p(\Delta_T)$ verifies $p(\Delta_T)=1-e^{-\vartheta\Delta_T}\leq\mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$ since

$$0<1-e^{-\underline{\mathfrak{T}}\Delta_T}\leq 1-e^{-\vartheta\Delta_T}\leq 1-e^{-\overline{\mathfrak{T}}\Delta_T}<1.$$

Moreover, we have

$$G'_m(x) = \frac{mx^{m-1}}{(1-x)^{m+1}\log(1-x)} + \frac{x^m}{(1-x)^{m+1}\big(\log(1-x)\big)^2},$$

then for all $m \geq 1$ $xG'_m(x)$ is continuous over $(0,1/2]$ and converges to 0 when $x \to 0$. We deduce

$$\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta}) - \widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]^{1/p} \\ \leq \mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}},K}\Delta_T^{-1}\mathbb{E}\big[\big\|\big(\widehat{p}_T - p(\Delta_T)\big)\widehat{P_{\Delta_T,m}}\big\|^p_{L_p(\mathcal{D})}\big]^{1/p}.$$

Cauchy-Schwarz inequality leads to

$$\mathbb{E}\big[\big\|\big(\widehat{p}_T - p(\Delta_T)\big)\widehat{P_{\Delta_T,m}}\big\|^p_{L_p(\mathcal{D})}\big]^2 \\ \leq \mathbb{E}\Big[\big\|\big(\widehat{p}_T - p(\Delta_T)\big)\big\|^{2p}_{2p}\Big]\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|^{2p}_{L_{2p}(\mathcal{D})}\Big],$$

where using part 1) of Theorem 1 and that $N_T \geq 1$ we have

$$\begin{aligned}\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|^{2p}_{L_{2p}(\mathcal{D})}\Big] &\leq \mathbb{E}\Big[\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\|^{2p}_{L_{2p}(\mathcal{D})}\Big] + \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|^{2p}_{L_{2p}(\mathcal{D})} \\ &\leq \mathfrak{C}\mathbb{E}[N_T^{-2\alpha(s,p,\pi)p}] + \mathfrak{M}^{2p} \\ &\leq \mathfrak{C} \qquad (3.33)\end{aligned}$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi)$. We apply Rosenthal's inequality (3.25) to conclude the proof : $\widehat{p}_T - p(\Delta_T)$ is the sum of independent and identically distributed centered random variables

$$\big(Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T}X_i\neq 0\}} - p(\Delta_T), i \in \{1,\dots,\lfloor T\Delta_T^{-1}\rfloor\}\big)$$

where $\mathbb{E}[|Y_i|^{2p}] \leq 2^{2p}\mathbb{E}\big[\mathbb{1}^{2p}_{\{\mathbf{D}^{\Delta_T}X_i\neq 0\}}\big] \leq \mathfrak{C}_{p,\overline{\mathfrak{T}}}\Delta_T$ and $\mathbb{E}[|Y_i|^2] \leq \mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$. Rosenthal's inequality (3.25) gives

$$\mathbb{E}\big[\|\widehat{p}_T - p(\Delta_T)\|^{2p}_{2p}\big] \\ \leq \mathfrak{C}_{p,\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\lfloor T\Delta_T^{-1}\rfloor^{-2p}\big(\lfloor T\Delta_T^{-1}\rfloor\Delta_T + (\lfloor T\Delta_T^{-1}\rfloor\Delta_T)^p\big). \qquad (3.34)$$

It follows from (3.33) and (3.34) that

$$\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta}) - \widehat{f}^K_{T,\Delta_T}\|_{L_p(\mathcal{D})}\big]^{1/p} \\ \leq \mathfrak{C}\Delta_T^{-1}\lfloor T\Delta_T^{-1}\rfloor^{-1}\big(T^{1/(2p)} + T^{1/2}\big),$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,\underline{\mathfrak{T}},\overline{\mathfrak{T}},K)$. We deduce for $p \geq 1$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta}) - \widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p} \\ \leq \mathfrak{C}\big(T^{-(1-1/(2p))} + T^{-1/2}\big)$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,\underline{\mathfrak{T}},\overline{\mathfrak{T}},K)$ and which is negligible compared to $T^{-\alpha(s,p,\pi)}$ since $\alpha(s,p,\pi) \leq 1/2$. The proof of Theorem 1 is now complete.

3.6 Appendix

3.6.1 Proof of Proposition 1

Let $x \in \mathbb{R}$, we have by stationarity of the increments of the process X

$$\begin{aligned}\mathbb{P}(\mathbf{D}^{\Delta} X_{S_1} \leq x) &= \mathbb{P}(X_\Delta \leq x | X_\Delta \neq 0) \\ &= \sum_{m=0}^{\infty} \mathbb{P}(X_\Delta \leq x | R_\Delta = m, R_\Delta \neq 0)\mathbb{P}(R_\Delta = m) \\ &= \sum_{m=1}^{\infty} p_m(\Delta)\mathbb{P}(X_\Delta \leq x | R_\Delta = m)\end{aligned}$$

where $\mathbb{P}(X_\Delta \leq x | R_\Delta = m) = \int_{-\infty}^{x} f^{\star m}(y)dy$ for $m \geq 1$. It follows

$$\mathbb{P}(\mathbf{D}^{\Delta} X_{S_1} \leq x) = \int_{-\infty}^{x} \mathbf{P}_\Delta[f](y)dy.$$

Immediate computation give the expression of $p_m(\Delta)$. For the control of $p_1(\Delta)$ the assertion $p_1(\Delta) \leq 1$ is immediate since $p_1(\Delta)$ is a probability. Moreover we have

$$\exp\big(\vartheta\Delta\big) - 1 = \vartheta\Delta\Big(1 + \vartheta\Delta \sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta\big)^{m-2}}{m!}\Big),$$

where

$$g(\Delta) := \sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta\big)^{m-2}}{m!} = \frac{1}{\big(\vartheta\Delta\big)^2}\big(\exp\big(\vartheta\Delta\big) - 1 - \vartheta\Delta\big) \longrightarrow \frac{1}{2} \quad \text{as } \Delta \to 0.$$

Since g is continuous, there exists $\Delta_0 > 0$ such that for all $\Delta \leq \Delta_0$ we have $g(\Delta) \leq 1$. It follows for $\Delta \leq \Delta_0$ that

$$p_1(\Delta) \geq \frac{1}{1 + \vartheta\Delta} \geq 1 - \vartheta\Delta.$$

3.6.2 Proof of Lemma 1

Let $\mathbf{F}[f]$ denote the Fourier transform of f and take h such that $h = \mathbf{P}_\Delta[f]$. Using the one-to-one mapping between densities and their Fourier transform we show the relation for the Fourier transforms. The linearity of the Fourier transform and the relation $\mathbf{F}[f \star g] = \mathbf{F}[f]\mathbf{F}[g]$ give

$$\mathbf{F}[h] = \mathbf{F}\big[\mathbf{P}_\Delta[f]\big] = \frac{1}{e^{\vartheta\Delta} - 1} \sum_{m=1}^{\infty} \frac{\big(\vartheta\Delta\big)^m}{m!} \mathbf{F}[f]^m = \frac{\big(\exp(\vartheta\Delta\mathbf{F}[f]) - 1\big)}{e^{\vartheta\Delta} - 1},$$

from which we deduce

$$\mathbf{F}[f] = \frac{\log\left(1 + (e^{\vartheta\Delta} - 1)\mathbf{F}[h]\right)}{\vartheta\Delta} = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \frac{(e^{\vartheta\Delta} - 1)^m}{\vartheta\Delta} \mathbf{F}[h]^m$$

as $\left\|(e^{\vartheta\Delta} - 1)\mathbf{F}[h]\right\|_\infty < \left\|e^{\vartheta\Delta} - 1\right\|_\infty < 1$ holds for $\Delta \leq \log 2$. We take the inverse Fourier transform of the equality to obtain the result.

Chapitre 4

Estimation of a renewal reward process at microscopic scales

4.1 Introduction

4.1.1 Motivation and statistical setting

Renewal reward processes are pure jump processes used in many application fields, for instance in seismology (see Alvarez [3] or Helmstetter *et al.* [52]), to model rainfall (see Rodriguez-Iturbe *et al.* [88]) or in mathematical insurance and finance (see for instance Scalas *et al.* [93, 94] or Masolivier *et al.* [71]). Although many papers are devoted to the estimation of a discretely observed Lévy process (see for instance Bec and Lacour [9], Comte and Genon-Catalot [22, 25], Figueroa-López [38] and Chapter 3 for the high frequency case and Neumann and Reiß [83] and Comte and Genon-Catalot [24] for the low frequency one), to the knowledge of the author, little exists on the estimation of a discretely observed renewal reward process. Vardi [102] estimates the density of a renewal process without rewards from the continuous observation of several independent trajectories. In this Chapter we estimate the compound law of a renewal reward process when one trajectory is observed at a sampling rate that goes to 0 arbitrarily slowly.

Let $J_1, \ldots, J_i$ be nonnegative independent random variables where $J_2, \ldots, J_i$ are identically distributed. Define T_i the time of the ith jump as $T_i = J_1 + ... + J_i$, $i \geq 1$. The associated counting process or renewal process R is

$$R_t = \sum_{i=1}^{\infty} \mathbb{1}_{T_i \leq t}, \quad t \geq 0.$$

The Poisson process is a particular case of a renewal process, corresponding to exponentially distributed interarrivals $\big(J_i\big)$. That latter case excepted, R does not have independent increments and is usually not stationary *i.e.* for all positive t, h the law of $R_{t+h} - R_t$ depends on

t. Assume that the common distribution τ of the $\big(J_i\big)$ has finite expectation

$$\mu = \int_0^\infty t\tau(dt) < \infty,$$

define the distribution

$$\tau_0(x) = \frac{1 - \int_0^x \tau(dt)}{\mu}. \tag{4.1}$$

The process R is stationary if and only if J_1 has density τ_0 (see Lindvall [70] p.70). Define the renewal reward process X as

$$X_t = \sum_{i=1}^{R_t} \xi_i, \quad t \geq 0$$

where the $\big(\xi_i\big)$ are independent and identically distributed random variables, independent of the interarrivals $\big(J_i\big)$. Renewal reward processes also correspond to decoupled continuous time random walks.

Assume that we have discrete observations of the process X over $[0, T]$ at times $i\Delta$ for some $\Delta > 0$

$$\big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big). \tag{4.2}$$

We focus on the *microscopic regime* namely

$$\Delta = \Delta_T \to 0 \qquad \text{as} \ \ T \to \infty,$$

and work under the following assumption.

Assumption 1. *The law of the ξ_i has density f which is absolutely continuous with respect to the Lebesgue measure.*
The law of the J_i, $i \geq 2$ has density τ which is absolutely continuous with respect to the Lebesgue measure and J_1 has density τ_0.

The necessity of the last part of Assumption 1 is discussed in Section 4.5.

We denote by $\mathcal{F}(\mathbb{R})$ the space of densities with respect to the Lebesgue measure supported by $\mathbb{R}$. We investigate the nonparametric estimation of the density f on a compact interval $\mathcal{D}$ of $\mathbb{R}$ from the observations (4.2). To that end we use wavelet threshold density estimators and study their rate of convergence, uniformly over Besov balls, for the following loss function

$$\big(\mathbb{E}\big[\|\widehat{f} - f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}, \tag{4.3}$$

where $\widehat{f}$ is an estimator of f, $p \geq 1$ and $\|.\|_{L_p(\mathcal{D})}$ denotes L_p loss over the compact set $\mathcal{D}$. We do not assume the interarrival distribution τ to be known : it is a nuisance parameter.

We estimate f from the increments of X, which are dependent. By Assumption 1, on the event $\{X_{i\Delta} - X_{(i-1)\Delta} = 0\}$ no jump occurred between $(i-1)\Delta$ and $i\Delta$ so that the increment $X_{i\Delta} - X_{(i-1)\Delta}$ gives no information on f. In the microscopic regime $\Delta = \Delta_T \to 0$ many increments are zero, therefore to estimate f we focus on the nonzero increments. We denote by N_T their number over $[0, T]$. In that statistical context different difficulties arise ; the number of data N_T used for the estimation is random, the increments are dependent, but more importantly on the event $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the density of $X_{i\Delta} - X_{(i-1)\Delta}$ is not f. Indeed even if Δ is small there is always a positive probability that more than one jump occurred between $(i-1)\Delta$ and $i\Delta$. Conditional on $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the law of $X_{i\Delta} - X_{(i-1)\Delta}$ has density given by (see Proposition refPropDefOperator Du3 below)

$$\mathbf{P}_\Delta[f](x) = \sum_{m=1}^{\infty} \mathbb{P}\big(R_\Delta = m \big| R_\Delta \neq 0\big) f^{\star m}(x), \qquad \text{for } x \in \mathbb{R}, \tag{4.4}$$

where $\star$ is the convolution product and $f^{\star m} = f \star \ldots \star f$, m times. Hereafter Lemma 1 gives for Δ small enough

$$1 - 2\tau(0)\Delta \leq \mathbb{P}\big(R_\Delta = m \big| R_\Delta \neq 0\big) \leq 1. \tag{4.5}$$

We deduce from (4.5) the decomposition

$$\mathbf{P}_\Delta[f] = f + r(\Delta),$$

where $r(\Delta)$ is a deterministic remainder of the order of Δ. We will see in Theorem 1 that if $\Delta = \Delta_T$ goes to 0 fast enough, namely $T\Delta_T^2 = O(1)$ (up to logarithmic factor in T) $r(\Delta)$ is negligible and it is possible to estimate f with optimal rates by ignoring the remainder $r(\Delta)$. Otherwise, when there exists $0 < \delta < 1$ such that $T\Delta_T^2 = O(T^\delta)$ (up to logarithmic factors in T) the remainder $r(\Delta)$ is no longer negligible. The condition $\delta < 1$ ensures that Δ_T goes to 0 as T tends to infinity. In the sequel we distinguish two different regimes that will be treated separately.

– *Fast microscopic rates* when –up to logarithmic factors in T–
$$T\Delta_T^2 = O(1).$$
– *Slow microscopic rates* when there exists $0 < \delta < 1$ such that –up to logarithmic factors in T–
$$T\Delta_T^2 = O(T^\delta).$$

Since all the results of the Chapter are given up to logarithmic factors in T, fast and slow microscopic rates cover all vanishing behaviours for $\Delta = \Delta_T$. We try to answer the following question : Is it possible to construct an adaptive wavelet estimator of f in fast and slow microscopic regimes which is optimal ? Papers which estimate nonparametrically the Lévy measure from a discretely observed Lévy process attain optimal rate estimators only for fast microscopic rates (see for instance Bec and Lacour [9], Comte and Genon-Catalot [22, 24, 25] and Figueroa-López [38]).

4.1.2 Our Results

In Section 4.2 we estimate f in the fast microscopic regime, the estimation procedure is based on the approximation

$$f \approx \mathbf{P}_\Delta[f].$$

We construct an adaptive wavelet threshold density estimator from the observations (4.2). It achieves the minimax rate of convergence which is $T^{-\alpha(s,p,\pi)}$ if f is of regularity s measured with the L_π norm, $\pi > 0$, and where $\alpha(s,\pi,p) \leq 1/2$ (see (4.16) hereafter). That procedure does not depend on the unknown interarrival density τ apart from Assumptions 1 and 3 hereafter.

In Section 4.3 we estimate f in the slow microscopic regime, the estimation procedure is the analogue of the one used in Chapter 3. The starting point is that

$$f = \mathbf{P}_\Delta^{-1}\big[\mathbf{P}_\Delta[f]\big],$$

and we proceed in two steps to estimate f. The first step is the computation of the inverse of the operator $\mathbf{P}_\Delta$ defined in (4.4). That step can be referred as decompounding as introduced in Buchmann and Grübel [16] or van Es *et al.* [103]. That inverse cannot be explicitly calculated, contrary to Chapter 3, but can be approached using a fixed point method. Indeed f is a fixed point of the operator

$$\mathbf{H}_{\Delta,f} : h \to \mathbf{P}_\Delta[f] + h - \mathbf{P}_\Delta[h]$$

which is a contraction if h and f verifies suitable smoothness properties (see Proposition 2 below). If we denote by $\circ$ the composition product, then the Banach fixed point theorem guarantees that for K in $\mathbb{N}$ and $p \geq 1$,

$$\big\|\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big] - f\big\|$$

is small in a sense that we precise later. Next we observe that the Taylor expansion of order K in Δ of $\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big]$ takes the form

$$\sum_{m=1}^{K+1} l_m(\Delta)\mathbf{P}_\Delta[f]^{\star m}, \tag{4.6}$$

where the $\big(l_m(\Delta)\big)$ depend on the unknown interarrival density τ (see Proposition 1 below). If τ is described by an unknown parameter $\vartheta \in \mathbb{R}$ then $l_m(\Delta) = l_m(\Delta, \vartheta)$ is estimated by plugging an estimator of ϑ.

The second step consists in estimating the densities $\mathbf{P}_\Delta[f]^{\star m}$, for $m = 1, \ldots, K+1$. For that we focus on the N_T nonzero increments which have density $\mathbf{P}_\Delta[f]$. The difficulty here is that we have N_T dependent observations where N_T is a random sum of dependent variables. The dependency of the increments is treated using that at each renewal times the renewal process forgets its past. To cope with the randomness of N_T, we prove that N_T/T concentrates for T large enough around a deterministic limit using Bernstein type inequalities for dependent data (see Lemma 3 in Section 4.6 and Dedecker *et. al.* [31]). In Theorem 2 we show that wavelet threshold estimators of $\mathbf{P}_\Delta[f]^{\star m}$ attain a rate of convergence –up to logarithmic factors– in $T^{-\alpha(s,\pi,p)}$. We inject those estimators into (4.6) and obtain an estimator of f that we call *estimator corrected at order* K.

The study of the rate of convergence of the estimator corrected at order K requires to control two distinct error terms. A deterministic one due the first step which is the error made when approximating f by (4.6). And a statistical one due to the replacement of the $\mathbf{P}_\Delta[f]^{\star m}$ by estimators in the second step. The deterministic error decreases when K increases. We choose K sufficiently large for the deterministic error term to be negligible compared to the statistical one. We give in Theorem 2 an upper bound for the rate of convergence of the estimator corrected at order K which is in –up to logarithmic factors–

$$\max\{T^{-\alpha(s,\pi,p)}, \Delta_T^{K+1}\}.$$

Since $\alpha(s,\pi,p) \leq 1/2$ if there exists K_0 such that

$$T\Delta_T^{2K_0+2} \leq 1,$$

the estimator corrected at order K_0 attains the optimal rate.

Remark 1. *There is a slight difference of methodology between fast and slow microscopic rates to estimate f. For fast rates we do not have to estimate the density τ to build an estimator of f. But to build a minimax estimator of f in slow microscopic regimes, we need to estimate τ as well, in that case we assume that τ depends on an unknown parameter.*

The Chapter is organised as follows. In Section 4.2 we give an adaptive minimax estimator of f in the fast microscopic regime. In that Section we also define wavelet functions and Besov spaces that are used for the estimation and describe the law of the increments. Those results are also used in Section 4.3 where we give an adaptive minimax estimator of f in the slow microscopic regime. In both Sections 4.2 and 4.3 we give upper bounds for the rate of convergence of the estimator of f for the L_p loss defined in (4.3), $p \geq 1$, uniformly over Besov balls. In Section 4.4, a numerical example illustrates the behavior of the estimators of f introduced in Sections 4.2 and 4.3. Finally Section 4.6 is dedicated to the proofs.

4.2 Estimation of f in the fast microscopic regime

4.2.1 Preliminary on Besov spaces and wavelet thresholding

For the estimation, we use wavelet threshold density estimators and study their performance uniformly over Besov balls. In this paragraph we reproduce some classical results on

Besov spaces, wavelet bases and wavelet threshold estimators (see Cohen [19], Donoho *et al.* [30] or Kerkyacharian and Picard [63]) that we use in the next sections.

Wavelets and Besov spaces

We describe the smoothness of a function with Besov spaces on $\mathcal{D}$. We recall here some well documented results on Besov spaces and their connection to wavelet bases (see Cohen [19], Donoho *et al.* [30] or Kerkyacharian and Picard [63]). Let $\left(\psi_\lambda\right)_\lambda$ be a regular wavelet basis adapted to the domain $\mathcal{D}$. The multi-index λ concatenates the spatial index and the resolution level $j = |\lambda|$. Set $\Lambda_j := \{\lambda, |\lambda| = j\}$ and $\Lambda = \cup_{j\geq -1}\Lambda_j$, for f in $L_p(\mathbb{R})$ we have

$$f = \sum_{j\geq -1}\sum_{\lambda\in\Lambda_j}\langle f, \psi_\lambda\rangle\psi_\lambda, \tag{4.7}$$

where $j = -1$ incorporates the low frequency part of the decomposition and $\langle ., \rangle$ denotes the usual L_2 inner product. For $s > 0$ and $\pi \in (0, \infty]$ a function f belongs to the Besov space $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ if the norm

$$\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} := \|f\|_{L_\pi(\mathcal{D})} + \|f^{(n)}\|_{L_\pi(\mathcal{D})} + \left\|\frac{w^2_\pi(f^{(n)}, t)}{t^a}\right\|_{L_\infty(\mathcal{D})} \tag{4.8}$$

is finite, where $s = n + a$, $n \in \mathbb{N}$ and $a \in (0, 1]$, w is the modulus of continuity defined by

$$w^2_\pi(f, t) = \sup_{|f|\leq t}\left\|\mathbf{D}^h\mathbf{D}^h[f]\right\|_{L_\pi(\mathcal{D})}$$

and $\mathbf{D}^h[f](x) = f(x-h) - f(x)$. Equivalently we can define Besov space in term of wavelet coefficients (see Härdle *et. al.* [49] p. 123), f belongs to the Besov space $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ if the quantity

$$\sup_{j\geq -1} 2^{j(s+1/2-1/\pi)}\Big(\sum_{\lambda\in\Lambda_j}|\langle f, \psi_\lambda\rangle|^\pi\Big)^{1/\pi}$$

is finite, with usual modifications if $\pi = \infty$.

We need additional properties on the wavelet basis $\left(\psi_\lambda\right)_\lambda$, which are listed in the following assumption.

Assumption 2. *For $p \geq 1$,*

– *We have for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1}2^{|\lambda|(p/2-1)} \leq \|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{|\lambda|(p/2-1)}.$$

– *For some $\mathfrak{C} > 0$, $\sigma > 0$ and for all $s \leq \sigma$, $J \geq 0$, we have*

$$\left\|f - \sum_{j\leq J}\sum_{\lambda\in\Lambda_j}\langle f, \psi_\lambda\rangle\psi_\lambda\right\|_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{-Js}\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}. \tag{4.9}$$

– *If $p \geq 1$, for some $\mathfrak{C} \geq 1$ and for any sequence of coefficients $\big(u_\lambda\big)_{\lambda\in\Lambda}$,*

$$\mathfrak{C}^{-1}\Big\| \sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda\Big\|_{L_p(\mathcal{D})} \leq \Big\|\Big(\sum_{\lambda\in\Lambda} |u_\lambda\psi_\lambda|^2\Big)^{1/2}\Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C}\Big\| \sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda\Big\|_{L_p(\mathcal{D})}. \tag{4.10}$$

– *For any subset $\Lambda_0 \subset \Lambda$ and for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1} \sum_{\lambda\in\Lambda_0} \|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \int_{\mathcal{D}} \Big(\sum_{\lambda\in\Lambda_0} |\psi_\lambda(x)|^2\Big)^{p/2} \leq \mathfrak{C} \sum_{\lambda\in\Lambda_0} \|\psi_\lambda\|^p_{L_p(\mathcal{D})}. \tag{4.11}$$

Property (4.9) ensures that definition (4.8) of Besov spaces matches the definition in terms of linear approximation. Property (4.10) ensures that $\big(\psi_\lambda\big)_\lambda$ is an unconditional basis of L_p and (4.11) is a super-concentration inequality (see Kerkyacharian and Picard [63] p. 304 and p. 306).

Wavelet threshold estimator

Let (ϕ, ψ) be a pair of scaling function and mother wavelet that generate a basis $\big(\psi_\lambda\big)_\lambda$ satisfying Assumption 2 for some $\sigma > 0$. We rewrite (4.7)

$$f = \sum_{k\in\Lambda_0} \alpha_{0k}\phi_{0k} + \sum_{j\geq 1}\sum_{k\in\Lambda_j} \beta_{jk}\psi_{jk},$$

where $\phi_{0k}(\bullet) = \phi(\bullet - k)$ and $\psi_{jk}(\bullet) = 2^{j/2}\psi(2^j \bullet - k)$ and

$$\alpha_{0k} = \int \phi_{0k}(x)f(x)dx$$
$$\beta_{jk} = \int \psi_{jk}(x)f(x)dx.$$

For every $j \geq 0$, the set Λ_j has cardinality 2^j and incorporates boundary terms that we choose not to distinguish in the notation for simplicity. An estimator of a function f is obtained when replacing the (α_{0k}) and (β_{jk}) by estimated values. In the sequel we uses (γ_{jk}) to design either (α_{0k}) or (β_{jk}) and (g_{jk}) for the wavelet functions (ϕ_{0k}) or (ψ_{jk}).

We consider classical hard threshold estimators of the form

$$\widehat{f}(\bullet) = \sum_{k\in\Lambda_0} \widehat{\alpha_{0k}}\phi_{0k}(\bullet) + \sum_{j=1}^{J}\sum_{k\in\Lambda_j} \widehat{\beta_{jk}}\mathbb{1}_{\{|\widehat{\beta_{jk}}|\geq\eta\}}\psi_{jk}(\bullet),$$

where $\widehat{\alpha_{0k}}$ and $\widehat{\beta_{jk}}$ are estimators of α_{0k} and β_{jk}, J and η are respectively the resolution level and the threshold, possibly depending on the data. Thus to construct $\widehat{f}$ we have to specify estimators $(\widehat{\gamma_{jk}})$ of the (γ_{jk}) and the coefficients J and η.

4.2.2 Construction of the estimator

Assume that we have $\lfloor T\Delta^{-1}\rfloor$ discrete data at times $i\Delta$ for some $\Delta > 0$ of the process X

$$\big(X_\Delta, \dots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big).$$

Introduce the increments

$$\mathbf{D}^\Delta X_i = X_{i\Delta} - X_{(i-1)\Delta}, \qquad \text{for } i = 1, \dots, \lfloor T\Delta^{-1}\rfloor,$$

where $X_0 = 0$. By Assumption 1, they are identically distributed but not independent.

Proposition 1. *The density of the increment $\mathbf{D}^\Delta X_1$ is*

$$\big(1 - p(\Delta)\big)\delta_0 + p(\Delta)\mathbf{P}_\Delta[f]$$

where δ_0 is the Dirac delta function, $p(\Delta) = \mathbb{P}(R_\Delta \neq 0)$ and

$$\mathbf{P}_\Delta[f] = \sum_{m=1}^{\infty} p_m(\Delta) f^{\star m}, \tag{4.12}$$

where $\star$ is the convolution product, $f^{\star m}$ is f convoluted m times and

$$p_m(\Delta) = \mathbb{P}\big(R_\Delta = m | R_\Delta \neq 0\big).$$

It is straightforward to verify that the operator $\mathbf{P}_\Delta$ is a mapping from $\mathcal{F}(\mathbb{R})$ to itself. The following Lemma gives a polynomial control of the coefficients $\big(p_m(\Delta)\big)$. It is widely used in Sections 4.2 and 4.3 and does not depend on the rate at which Δ_T decays to 0.

Lemma 1. *Assume $\tau(0) > 0$ and let Δ_0 be such that*

$$\int_0^{\Delta_0} \tau(t)dt \leq \frac{1}{2} \quad \text{and} \quad \sup_{t\in[0,\Delta_0]} \tau(t) \leq 2\tau(0).$$

For all $\Delta \leq \Delta_0$ we have

$$1 - 2\tau(0)\Delta \leq p_1(\Delta) \leq 1,$$

and for $m \geq 2$

$$0 \leq p_m(\Delta) \leq 2\frac{\big(2\tau(0)\big)^{m-1}}{m!}\Delta^{m-1},$$

where the $\big(p_m(\Delta)\big)$ are defined in Proposition 1.

Remark 2. *The assumption $\tau(0)>0$ in Lemma 1 ensures that the given inequalities are sharp. In the Poisson case it is always true since $\tau(0)$ is the positive intensity. In the renewal case we may have $\tau(0)=0$, if so two cases must be distinguished. The first one is when τ has infinitely many derivatives null at 0; it is the case if τ is bounded away from 0. Then straightforward computations give for any K in $\mathbb{N}$: $p_1(\Delta)=1+O\big(\Delta^K\big)$, thus the procedure of Section 4.2 enables to achieve optimal rates even in slow microscopic regimes. It is not the purpose of this Chapter. The second case is $\tau(0)=0$ but there exists l_0 in $\mathbb{N}$ such that $\tau^{(l_0)}(0)>0$, then Lemma 1 can be adapted replacing $\tau(0)$ by $\tau^{(l_0)}(0)$ and Δ by Δ^{l_0}. In the sequel we assume that $\tau(0)>0$ and leave to the reader the changes to be made when $\tau(0)=0$.*

In this Section we consider the regimes for which $\Delta=\Delta_T$ is such that $T\Delta_T^2=O(1)$, up to logarithmic factors in T. To estimate f, we use the approximation $\mathbf{P}_{\Delta_T}[f]\approx f$. It is equivalent to consider that nonzero increments are realisations of f. We construct wavelet threshold density estimators of $\mathbf{P}_\Delta[f]$ from the observations

$$\big(\mathbf{D}^\Delta X_i, i=1,\ldots,\lfloor T\Delta^{-1}\rfloor\big).$$

Define the wavelet coefficients

$$\widehat{\gamma}_{jk}=\frac{1}{N_T}\sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} g_{jk}\big(\mathbf{D}^\Delta X_i\big)\mathbb{1}_{\left\{\mathbf{D}^\Delta X_i\neq 0\right\}}. \tag{4.13}$$

Let $\eta>0$ and $J\in\mathbb{N}\setminus\{0\}$, the estimator $\widehat{P_\Delta}$ of $\mathbf{P}_\Delta[f]$ is for x in $\mathcal{D}$

$$\widehat{P_\Delta}(x)=\sum_k\widehat{\alpha}_{0k}\phi_{0k}(x)+\sum_{j=0}^J\sum_k\widehat{\beta}_{jk}\mathbb{1}_{\left\{|\widehat{\beta}_{jk}|\geq\eta\right\}}\psi_{jk}(x). \tag{4.14}$$

Definition 1. *We define $\widehat{f}_{T,\Delta}$ an estimator of f for x in $\mathcal{D}$ as*

$$\widehat{f}_{T,\Delta}(x)=\widehat{P_\Delta}(x). \tag{4.15}$$

4.2.3 Convergence rates

We estimate densities f which verify a smoothness property in term of Besov balls

$$\mathcal{F}(s,\pi,\mathfrak{M})=\big\{f\in\mathcal{F}(\mathbb{R}),\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\leq\mathfrak{M}\big\},$$

where $\mathfrak{M}$ is a positive constant. We are interested in estimating f on the compact interval $\mathcal{D}$, that is why we only impose that its restriction to $\mathcal{D}$ belongs to a Besov ball.

Assumption 3. *Assume that there exist $(\mathfrak{A},\mathfrak{a},\mathfrak{g})$ positive constants such that*

$$\tau(x)\leq\mathfrak{A}\exp\big(-\mathfrak{a}x^{\mathfrak{g}}\big),\qquad\forall x\in[0,\infty).$$

Assumption 3 is a technical condition which ensures that τ has moments of all order. It is used in the proofs to replace N_T/T by its asymptotic deterministic limit. Compactly supported densities and densities with subexponential tails satisfies Assumption 3.

Theorem 1. *We work under Assumptions 1, 2 and 3, let Δ_T be such that $T\Delta_T^2 = O(1)$ up to logarithmic factors in T. Let $\pi > 0$, $\sigma > s > 1/\pi$, $p \geq 1 \wedge \pi$ and $\widehat{P_{\Delta_T}}$ be the wavelet threshold estimator of $\mathbf{P}_{\Delta_T}[f]$ on $\mathcal{D}$ constructed from (ϕ, ψ) and defined in (4.14). Take J such that*

$$2^J T^{-1} \log\left(T^{1/2}\right) \leq 1,$$

and

$$\eta = \kappa T^{-1/2}\sqrt{\log\left(T^{1/2}\right)},$$

for some $\kappa > 0$. Let

$$\alpha(s,p,\pi) = \min\left\{\frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2\big(s+1/2-1/\pi\big)}\right\}. \tag{4.16}$$

1) The estimator $\widehat{P_{\Delta_T}}$ verifies for large enough T and sufficiently large $\kappa > 0$

$$\sup_{\mathbf{P}_{\Delta_T}[f]\in\mathcal{F}(s,\pi,\mathfrak{M})} \Big(\mathbb{E}\big[\big\|\widehat{P_{\Delta_T}} - \mathbf{P}_{\Delta_T}[f]\big\|_{L_p(\mathcal{D})}^p\big]\Big)^{1/p} \leq \mathfrak{C}T^{-\alpha(s,p,\pi)},$$

up to logarithmic factors in T and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, \mu$.
2)The estimator $\widehat{f}_{T,\Delta_T}$ defined in (4.15) verifies for T large enough, sufficiently large $\kappa > 0$ and any positive constants $\underline{\mathfrak{a}} < \overline{\mathfrak{a}}$

$$\sup_{(\mu,\tau(0))\in[\underline{\mathfrak{a}},\overline{\mathfrak{a}}]^2} \; \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})} \Big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T} - f\|_{L_p(\mathcal{D})}^p\big]\Big)^{1/p} \leq \mathfrak{C}T^{-\alpha(s,p,\pi)},$$

up to logarithmic factors in T, where $\mu = \int t\tau(t)dt$ and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{a}}$ and $\overline{\mathfrak{a}}$.

The proof of Theorem 1 is postponed to Section 4.6.2. Theorem 1 guarantees that when $\Delta = \Delta_T$ tends rapidly to 0, namely $T\Delta_T^2 = O(1)$, the approximation $f \approx \mathbf{P}_{\Delta_T}[f]$ enables to achieve minimax rates of convergence (see Section 4.5). The estimator does not depend on τ.

4.3 Estimation of f in the slow microscopic regime

In this Section we consider the regimes for which there exists $0 < \delta < 1$ with $T\Delta_T^2 = O(T^\delta)$, up to logarithmic factors in T.

4.3.1 Construction of the estimator

We construct the estimator corrected at order K, following the estimation procedure described in Section 4.1.2.

Construction of the inverse

Define the space

$$\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N}) = \Big\{h, \|h\|_{L_1(\mathcal{D})} \leq \mathfrak{O}, \|h\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \mathfrak{N}\Big\},$$

where $\mathfrak{O}$ is any constant strictly greater than 1 and $\mathfrak{N}$ is a positive constant strictly greater than $\mathfrak{M}$. The space $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ is a subset of $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ which is a Banach space if equipped with the Besov norm (4.8).

First we approach the inverse of $\mathbf{P}_\Delta$ with a fixed point method. Consider the mapping $\mathbf{H}_{\Delta,f}$ defined for h in $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ by

$$\mathbf{H}_{\Delta,f}[h] := \mathbf{P}_\Delta[f] + h - \mathbf{P}_\Delta[h]. \tag{4.17}$$

We immediately verify that f is a fixed point : $\mathbf{H}_{\Delta,f}[f] = f$. The constraints $1 < \mathfrak{O}$ and $\mathfrak{M} < \mathfrak{N}$ ensure that if f is in $\mathcal{F}(s,\pi,\mathfrak{M})$, then $\mathbf{H}_{\Delta,f}[h]$ sends elements of $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ into itself (see Proposition 2). The following Proposition guarantee that the definition of the operator (4.17) matches the assumptions of the Banach fixed point theorem.

Proposition 2. *The following properties hold.*
1) Let $\pi \geq 1$*, the space* $\big(\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N}), \|.\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\big)$ *is a closed set of a Banach space and is then complete.*
2) The mapping $\mathbf{H}_{\Delta,f}$ *sends elements of* $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ *into itself and is a contraction. For all* $h_1, h_2 \in \mathcal{H}(s,\pi,\mathfrak{O})$ *we have that*

$$\big\|\mathbf{H}_{\Delta,f}[h_1] - \mathbf{H}_{\Delta,f}[h_2]\big\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \mathfrak{K}(\Delta)\|h_1 - h_2\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})},$$

where

$$\mathfrak{K}(\Delta) = 2\mathfrak{O}(e^{2\tau(0)\Delta} - 1) + 2\tau(0)\Delta. \tag{4.18}$$

Moreover since $\Delta_T \to 0$ *we have*

$$\mathfrak{K}(\Delta_T) \leq \mathfrak{C}\Delta_T < 1 \tag{4.19}$$

for some positive constant $\mathfrak{C}$ *depending on* $\tau(0)$ *and* $\mathfrak{O}$*.*

Proposition 2 enables to apply the Banach fixed point theorem ; we derive that f is the unique fixed point of $\mathbf{H}_{\Delta,f}$ and from any initial point h_0 in $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ we have

$$\big\|f - \mathbf{H}^{\circ K}_{\Delta,f}[h_0]\big\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \to 0 \quad \text{as} \quad K \to \infty,$$

where $\circ$ stands for the composition product and $\mathbf{H}^{\circ K}_{\Delta,f}$ is $\mathbf{H}_{\Delta,f} \circ \ldots \circ \mathbf{H}_{\Delta,f}$, K times. We choose $h_0 = \mathbf{P}_\Delta[f]$ as a starting point (Lemma 2 in Section 4.6 ensures that $\mathbf{P}_\Delta[f]$ belongs to $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$).

Proposition 3. *Let $\pi \geq 1$ and define the operator $\mathbf{L}_{\Delta,K}$ as the Kth degree Taylor polynomial of $\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big]$ in Δ. It verifies for $p \geq 1$*

$$\Big\|\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big] - \mathbf{L}_{\Delta,K}\Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C}\Delta^{K+1} \tag{4.20}$$

where $\mathfrak{C}$ is a positive constant depending on $\tau(0)$, $\mathfrak{M}$ and $\mathfrak{O}$. Moreover we have for T large enough

$$\mathbf{L}_{\Delta,K} = \sum_{m=1}^{K+1} l_m(\Delta)\mathbf{P}_\Delta[f]^{\star m}, \tag{4.21}$$

where for $m = 1, \dots, K+1$ we have $|l_m(\Delta)| \leq \mathfrak{C}\Delta^{m-1}$ where $\mathfrak{C}$ is a positive constant that depends on $\tau(0)$ and K.

Construction of estimators of the $\mathbf{P}_\Delta[f]^{\star m}$

Consider the increments $\big(\mathbf{D}^\Delta X_i = X_{i\Delta} - X_{(i-1)\Delta}, i = 1, \dots, \lfloor T\Delta^{-1}\rfloor\big)$ introduced earlier and define the nonzero ones using

$$\begin{aligned} S_1 &= \inf\big\{j, \mathbf{D}^\Delta X_j \neq 0\big\} \wedge \lfloor T\Delta^{-1}\rfloor \\ S_i &= \inf\big\{j > S_{i-1}, \mathbf{D}^\Delta X_j \neq 0\big\} \wedge \lfloor T\Delta^{-1}\rfloor \quad \text{for } i \geq 1, \end{aligned}$$

where S_i is the random index of the ith jump. Let

$$N_T = \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} \mathbb{1}_{\{\mathbf{D}^\Delta X_i \neq 0\}}$$

the random number of nonzero increments observed over $[0,T]$. By Assumption 1, on the event $\{\mathbf{D}^\Delta X_i = 0\}$, no jump occurred between $(i-1)\Delta$ and $i\Delta$. In the microscopic regime when $\Delta = \Delta_T \to 0$ as T goes to infinity many increments are null and convey no information about f, hence for the estimation of f we focus on the nonzero ones

$$\big(\mathbf{D}^\Delta X_{S_1}, \dots, \mathbf{D}^\Delta X_{S_{N_T}}\big).$$

They are identically distributed of density given by (4.12) ; Lemma 1 still applies.

We construct wavelet threshold density estimators of the $K+1$ first convolution powers of $\mathbf{P}_\Delta[f]$; define the wavelet coefficients for $m \geq 1$

$$\widehat{\gamma}_{jk}^{(m)} = \frac{1}{N_{T,m}} \sum_{i=1}^{N_{T,m}} g_{jk}\Big(\mathbf{D}_m^\Delta X_{S_i}\Big), \tag{4.22}$$

where $N_{T,m} = \lfloor N_T/m \rfloor \geq 1$ for large enough T and

$$\mathbf{D}_m^{\Delta} X_{S_i} = \mathbf{D}^{\Delta} X_{S_i} + \mathbf{D}^{\Delta} X_{S_{N_{T,m}+i}} + \cdots + \mathbf{D}^{\Delta} X_{S_{(m-1)N_{T,m}+i}}.$$

Let $\eta > 0$ and $J \in \mathbb{N} \setminus \{0\}$, define $\widehat{P_{\Delta,m}}$ the estimator of $\mathbf{P}_{\Delta}[f]^{\star m}$ over $\mathcal{D}$ for $m \geq 1$

$$\widehat{P_{\Delta,m}}(x) = \sum_k \widehat{\alpha}_{0k}^{(m)} \phi_{0k}(x) + \sum_{j=0}^{J} \sum_k \widehat{\beta}_{jk}^{(m)} \mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{(m)}| \geq \eta\right\}} \psi_{jk}(x), \quad x \in \mathcal{D}. \tag{4.23}$$

As mentioned earlier τ is a nuisance that needs to be estimated. To simplify the problem, we make the following parametric assumption on τ.

Assumption 4. *Assume there exists ϑ in Θ a compact subset of $\mathbb{R}$ such that*

$$\tau(x) = \tau_1(x, \vartheta), \qquad \forall x \in [0, \infty),$$

where τ_1 is known, $\tau_1(0, \vartheta) > 0$ and $\vartheta \to \tau_1(., \vartheta)$ is C^1 . Assume there exists q from Θ to $[0,1]$, invertible, such that $q(\vartheta) = \mathbb{P}(R_{\Delta} \neq 0)$ and whose inverse q^{-1} is bounded.

Assumption 4 enables to estimate the unknown coefficients $\big(p_m(\Delta)\big)$ and $\big(l_m(\Delta)\big)$, and to compute the estimator of f defined hereafter.

Definition 2. *Let $\widehat{f}_{T,\Delta}^{K}$ be the estimator corrected at order K defined for K in $\mathbb{N}$ and x in $\mathcal{D}$ as*

$$\widehat{f}_{T,\Delta}^{K}(x) = \sum_{m=1}^{K+1} l_m(\Delta, \widehat{\vartheta_T}) \widehat{P_{\Delta,m}}(x), \tag{4.24}$$

where

$$\widehat{\vartheta_T} = q^{-1}\Big(\frac{1}{\lfloor T\Delta^{-1} \rfloor} \sum_{i=1}^{\lfloor T\Delta^{-1} \rfloor} \mathbb{1}_{\mathbf{D}^{\Delta} X_i \neq 0}\Big)$$

and the $l_m(\Delta, \vartheta)$ are defined in Proposition 3.

Remark 3. *When Δ_T satisfies $T\Delta_T^2 = O(1)$, $\widehat{f}_{T,\Delta}^{0}$ defined in (4.24) with $K = 0$ and $\widehat{f}_{T,\Delta}$ defined in (4.15) coincides. Indeed, when $K = 0$ we have $\mathbf{L}_{\Delta,0} = \mathbf{P}_{\Delta}[f]$ (see Proposition 3).*

4.3.2 Convergence rates

Theorem 2. *We work under Assumptions 1, 2, 4 and 3 and assume that there exists $0 < \delta < 1$ such that*

$$T\Delta_T^2 = O(T^{\delta}),$$

up to logarithmic factors in T. Let $\pi \geq 1$, $\sigma > s > 1/\pi$, $p \geq 1$ and $\widehat{P_{\Delta_T,m}}$ be the threshold wavelet estimator of $\mathbf{P}_{\Delta_T}[f]^{\star m}$ on $\mathcal{D}$ constructed from (ϕ, ψ) and defined in (4.23). Take J such that

$$2^J T^{-1} \log\left(T^{1/2}\right) \leq 1,$$

and

$$\eta = \kappa T^{-1/2} \sqrt{\log\left(T^{1/2}\right)},$$

for some $\kappa > 0$.
1) For $m \geq 1$ the estimator $\widehat{P_{\Delta_T,m}}$ of $\mathbf{P}_{\Delta_T}[f]^{\star m}$ verifies for sufficiently large $\kappa > 0$

$$\sup_{\mathbf{P}_{\Delta_T}[f]^{\star m} \in \mathcal{F}(s,\pi,\mathfrak{M})} \left(\mathbb{E}\left[\left\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\right\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p} \leq \mathfrak{C} T^{-\alpha(s,p,\pi)},$$

up to logarithmic factors in T, where $\alpha(s,p,\pi)$ is defined in (4.16) and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi$ and ϑ.
2) The estimator corrected at order K $\widehat{f}_{T,\Delta}^K$ for $K \in \mathbb{N}$ defined in (4.24) verifies for T large enough, sufficiently large $\kappa > 0$ and any compact set $\Theta \subset \mathbb{R}$

$$\sup_{\vartheta \in \Theta} \sup_{f \in \mathcal{F}(s,\pi,\mathfrak{M})} \left(\mathbb{E}\left[\|\widehat{f}_{T,\Delta}^K - f\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p} \leq \mathfrak{C} \max\left(T^{-\alpha(s,p,\pi)}, \Delta_T^{K+1}\right),$$

up to logarithmic factors in T and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi$ and K.

The proof of Theorem 2 is postponed to Section 4.6.4. Since $\alpha(s,p,\pi) \leq 1/2$, Theorem 2 ensures that whenever Δ_T and T are polynomially related it is always possible to find K_0 such that the estimator corrected at order K_0 achieves the minimax rate of convergence (see Section 4.5). If Δ_T decays slower than any power of $1/T$, for instance if it decreases logarithmically with T, the estimator corrected at order K still provide a consistent estimator of f.

4.4 A numerical example

In this Section we illustrate the results of Theorems 1 and 2. In both cases we compare the performances of our estimator with an oracle : the wavelet estimator we would compute in the idealised framework where all the jumps are observed

$$\widehat{f}^{Oracle}(x) = \sum_k \widehat{\alpha}_{0k}^{Oracle} \phi_{0k}(x) + \sum_{j=0}^{J} \sum_k \widehat{\beta}_{jk}^{Oracle} \mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{Oracle}| \geq \eta\right\}} \psi_{jk}(x),$$

where

$$\widehat{\alpha}_{0k}^{Oracle} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i) \quad \text{and} \quad \widehat{\beta}_{jk}^{Oracle} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i),$$

R_T being the value of the renewal process R at time T and (ξ_i) the jumps. The parameters J and η as well as the wavelet bases (ϕ, ψ) are the same for all the estimators.

We consider a renewal process with a $Beta(1, \vartheta)$ interarrival density τ. We have $\vartheta = 3$, the first shape parameter is set to 1 to ensure the condition $0 < \tau_1(0, \vartheta) < \infty$. We estimate the compound law given by $f(x) = (1-a)f_1(x) + af_2(x)$, where f_1 is the uniform distribution over $[-2, 2]$ and f_2 is a Laplace with location parameter 1 and scale parameter 0.5, we take $a = 0.5$. We estimate the mixture f on $\mathcal{D} = [-10, 10]$ with the estimator corrected at order K for different values of K and study the results with the L_2 error. We also compare them with the oracle $\widehat{f}^{Oracle}$. Wavelet estimators are based on the evaluation of the first wavelet coefficients, to perform those we use Symlets 4 wavelet functions and a resolution level $J = 10$. Moreover we transform the data in an equispaced signal on a grid of length 2^L with $L = 8$, it is the binning procedure (see Härdle *et al.* [49] Chap. 12). The threshold is chosen as in Theorems 1 and 2. The estimators we obtain take the form of a vector giving the estimated values of the density f on the uniform grid $[-10, 10]$ with mesh 0.01. We use the wavelet toolbox of Matlab.

4.4.1 Illustration in the fast microscopic case

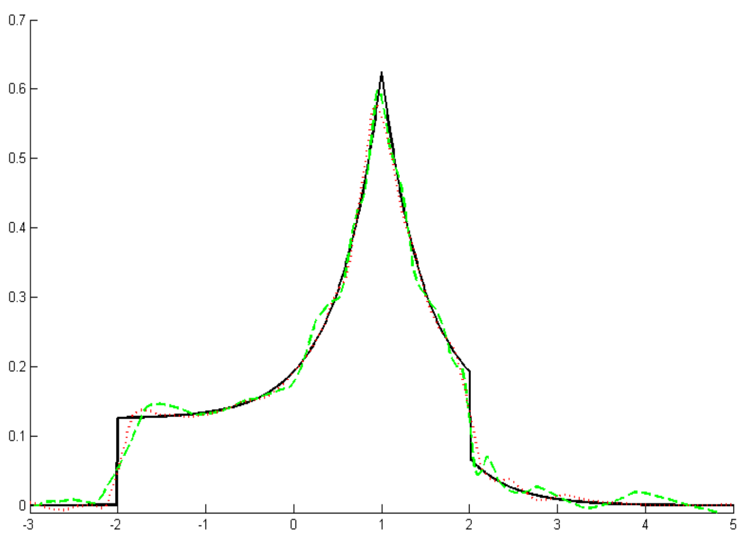

FIGURE 4.1 – Estimators of the density f (plain dark) for $T = 10000$ and $\Delta = 0.01$: the oracle (dotted red) and the estimator $\widehat{f}_{T,\Delta}$ (dashed green).

In this case we choose $\Delta = T^{-1/2}$. Figure 4.1 represents the estimator $\widehat{f}_{T,\Delta}$ of Definition 1 and the oracle. The estimators are evaluated on the same trajectory. They are quite hard to distinguish, what is confirmed by the comparison of their L_2 losses.

We approximate the L_2 errors by Monte Carlo. For that we compute $M = 1000$ times each estimator (for $T = 10000$ and $\Delta = 0.01$) and approximate the L_2 loss by

$$\frac{1}{M}\sum_{i=1}^{M}\Big(\sum_{p=0}^{2000}\big(\widehat{f}(-10+0.01p)-f(-10+0.01p)\big)^2\times 0.01\Big).$$

For each Monte Carlo iteration the estimators are evaluated on the same trajectory. The results are reproduced in the following table.

Estimator	Oracle	$\widehat{f}_{T,\Delta}$
L_2 error ($\times 10^{-2}$)	0.19	0.20
Standard deviation ($\times 10^{-5}$)	0.45	0.46

4.4.2 Illustration in the slow microscopic case

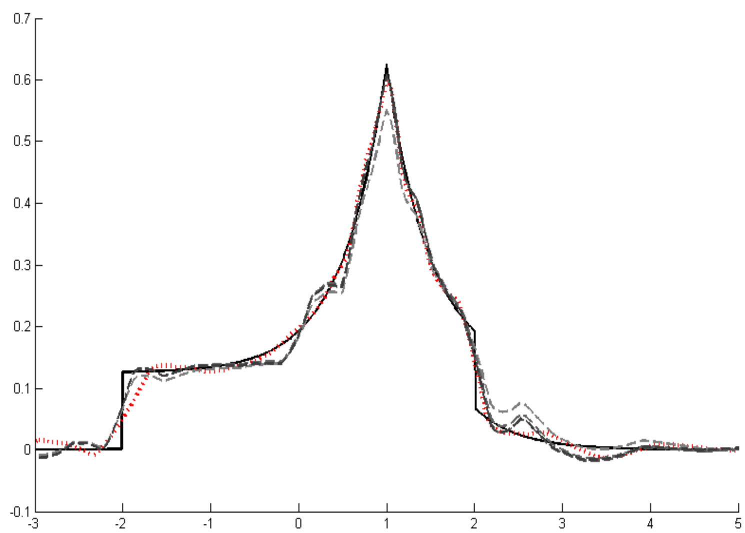

FIGURE 4.2 – Estimators of the density f (plain dark) for $T = 10000$ and $\Delta = 0.1$: the oracle (dotted red) and the estimator $\widehat{f}^K_{T,\Delta}$ for $K = 0, 1, 2, 3$ (dashed light to dark grey).

We now study the behaviour of the estimator corrected at order K for different values of K. We choose $T = 10000$ and $\Delta = 0.1$. In that case $T\Delta^2$ is large but $T\Delta^4$ is 1. According to Theorem 2 we should observe that the estimator corrected at order 2, behaves as the oracle. Figure 4.2 represents the estimators $\widehat{f}^K_{T,\Delta}$ defined in Definition 2 for $K \in \{0, 1, 2, 3\}$ and the oracle. The estimators are evaluated on the same trajectory. They all manage to reproduce the shape of the density f, and graphically apart from the estimator corrected at order 0 they are difficult to distinguish. We compare their L_2 losses in the following tabular.

Estimator	Oracle	$K=0$	$K=1$	$K=2$	$K=3$
L_2 error ($\times 10^{-2}$)	0.19	0.52	0.30	0.30	0.30
Standard deviation ($\times 10^{-5}$)	0.43	0.78	0.75	0.75	0.75

This confirms that there is an actual gain in considering the estimator corrected at order 1 instead of the uncorrected one. In the following table we estimate the $\big(p_m(\Delta)\big)$ defined in Proposition 1.

Estimated quantity	$\widehat{p_1}$	$\widehat{p_2}$	$\widehat{p_3}$
Estimation	0.85	0.13	0.01
Standard deviation ($\times 10^{-3}$)	0.92	0.74	0.16

It turns out that making no correction is equivalent to estimate a density on a data set where 15% of the observations are realisations of a law which is not target. This explains why it is relevant to take them into account when estimating f. Considering more than 1 or 2 corrections is unnecessary as the L_2 losses get stable afterwards. The L_2 loss of the oracle is strictly lower than the loss of the estimator corrected at order K, even for large K. That difference is explained by the fact that to estimate the mth convolution power we do not use N_T data points but $N_{T,m} = \lfloor N_T/m \rfloor$. Therefore we do not loose in terms of rate of convergence, but we surely deteriorate the constants in comparison with the oracle.

4.5 Discussion and Conclusion

Attainable rates. Without loss of generality, assuming T is an integer if we observe T independent realisations of the density f, it is possible to achieve the minimax rates of convergence $T^{-\alpha(s,\pi,p)}$ (see for instance Donoho *et al.* [30]). When the process X is continuously observed over $[0,T]$, we have R_T independent and identically distributed realisations of f. Moreover for T large enough, the elementary renewal theorem guarantees that R_T is of the order of T (see for instance Lindvall [70]). It follows that the estimators of f given in Sections 4.2 and 4.3 enables to attain the minimax rates of convergence of an experiment where X is continuously observed.

Comparison with a previous work. The results of this Chapter are the generalisation to the renewal reward case of Chapter 3 ; a compound Poisson process is a particular renewal reward process and Theorems 1 and 2 enable to recover the results of Chapter 3. However in this Chapter we do not have an explicit formula for the estimator corrected at order K but only a construction method. In the Poisson case it is much more simpler to apply the results of Chapter 3.

Extension to the case where Δ is fixed. We established Theorem 2 for Δ_T vanishing to 0. Since the approximation of the inverse depends only on the fact that $\mathbf{H}_{\Delta,f}$ is a contraction, the method remains valid for Δ's such that $\mathfrak{K}(\Delta)$ defined in (4.18) is strictly lower than 1. Which means that we can expand the results to cases where Δ does not go to 0 but satisfies $\mathfrak{K}(\Delta) < 1$. The value of the maximum value Δ_1 satisfying the former inequality depends on $\tau(0)$ and $\mathfrak{O}$ but is not only determined by (4.18). Another hidden condition on Δ have to be satisfied for $\mathbf{H}_{\Delta,f}$ to send elements of $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ into itself. Then to find Δ_1 one has to solve an optimisation program with constraints to find Δ_1 and $\mathfrak{O}$ giving the maximum coverage for Δ. To get an idea of the value of Δ_1 we use the function `NMaximize` of `Mathematica` and find that one should take $\mathfrak{O} = 1.645$ and $\Delta_1 = 0.071/\tau(0) > 0$, which is positive. The results of Theorem 2 should generalise to for all $\Delta_T \to \Delta_\infty$ such that $\Delta_\infty < \Delta_1$ and for $K \in \mathbb{N}$ the rate of convergence for the estimator corrected at order K is bounded by

$$\max\left\{T^{-\alpha(s,p,\pi)}, \Delta_\infty^{K+1}\right\}.$$

However to achieve suitable rates theoretically one should consider larger K, therefore the dependency in K in the constants need to be handled carefully. In practice for $T = 10000$ and $\Delta = 0.1$ considering $K = 2$ appears sufficient to have $T^{-\alpha(s,p,\pi)}$ predominant compared to Δ^{K+1}.

Discussion on Assumptions 1 and 4. In the present Chapter we made two simplifying assumptions on the interarrival density τ. First we assume that J_1 was distributed according to τ_0 to work with a process with stationary increments. In fact if τ has finite expectation this assumption is not necessary since asymptotically the process has stationary increments (see Lindvall [70]). The second assumption is that τ is described by a 1-dimensional parameter ϑ. Generalising the result to a d-dimensional parameter should be possible at small cost, but removing all parametric assumption on τ would demand to solve a nonstandard nonparametric program for τ from the observations (4.2) : observations (4.2) only give access to truncated values of realisations of τ spaced of more than Δ. Then the problem of estimating τ from (4.2) should be considered separately.

Other generalisations. We constructed in the microscopic regime an adaptive minimax estimator of the jump density of a renewal reward process. The methodology presented here should adapt to any process defined similarly to X but whose counting process has stationary increments and manageable dependencies. We consider in the present Chapter a renewal counting measure since we are interested in expanding the methodology to other regimes of Δ, namely when $\Delta = \Delta_T$ tends to a constant (intermediate regime) or to infinity (macroscopic regime). The macroscopic regime is of special interest since the observed process presents diffusive or anomalous asymptotic behaviour determined by the laws f and τ (see for instance Meerschaert and Scheffler [73, 74] or Kotulski [66]) and many applications have a model based on a macroscopically observed renewal reward processes. For instance in physics where they are used to model particle motion (see Watkins and Credgington [108] or Cuppen *et al.* [27]), in biology to model the proliferation of tumor cells (see Fedotov and Iomin [35]) or lipid granule motion (see Jeon *et al.* [60]), they are also used to model records (see Sabhapandit [90]).

4.6 Proofs

In the sequel $\mathfrak{C}$ denotes a constant which may vary from line to line.

4.6.1 Preliminaries

In this Section we introduce two technical lemmas used in to prove Theorems 1 and 2.

Lemma 2. *If f belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$ then for $m\geq 1$, $\mathbf{P}_\Delta[f]^{\star m}$ also belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$.*

To prove Theorem 1, we use Lemma 2 for $m=1$, and for Theorem 2 we apply Lemma 2 for $m\in\{1,\ldots,K+1\}$.

Proof of Lemma 2. It is straightforward to derive $\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{L_1(\mathbb{R})}=1$. The remainder of the proof is a consequence of the following result : Let $f\in\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ and $g\in L_1$ we have

$$\|f\star g\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\leq\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\|g\|_{L_1(\mathbb{R})}. \tag{4.25}$$

To prove (4.25) we use the definition of the Besov norm (4.8) ; the result is a consequence of Young's inequality and elementary properties of the convolution product. First Young's inequality gives

$$\|f_1\star f_2\|_{L_\pi(\mathbb{R})}\leq\|f_1\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}. \tag{4.26}$$

Then the differentiation property of the convolution product leads for $n\geq 1$ to

$$\left\|\frac{d^n}{dx^n}(f_1\star f_2)\right\|_{L_\pi(\mathcal{D})}=\left\|\Big(\frac{d^n}{dx^n}f_1\Big)\star f_2\right\|_{L_\pi(\mathbb{R})}\leq\left\|\frac{d^n}{dx^n}f_1\right\|_{L_\pi(\mathcal{D})}\|f_2\|_{L_1(\mathbb{R})}. \tag{4.27}$$

Finally translation invariance of the convolution product enables to get

$$\begin{aligned}\left\|\mathbf{D}^h\mathbf{D}^h[(f_1\star f_2)^{(n)}]\right\|_{L_\pi(\mathcal{D})}&=\left\|(\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}])\star f_2\right\|_{L_\pi(\mathcal{D})}\\&\leq\left\|\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}]\right\|_{L_\pi(\mathcal{D})}\|f_2\|_{L_1(\ R)}.\end{aligned} \tag{4.28}$$

Inequality (4.25) is then obtained by bounding $\|f\star g\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}$ using (4.26), (4.27) and (4.28). To complete the proof of Lemma 2, we apply $m-1$ times (4.25) which leads to

$$\forall m\in\mathbb{N}\setminus\{0\},\qquad\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\leq\left\|\mathbf{P}_\Delta[f]\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}.$$

The triangle inequality gives $\|\mathbf{P}_\Delta[f]^{\star m}\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\leq\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}\leq\mathfrak{M}$ which concludes the proof. □

The estimators of the convolution powers of $\mathbf{P}_\Delta[f]$ depend on N_T which is random and depends on the $\mathbf{D}_m^\Delta X_i$.

Lemma 3. *Work under Assumption 3 and let $p(\Delta) = \mathbb{P}(R_\Delta \neq 0)$ and Δ_1 be such that $\int_0^{\Delta_1} \tau(x)dx \leq \frac{1}{2}$. Then for all $\lambda > 0$ and $\Delta \leq \Delta_1$ we have*

$$\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta^{-1}\rfloor} - p(\Delta)\Big| > \lambda\Delta\Big) \leq \exp\Big(- \mathfrak{C}\sqrt{T}\Delta\Big),$$

where $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \mathfrak{g}, \mu, \lambda$.

Remark 4. *The result of Lemma 3 is interesting only in slow microscopic regimes where $\sqrt{T}\Delta \to \infty$.*

Proof of Lemma 3. We have

$$\frac{N_T}{\lfloor T\Delta^{-1}\rfloor} - p(\Delta) = \frac{1}{\lfloor T\Delta^{-1}\rfloor} \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} Y_i,$$

where

$$Y_i = \mathbb{1}_{\{\mathbf{D}^\Delta X_i \neq 0\}} - p(\Delta), \quad i = 1, \ldots, \lfloor T\Delta^{-1}\rfloor$$

are centered random variables, bounded by $M = 1 - p(\Delta)$ and such that

$$\mathbb{E}[Y_i^2] \leq p(\Delta).$$

To show the result, we apply Theorem 4.5 of Dedecker *et. al.* [31] which is a Bernstein-type inequality for dependent data. We have to verify conditions (4.4.16) and (4.4.17) of Theorem 4.5 of [31]. With their notation, condition (4.4.16) ensures that for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ such that

$$1 \leq s_1 \leq \ldots \leq s_u \leq t_1 \leq \ldots \leq t_v \leq \lfloor T\Delta^{-1}\rfloor$$

we have

$$\big|Cov(Y_{s_1} \ldots Y_{s_u}, Y_{t_1} \ldots Y_{t_v})\big| \leq K^2 M^{u+v-2} uv\rho(t_1 - s_u),$$

for some positive constant K and a nonincreasing function ρ satisfying (4.4.17) namely

$$\sum_{s=0}^{\infty} (s+1)^k \rho(s) \leq L_1 L_2^k (k!)^\nu, \quad \forall k \geq 0,$$

where L_1, L_2 and ν are positive constants.

Since X is a renewal process, $Y_{s_1} \ldots Y_{s_u}$ and $Y_{t_1} \ldots Y_{t_v}$ are independent if there exists r such that $s_u < r < t_1$ and $Y_r = 1 - p(\Delta)$ *i.e* there is a jump between Y_{s_u} and Y_{t_1}. For the covariance to be nonzero it is necessary that no jump occurred between $s_u\Delta$ and $(t_1-1)\Delta$. Let $s = t_1 - s_u - 1$ using that R is stationary we get an upper bound for ρ

$$\rho(t_1 - s_u) \leq \mathbb{P}\big(R_{(t_1-1)\Delta} - R_{s_u\Delta} = 0\big) = \mathbb{P}\big(R_{s\Delta} = 0\big) = \int_{s\Delta}^{\infty} \tau_0(x)dx \tag{4.29}$$

which decreases with s. Moreover since the Y_i are centered and bounded by $M \leq 1$ we have by Cauchy-Schwarz and $\mathbb{E}[Y_i^2] \leq p(\Delta)$

$$\big|Cov(Y_{s_1}\dots Y_{s_u}, Y_{t_1}\dots Y_{t_v})\big| \leq \big|Cov(Y_{s_1}, Y_{t_v})\big| \leq \sqrt{\mathbb{E}\big[Y_{s_1}^2\big]\mathbb{E}\big[Y_{t_v}^2\big]} \leq p(\Delta).$$

We deduce that condition (4.4.16) is fulfilled with $K = p(\Delta)^{1/2}$ and the nonincreasing sequence ρ.

Next we show that ρ satisfies (4.4.17), using Assumption 3 and (4.29) we get for $s \geq 1$

$$\rho(s) \leq \frac{1}{\mu}\int_{s\Delta_T}^{\infty}\big(1 - \int_0^x \tau(t)dt\big)dx \leq \mathfrak{C}\exp\big(-\mathfrak{a}(s\Delta)^{\mathfrak{g}'}\big),$$

where $\mathfrak{g} < \mathfrak{g}'$ and $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \mu, \mathfrak{g}$. Which leads to for $k \geq 0$

$$\begin{aligned}\sum_{s=1}^{\infty} s^k\rho(s) &\leq \sum_{s=1}^{1/\Delta} s^k + \mathfrak{C}\sum_{s=1/\Delta}^{\infty} s^k \exp\big(-\mathfrak{a}(s\Delta)^{\mathfrak{g}'}\big)\\ &\leq \Delta^{-(k+1)} + \mathfrak{C}\Delta^{-k}\sum_{s'=1}^{\infty} s'^k \exp\big(-\mathfrak{a}(s')^{\mathfrak{g}'}\big)\\ &\leq \mathfrak{C}\Delta^{-(k+1)}\end{aligned} \tag{4.30}$$

where $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \mathfrak{g}, \mu$, condition (4.4.17) follows with $L_1 = \mathfrak{C}\Delta^{-1}$, $L_2 = \Delta^{-1}$ and $\nu = 0$.

We can now apply Theorem 4.5 of [31] which gives for all $\lambda > 0$

$$\begin{aligned}&\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta^{-1}\rfloor} - p(\Delta)\Big| > \lambda\Delta\Big)\\ &\quad\leq 2\exp\bigg(-\frac{\lfloor T\Delta^{-1}\rfloor^2\Delta^2\lambda^2}{2\big(\lfloor T\Delta^{-1}\rfloor p(\Delta) + (\lfloor T\Delta^{-1}\rfloor\Delta\lambda)^{3/2}\sqrt{2^5\Delta^{-2}}\big)}\bigg).\end{aligned}$$

Finally we denote by F the cumulative density function of τ and use that

$$1 - p(\Delta) = \mathbb{P}(J_1 \geq \Delta) = \frac{1}{\mu}\int_0^{\Delta}\big(1 - F(u)\big)du,$$

since J_1 has distribution (4.1), and derive that there exists $\Delta_1 > 0$ such that $F(\Delta_1) \leq \frac{1}{2}$ and for all $\Delta \leq \Delta_1$ we have

$$\frac{\Delta}{2\mu} \leq 1 - p(\Delta) \leq \frac{\Delta}{\mu}. \tag{4.31}$$

We derive for $\Delta \leq \Delta_1$

$$\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta^{-1}\rfloor} - p(\Delta)\Big| > \lambda\Delta\Big) \leq \exp\Big(-\mathfrak{C}\sqrt{T}\Delta\Big),$$

where $\mathfrak{C}$ depends on $\lambda, \mathfrak{A}, \mathfrak{a}, \mathfrak{g}, \mu$. The proof is now complete. □

4.6.2 Proof of Theorem 1

Preliminary

In the estimator the quantity N_T appears, but we are in a regime where Lemma 3 cannot be applied as such. We introduce the following Lemma.

Lemma 4. *Let Δ_T be such that $T\Delta_T^2 = O(1)$, there exist $\mathfrak{c}$ and $\delta > 0$ such that for all $r > 0$*

$$\mathbb{E}\Big[\Big(\tfrac{(1-p(\Delta_T))\lfloor T\Delta_T^{-1}\rfloor}{N_T}\Big)^r\Big] \leq \mathfrak{c} \tag{4.32}$$

$$\mathbb{P}\Big(\tfrac{(1-p(\Delta_T))\lfloor T\Delta_T^{-1}\rfloor}{N_T} > \mathfrak{c}\Big) \leq \mathfrak{C}\exp\big(-T^\delta\big), \tag{4.33}$$

where $p(\Delta_T)$ is defined in Proposition 1, $\mathfrak{c}$ and $\mathfrak{C}$ are positive constants that depend on $\mathfrak{U}$, $\mathfrak{a}$, $\mathfrak{g}$, μ and r.

Proof of Lemma 4. Notice that if Δ_T is such that $T\Delta_T^2 = O(1)$, there exist $\widetilde{\Delta}_T$ such that for large enough T we have $\Delta_T \leq \widetilde{\Delta}_T$ and $T\Delta_T^2 = O(T^\delta)$ for some $\delta \in (0,1)$. The proof is a consequence of Lemma 3 applied to $\widetilde{N}_T$ which verify $\widetilde{N}_T \leq N_T$ for all T. □

Proof of part 1) of Theorem 1

To prove part 1) of Theorem 1 we apply the general results of Kerkyacharian and Picard [63]. For that we establish some technical lemmas.

Lemma 5. *Let $2^j \leq T$, then for $p \geq 1$ we have*

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk} - \gamma_{jk}\big|^p\big] \leq \mathfrak{C}_{p,\|g\|_{L_p(\mathbb{R})},\mathfrak{M}} T^{-p/2},$$

where $\widehat{\gamma}_{jk}$ is defined in (4.13) *and*

$$\gamma_{jk} = \int g_{jk}(y)\mathbf{P}_\Delta[f](y)dy. \tag{4.34}$$

Proof of Lemma 5. The proof is obtained with Rosenthal's inequality : let $p \geq 1$ and let $(Y_1, \dots, Y_n)$ be independent random variables such that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\big[|Y_i|^p\big] < \infty$. Then there exists $\mathfrak{C}_p$ such that

$$\mathbb{E}\Big[\Big|\sum_{i=1}^n Y_i\Big|^p\Big] \leq \mathfrak{C}_p\Big\{\sum_{i=1}^n \mathbb{E}\big[|Y_i|^p\big] + \Big(\sum_{i=1}^n \mathbb{E}\big[|Y_i|^2\big]\Big)^{p/2}\Big\}. \tag{4.35}$$

To prove Lemma 5 we first use Cauchy Schwarz inequality to get

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}-\gamma_{jk}\big|^{p}\big] \leq \sqrt{\mathbb{E}\Big[\Big|\tfrac{(1-p(\Delta_T))\lfloor T\Delta_T^{-1}\rfloor}{N_T}\Big|^{2p}\Big]}\sqrt{\mathbb{E}\big[\big|\widetilde{\gamma}_{jk}-\gamma_{jk}\big|^{2p}\big]} \tag{4.36}$$

where

$$\widetilde{\gamma}_{jk}-\gamma_{jk} = \frac{1}{(1-p(\Delta_T))\lfloor T\Delta_T^{-1}\rfloor}\Big(\sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} g_{jk}\big(\mathbf{D}^{\Delta}X_i\big)\mathbb{1}_{\{\mathbf{D}^{\Delta}X_i\neq 0\}}-\gamma_{jk}\Big). \tag{4.37}$$

According to Proposition 1 the $\big(\mathbf{D}^{\Delta_T}X_i\big)$ have distribution

$$f_{\mathbf{D}^{\Delta_T(x)}X_1} = p(\Delta_T)\delta_0(x)+\big(1-p(\Delta_T)\big)\mathbf{P}_{\Delta_T}[f](x), \qquad x\in\mathcal{D}$$

where δ_0 is the Dirac delta function and $p(\Delta_T)=\mathbb{P}(R_{\Delta_T}=0)$. We derive

$$\mathbb{E}\big[\,\widetilde{gamma}_{jk}\big] = \int g_{jk}(z)\mathbb{1}_{\{z\neq 0\}}\frac{f_{\mathbf{D}^{\Delta_T}X_1}(z)}{1-p(\Delta_T)}dz = \int g_{jk}(z)\mathbf{P}_{\Delta_T}[f](z)dz = \gamma_{jk}.$$

Then $\widetilde{\gamma}_{jk}-\gamma_{jk}$ is a sum of centered and identically distributed random variables, define

$$Z_i = \frac{1}{1-p(\Delta_T)}g_{jk}\big(\mathbf{D}^{\Delta_T}X_i\big)\mathbb{1}_{\{\mathbf{D}^{\Delta_T}X_i\neq 0\}}.$$

Since X is a renewal reward process, nonzero and nonconsecutive Z_i are independent, then if we separate the sum in two sums of nonzero and nonconsecutive indices we can apply Rosenthal's inequality for independent variables to each sum, it will not affect the rates but the constant will be modified. For $r\geq 1$ we have by the convex inequality

$$\begin{aligned}\mathbb{E}\big[\big|Z_i-\mathbb{E}[Z_i]\big|^{r}\big] &\leq 2^r\mathbb{E}\big[\big|Z_i\big|^{r}\big]\\ &\leq \frac{2^r2^{jr/2}}{\big(1-p(\Delta_T)\big)^r}\int |g(2^jy-k)|^r\mathbb{1}_{\{y\neq 0\}}f_{\mathbf{D}^{\Delta_T}}(y)dy\\ &= \frac{2^r2^{j(r/2-1)}}{\big(1-p(\Delta_T)\big)^{r-1}}\int |g(z)|^r\mathbf{P}_{\Delta_T}[f]\Big(\frac{z+k}{2^j}\Big)dz,\end{aligned}$$

where we made the substitution $z=2^jy-k$. Lemma 2 and Sobolev embeddings (see [19, 30, 49])

$$\mathcal{B}^s_{\pi\infty}\hookrightarrow\mathcal{B}^{s'}_{r\infty} \quad \text{and} \quad \mathcal{B}^{s'}_{\pi\infty}\hookrightarrow\mathcal{B}^{s}_{\infty\infty}, \tag{4.38}$$

where $r>\pi$, $s\pi>1$ and $s'=s-1/\pi+1/r$, give $\big\|\mathbf{P}_{\Delta_T}[f]\big\|_\infty\leq\mathfrak{M}$. It follows that

$$\mathbb{E}\big[\big|Z_i-\mathbb{E}[Z_i]\big|^{r}\big] \leq 2^r2^{j(r/2-1)}\|g\|^r_{L_r(\mathbb{R})}\mathfrak{M}/(1-p(\Delta_T))^{r-1}$$

and

$$\mathbb{E}\big[\big|Z_i - \mathbb{E}[Z_i]\big|^2\big] \leq \mathfrak{M}/(1-p(\Delta_T))$$

since $\|g\|_{L_2(\mathbb{R})} = 1$. Rosenthal's inequality (4.35) gives for $r \geq 1$

$$\mathbb{E}\big[\big|\widetilde{\gamma}_{jk} - \gamma_{jk}\big|^r\big] \leq \mathfrak{C}_r\Big\{2^r\Big(\frac{2^j}{A_T}\Big)^{\frac{r}{2}-1}\|g\|^r_{L_r(\mathbb{R})}\mathfrak{M} + \mathfrak{M}^{r/2}\Big\}A_T^{-\frac{r}{2}},$$

where $A_T = \lfloor T\Delta_T^{-1}\rfloor(1-p(\Delta_T))$. To conclude it follows from (4.31) that

$$\frac{T}{2\mu} \leq A_T \leq \frac{T}{\mu}$$

and using $2^j \leq T$ we derive

$$\mathbb{E}\big[\big|\widetilde{\gamma}_{jk} - \gamma_{jk}\big|^r\big] \leq \mathfrak{C}T^{-r/2}, \tag{4.39}$$

where $\mathfrak{C}$ depends on r, $\|g\|_{L_r(\mathbb{R})}$, $\mathfrak{M}$ and μ. Then (4.32) of Lemma 4, (4.39) and (4.36) completes the proof. □

Lemma 6. *Choose j and c such that*

$$2^jT^{-1}\log(T^{1/2}) \leq 1 \text{ and } c^2 \geq 32\mu\Big(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\Big).$$

For all $r \geq 1$, let $\kappa_r = cr$. We have

$$\mathbb{P}\Big(\big|\widehat{\gamma}_{jk} - \gamma_{jk}\big| \geq \frac{\kappa_r}{2}T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \leq T^{-r/2},$$

where $\widehat{\gamma}_{jk}$ is defined in (4.13) and γ_{jk} in (4.34).

Proof of Lemma 6. The proof is obtained with Bernstein's inequality. Consider $Y_1, \dots, Y_n$ independent random variables such that $|Y_i| \leq \mathfrak{A}$, $\mathbb{E}[Y_i] = 0$ and $b_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2]$. Then for any $\lambda > 0$,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n Y_i\Big| > \lambda\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2(b_n^2 + \frac{\lambda\mathfrak{A}}{3})}\Big). \tag{4.40}$$

To prove Lemma 6 we first have the following decomposition using (4.37)

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\kappa_r}{2}T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \\
&\quad = \mathbb{P}\Big(|\widetilde{\gamma}_{jk} - \gamma_{jk}| \geq \tfrac{N_T}{(1-p(\Delta_T)\lfloor T\Delta_T^{-1}\rfloor}\frac{\kappa_r}{2}T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \\
&\quad \leq \mathbb{P}\Big(|\widetilde{\gamma}_{jk} - \gamma_{jk}| \geq \mathfrak{c}\frac{\kappa_r}{2}T^{-1/2}\sqrt{\log(T^{1/2})}\Big) + \mathbb{P}\big(\tfrac{N_T}{(1-p(\Delta_T)\lfloor T\Delta_T^{-1}\rfloor} > \mathfrak{c}\big)
\end{aligned} \tag{4.41}$$

for some positive $\mathfrak{c}$. To bound the first term, we keep the notation Z_i introduced in the proof of Lemma 5, $\widetilde{\gamma}_{jk} - \gamma_{jk}$ is a sum of centered and identically distributed random variables bounded by $2^{j/2}\|g\|_\infty/(1-p(\Delta_T))$ which verify

$$\mathbb{E}\big[\big|Z_i - \mathbb{E}[Z_i]\big|^2\big] \leq \mathfrak{M}/(1-p(\Delta_T)).$$

After separating the sum to get two sums of nonzero and nonconsecutive indices we apply Bernstein's inequality (4.40) for independent variables to each sum, which modify the constants. It follows that

$$\begin{aligned}\mathbb{P}\Big(|\widetilde{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\mathfrak{c}\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big)\\ \leq & 2\exp\left(-\frac{\mathfrak{c}^2\kappa_r^2 T^{-1}\log(T^{1/2})\lfloor T\Delta_T^{-1}\rfloor\big(1-p(\Delta_T)\big)}{16\Big(\mathfrak{M} + \frac{\mathfrak{c}\kappa_r T^{-1/2}\sqrt{\log(T^{1/2})}2^{j/2}\|g\|_\infty}{6}\Big)}\right).\end{aligned}$$

Using that $2^j T^{-1}\log(T^{1/2}) \leq 1$ and (4.31) which gives

$$T^{-1}\lfloor T\Delta_T^{-1}\rfloor(1-p(\Delta_T)) \geq \frac{1}{2\mu},$$

we have

$$\begin{aligned}\mathbb{P}\Big(|\widetilde{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\mathfrak{c}\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big)\\ \leq & 2\exp\left(-\frac{c^2 r}{32\mu\big(\mathfrak{M} + \frac{\mathfrak{c}\kappa_r\|g\|_\infty}{6}\big)} r\log(T^{1/2})\right) \leq T^{-r/2},\end{aligned}$$

since $c^2 \geq 32\mu\big(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\big)$, and $\mathfrak{c}$ can be chosen greater to 1. Then (4.33) of Lemma 4, (4.39) and (4.41) completes the proof. □

Proof of of part 1) of Theorem 1. It is a consequence of Lemma 2, 5, 6 and of the general theory of wavelet threshold estimators of Kerkyacharian and Picard [63]. It suffices to have conditions (5.1) and (5.2) of Theorem 5.1 of [63], which are satisfied –Lemma 5 and 6– with $c(T) = T^{-1/2}$ and $\Lambda_n = c(T)^{-1}$ (with the notation of [63]). We can now apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [63] to obtain the result. □

Completion of the proof of Theorem 1

To prove part 2) of Theorem 1 we decompose the L_p loss as follows

$$\begin{aligned}&\big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\\ &\leq \big(\mathbb{E}\big[\big\|\widehat{f}_{T,\Delta_T} - \mathbf{P}_{\Delta_T}[f]\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p} + \big\|\mathbf{P}_{\Delta_T}[f] - f\big\|_{L_p(\mathcal{D})}.\end{aligned}$$

An upper bound for the first term is given by part 1) of Theorem 1

$$\big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T} - \mathbf{P}_{\Delta_T}[f]\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \leq \mathfrak{C}T^{-\alpha(s,p,\pi)}, \tag{4.42}$$

where $\mathfrak{C}$ continuously depends on μ, and on $s, \pi, p, \mathfrak{M}, \phi, \psi$ and μ. Since

$$\mathbf{P}_{\Delta_T}[f] - f = -(1 - p_1(\Delta_T))f + \sum_{m=2}^{\infty} p_m(\Delta_T)\mathbf{P}_{\Delta_T}[f]^{\star m}$$

Lemma 1, Young's inequality, which gives $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathcal{D})} \leq \|f\|_{L_p(\mathcal{D})}$ and Sobolev embeddings (4.38), which give $\|f\|_{L_p(\mathcal{D})} \leq \mathfrak{M}$ and leads to the bound

$$\big\|\mathbf{P}_{\Delta_T}[f] - f\big\|_{L_p(\mathcal{D})} \leq 2\tau(0)\Delta_T + 2\mathfrak{M}\sum_{m=2}^{\infty} \frac{\big(2\tau(0)\Delta_T\big)^{m-1}}{m!} \leq \mathfrak{C}\Delta_T, \tag{4.43}$$

where $\mathfrak{C}$ continuously depends on $\tau(0)$ and $\mathfrak{M}$. We finish the proof noticing that (4.42) is predominant compared to (4.43) since $\alpha(s, p, \pi) \leq 1/2$ and $T\Delta_T^2 = O(1)$. Finally we take the supremum in μ and $\tau(0)$ over any compact of $(0, \infty)$ to render the constant independent of the unknown interarrival law τ. The proof is now complete.

4.6.3 Proof of Proposition 2

First we prove part 1) of Proposition 2. The set $\mathcal{H}(s, \pi, \mathfrak{O}, \mathfrak{N})$ is a subset of

$$\big(\mathcal{B}_{\pi\infty}^s(\mathcal{D}), \|.\|_{\mathcal{B}_{\pi\infty}^s(\mathcal{D})}\big)$$

which is a Banach space. We show that $\mathcal{H}(s, \pi, \mathfrak{O}, \mathfrak{N})$ is complete since it is a closed subset of a Banach space. For that we establish the following assertions; for all sequence $h_n \in \mathcal{H}(s, \pi, \mathfrak{O}, \mathfrak{N})$ such that there exists h with

$$\|h_n - h\|_{\mathcal{B}_{\pi\infty}^s(\mathcal{D})} \to 0, \quad \text{as } n \to \infty,$$

we have $h \in \mathcal{H}(s, \pi, \mathfrak{O}, \mathfrak{N})$ *i.e.* $\|h\|_{\mathcal{B}_{\pi\infty}^s(\mathcal{D})} \leq \mathfrak{N}$ and $\|h\|_{L_1(\mathcal{D})} \leq \mathfrak{O}$. The first inequality is immediate. The second one is a consequence of the compactness of $\mathcal{D}$. Indeed

$$\|h_n - h\|_{\mathcal{B}_{\pi\infty}^s(\mathcal{D})} \to 0, \quad \text{as } n \to \infty,$$

we have by definition of the Besov norm (4.8) that

$$\|h_n - h\|_{L_\pi(\mathcal{D})} \to 0, \quad \text{as } n \to \infty.$$

Since $\mathcal{D}$ is compact and $\pi \geq 1$ we derive from Hölder's inequality that

$$\|h_n - h\|_{L_1(\mathcal{D})} \to 0, \quad \text{as } n \to \infty$$

and then $\|h\|_{L_1(\mathcal{D})} \leq \mathfrak{O}$ follows. The proof of part 1) of Proposition 2 is now complete.

To prove part 2) of Proposition 2 we show that $\mathbf{H}_{\Delta,f}$ sends elements of $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ into $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ and that it is a contraction. We start with the first assertion, the triangle inequality gives for $h \in \mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$

$$\big\|\mathbf{H}_{\Delta,f}[h]\big\|_{L_1(\mathcal{D})} \leq \big\|\mathbf{P}_\Delta[f]\big\|_{L_1(\mathcal{D})} + \big(1-p_1(\Delta)\big)\|h\|_1 + \sum_{m=2}^{\infty} p_m(\Delta)\|h^{\star m}\|_{L_1(\mathcal{D})},$$

where $\big\|\mathbf{P}_\Delta[f]\big\|_{L_1(\mathcal{D})} \leq \big\|\mathbf{P}_\Delta[f]\big\|_{L_1(\mathbb{R})} = 1$. Immediate induction on Young's inequality leads to

$$\|h^{\star m}\|_{L_1(\mathcal{D})} \leq \|h\|_{L_1(\mathcal{D})}^m \leq \mathfrak{O}^m$$

since $h \in \mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ and with Lemma 1 we get

$$\big\|\mathbf{H}_{\Delta,f}[h]\big\|_1 \leq 1 + 2\mathfrak{O}\tau(0)\Delta + \frac{1}{\tau(0)\Delta}(e^{2\mathfrak{O}\tau(0)\Delta} - 1 - 2\mathfrak{O}\tau(0)\Delta) \leq \mathfrak{O}$$

for Δ small enough since $\mathfrak{O} > 1$. Similar computations and (4.25) give

$$\begin{aligned}\big\|\mathbf{H}_{\Delta,f}[h]\big\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq& \big\|\mathbf{P}_\Delta[f]\big\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} + \big(1-p_1(\Delta)\big)\|h\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \\ &+ \sum_{m=2}^{\infty} p_m(\Delta)\|h^{\star m}\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \\ \leq& \mathfrak{M} + 2\tau(0)\Delta\mathfrak{N} + \frac{\mathfrak{N}}{\tau(0)\Delta\mathfrak{O}}(e^{2\mathfrak{O}\tau(0)\Delta} - 1 - 2\mathfrak{O}\tau(0)\Delta) \leq \mathfrak{N},\end{aligned}$$

for Δ small enough since $\mathfrak{M} < \mathfrak{N}$. Then if h is in $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$, $\mathbf{H}_{\Delta,f}$ belongs to $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$.

For the contraction property, we have for all $h_1, h_2 \in \mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$

$$\begin{aligned}\mathbf{H}_{\Delta,f}[h_1] - \mathbf{H}_{\Delta,f}[h_2] =& (1-p_1(\Delta))(h_1-h_2) \\ &- (h_1-h_2) \star \sum_{m=2}^{\infty} p_m(\Delta) \sum_{q=0}^{m-1} h_1^{\star q} \star h_2^{\star m-1-q}\end{aligned} \tag{4.44}$$

Lemma 1 gives

$$0 \leq 1 - p_1(\Delta) \leq 2\tau(0)\Delta, \tag{4.45}$$

and with Young's inequality and since h_1 and h_2 belong to $\mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$ we get

$$\Big\|\sum_{m=2}^{\infty} p_m(\Delta) \sum_{q=0}^{m-1} h_1^{\star q} \star h_2^{\star m-1-q}\Big\|_{L_1(\mathcal{D})} \leq 2\mathfrak{O}\big(e^{2\tau(0)\mathfrak{O}\Delta} - 1\big) \tag{4.46}$$

for Δ small enough. Finally injecting (4.45) and (4.46) into (4.44) leads to the contraction property for all $h_1, h_2 \in \mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$

$$\big\|\mathbf{H}_{\Delta,f}[h_1] - \mathbf{H}_{\Delta,f}[h_2]\big\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \big(2\tau(0)\Delta + 2\mathfrak{O}\big(e^{2\tau(0)\mathfrak{O}\Delta} - 1\big)\big)\|h_1 - h_2\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})},$$

which concludes the proof.

4.6.4 Proof of Theorem 2

Proof of part 1) of Theorem 2

As for the proof of part 1) of Theorem 1 we apply the general results of Kerkyacharian and Picard [63] and first establish some technical lemmas.

Lemma 7. *Let $2^j \leq T$, then for $p \geq 1$ we have for all $m \geq 1$*

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p\big] \leq \mathfrak{C}_{p,m,\|g\|_{L_p(\mathbb{R})},\mathfrak{M},\mu,\tau} T^{-p/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (4.22) *and*

$$\gamma_{jk}^{(m)} = \int g_{jk}(y)\mathbf{P}_\Delta[f]^{\star m}(y)dy. \tag{4.47}$$

Proof of Lemma 7. For $m \geq 1$, $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is the sum of $\lfloor N_T/m \rfloor$ identically distributed random variables, where N_T is random. First we replace N_T by its deterministic asymptotic limit using the following decomposition

$$\begin{aligned}\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p\big] =& \mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p \mathbb{1}_{\big\{\big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\big| \geq \lambda\big\}}\big] \\ &+ \mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p \mathbb{1}_{\big\{\big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\big| < \lambda\big\}}\big].\end{aligned}$$

Take $\lambda = 1/4\mu$ and denote $n_m = \lfloor T/m\mu \rfloor$ and $n_m' = \lfloor T/(4m\mu) \rfloor$, we have with that (4.31) that

$$\begin{aligned}\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p\big] \leq& 2^{jp/2}\|g\|_\infty \mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| \geq \frac{\Delta_T}{4\mu}\Big) \\ &+ \mathbb{E}\Big[\Big|\frac{1}{n_m'}\sum_{i=1}^{n_m}\big(\mathbf{D}_m^\Delta X_{S_i} - \mathbb{E}\big[\mathbf{D}_m^\Delta X_{S_i}\big]\big)\Big|^p\Big].\end{aligned}$$

For the first term of the right hand part, $T\Delta_T = O(T^\delta)$, Lemma 3 and $2^j \leq T$ leads to

$$\mathbb{P}\Big(\big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\big| \geq \frac{\Delta_T}{4\mu}\Big) \leq \mathfrak{C}T^{p/2}\exp\Big(-\mathfrak{C}T^\delta\Big) \leq \mathfrak{C}\exp(-T^{\delta_p'}), \tag{4.48}$$

for some $\delta_p' < \delta$ and where $\mathfrak{C}$ depends on $p, \|g\|_\infty, \mathfrak{A}, \mathfrak{a}, \mu$. For the second term we apply Rosenthal's inequality (4.35). Since X is a renewal process the variables $\big(\mathbf{D}_m^\Delta X_{S_{2i}}\big)_i$ are independent but dependent of the variables $\big(\mathbf{D}_m^\Delta X_{S_{2i+1}}\big)_i$ which are independent. It ensures that the variables $\big(\mathbf{D}_m^\Delta X_{S_i}\big)$ are distributed according to $\mathbf{P}_{\Delta_T}[f]^{\star m}$. Moreover if we separate the sum $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ between odd and even indices we can apply Rosenthal's inequality for independent variables to each sum. For $p \geq 1$ we have by convex inequality

$$\begin{aligned}\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta} X_{S_i}) - \gamma_{jk}\big|^p\big] &\leq 2^p \mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta} X_{S_i})\big|^p\big] \\ &\leq 2^p 2^{jp/2} \int |g(2^j y - k)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}(y) dy \\ &\leq 2^p 2^{j(p/2-1)} \int |g(z)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}\Big(\frac{z+k}{2^j}\Big) dz,\end{aligned}$$

where we made the substitution $z = 2^j y - k$. Lemma 2 and Sobolev embeddings (4.38) give $\big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_\infty \leq \mathfrak{M}$. It follows that

$$\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta} X_{S_i}) - \gamma_{jk}\big|^p\big] \leq 2^p 2^{j(p/2-1)} \|g\|_{L_p(\mathbb{R})}^p \mathfrak{M}$$

and

$$\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta} X_{S_i}) - \gamma_{jk}\big|^2\big] \leq \mathfrak{M}$$

since $\|g\|_{L_2(\mathbb{R})} = 1$. We derive for $p \geq 1$

$$\begin{aligned}&\mathbb{E}\Big[\Big|\frac{1}{n_m'} \sum_{i=1}^{n_m} \big(\mathbf{D}_m^{\Delta} X_{S_i} - \mathbb{E}\big[\mathbf{D}_m^{\Delta} X_{S_i}\big]\big)\Big|^p\Big] \\ &\qquad\leq \mathfrak{C}_p \Big\{2^p \Big(\frac{2^j}{n_m}\Big)^{\frac{p}{2}-1} \|g\|_{L_p(\mathbb{R})}^p \mathfrak{M} + \mathfrak{M}^{p/2}\Big\} {n_m'}^{-p/2} \\ &\qquad\leq \mathfrak{C}_{p,m,\|g\|_{L_p(\mathbb{R})},\mathfrak{M},\mu} T^{-p/2}. \end{aligned} \tag{4.49}$$

It follows from (4.48) and (4.49) that

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p\big] \leq \mathfrak{C} \exp(-T^{\delta_p'}) + \mathfrak{C} T^{-\frac{p}{2}} \leq \mathfrak{C} T^{-\frac{p}{2}},$$

since the first term is negligible as $\delta' > 0$, where $\mathfrak{C}$ depends on $p, m, \|g\|_{L_p(\mathbb{R})}, \|g\|_\infty, \mathfrak{A}, \mathfrak{a}, \mathfrak{M}, \mu$. It concludes the proof. □

Lemma 8. *Choose j and c such that*

$$2^j T^{-1} \log(T^{1/2}) \leq 1 \text{ and } c^2 \geq 256 m\mu\Big(\mathfrak{M} + \frac{c\|g\|_\infty}{24}\Big).$$

For all $r \geq 1$ let $\kappa_r = cr$. We have for all $m \geq 1$

$$\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} T^{-1/2} \sqrt{\log(T^{1/2})}\Big) \leq T^{-r/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (4.22) and $\gamma_{jk}^{(m)}$ in (4.47).

Proof of Lemma 8. As for the proof of Lemma 7 we decompose as follow for $m \geq 1$

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \\
&\quad \leq \mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| \geq \frac{\Delta_T}{4\mu}\Big) \\
&\quad + \mathbb{P}\Big(\Big|\frac{1}{n'_m}\sum_{i=1}^{n_m}\big(\mathbf{D}_m^{\Delta} X_{S_i} - \mathbb{E}\big[\mathbf{D}_m^{\Delta} X_{S_i}\big]\big)\Big| \geq \frac{\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big),
\end{aligned}$$

where $n_m = \lfloor T/m\mu \rfloor$ and $n'_m = \lfloor T/(4m\mu) \rfloor$. From $T\Delta_T^2 = O(T^\delta)$ and Lemma 3 we derive

$$\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| \geq \frac{\Delta_T}{4\mu}\Big) \leq \exp\Big(-\mathfrak{C}T^\delta\Big), \tag{4.50}$$

where $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \mu$. For the second term we apply Bernstein's inequality (4.40) and as in the proof of Lemma 7 we separate the sum between odd and even indices to work with independent variables. We get

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\frac{1}{n'_m}\sum_{i=1}^{n_m}\big(\mathbf{D}_m^{\Delta} X_{S_i} - \mathbb{E}\big[\mathbf{D}_m^{\Delta} X_{S_i}\big]\big)\Big| \geq \frac{\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \\
&\quad \leq 2\exp\left(-\frac{\kappa_r^2 {n'_m}^2 T^{-1}\log(T^{1/2})}{16\Big(n_m\mathfrak{M} + \frac{\kappa_r n'_m T^{-1/2}\sqrt{\log(T^{1/2})}2^{j/2}\|g\|_\infty}{6}\Big)}\right) \\
&\quad \leq 2\exp\left(-\frac{c^2 r}{128 m\mu\Big(\mathfrak{M} + \frac{\kappa_r T^{-1/2}\sqrt{\log(T^{1/2})}2^{j/2}\|g\|_\infty}{24}\Big)} r\log(T^{1/2})\right).
\end{aligned}$$

With $2^j T^{-1}\log(T^{1/2}) \leq 1$ and $c^2 \geq 256 m\mu\big(\mathfrak{M} + \frac{c\|g\|_\infty}{24}\big)$ we have for $r \geq 1$

$$\mathbb{P}\Big(\Big|\frac{1}{n'_m}\sum_{i=1}^{n_m}\big(\mathbf{D}_m^{\Delta} X_{S_i} - \mathbb{E}\big[\mathbf{D}_m^{\Delta} X_{S_i}\big]\big)\Big| \geq \frac{\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \leq \mathfrak{C}T^{-r/2}, \tag{4.51}$$

where $\mathfrak{C}$ depends on $c, m, \|g\|_\infty, \mathfrak{M}, \mu$. It follows from (4.50) and (4.51) that

$$\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} T^{-1/2}\sqrt{\log(T^{1/2})}\Big) \leq \exp\Big(-\mathfrak{C}T^\delta\Big) + \mathfrak{C}T^{-r/2} \leq \mathfrak{C}T^{-r/2}$$

since the first term is negligible since $\delta > 0$ and where $\mathfrak{C}$ depends on $c, m, \|g\|_\infty, \mathfrak{M}, \mu$. It concludes the proof. □

Completion of the proof of part 1) of Theorem 2. It is a consequence of Lemma 2, 7, 8 and of the general theory of wavelet threshold estimators of Kerkyacharian and Picard [63]. It suffices to have conditions (5.1) and (5.2) of Theorem 5.1 of [63], which are satisfied –Lemma 7 and 8– with $c(T) = T^{-1/2}$ and $\Lambda_n = c(T)^{-1}$ (with the notation of [63]). We can now apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [63] to obtain the result. □

Completion of the proof of Theorem 2

To prove Theorem 2 we define for K in $\mathbb{N}$ and x in $\mathcal{D}$ the quantity

$$\widetilde{f}^K_{T,\Delta}(x) = \sum_{m=1}^{K+1} l_m(\Delta,\vartheta)\widehat{P_{\Delta,m}}(x).$$

It is the estimator of f one would compute if τ were known. We decompose the L_p error as follows

$$\begin{aligned}\left(\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \leq &\left(\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T} - \widetilde{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \\ &+ \left(\mathbb{E}\left[\|\widetilde{f}^K_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p},\end{aligned}$$

and control each term separately.

First we look at the second term

$$\begin{aligned}\left(\mathbb{E}\left[\|\widetilde{f}^K_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \leq &\left(\mathbb{E}\left[\|\widetilde{f}^K_{T,\Delta_T} - \mathbf{L}_{\Delta_T,K}\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \\ &+ \left\|\mathbf{L}_{\Delta_T,K} - \mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right]\right\|_{L_p(\mathcal{D})} \\ &+ \left\|\mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - f\right\|_{L_p(\mathcal{D})}.\end{aligned} \tag{4.52}$$

An upper bound for the first term is given by part 1) of Theorem 2, given the definition (4.21) of $\mathbf{L}_{\Delta_T,K}$ and the triangle inequality we derive

$$\left(\mathbb{E}\left[\|\widetilde{f}^K_{T,\Delta_T} - \mathbf{L}_{\Delta_T,K}\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \leq \mathfrak{C}T^{-\alpha(s,p,\pi)}, \tag{4.53}$$

where $\mathfrak{C}$ depends on $\vartheta, s, \pi, p, \mathfrak{M}, \phi, \psi$, and K. By (4.20), we have

$$\left\|\mathbf{L}_{\Delta_T,K} - \mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right]\right\|_{L_p(\mathcal{D})} \leq \mathfrak{C}\Delta_T^{K+1}, \tag{4.54}$$

where $\mathfrak{C}$ depends on ϑ, $\mathfrak{O}$ and $\mathfrak{M}$. For the last term we use the fixed point theorem's approximation, first we have to relate the L_p norm with the Sobolev one. The triangle inequality ensures that if f is in $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ then $\mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - f$ is in $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$. It follows using Sobolev embeddings (4.38) that

$$\left\|\mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - f\right\|_{L_p(\mathcal{D})} \leq \left\|\mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - f\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}.$$

We now use the approximation given by the Banach fixed point theorem

$$\left\|\mathbf{H}^{\circ K}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - f\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \mathfrak{K}(\Delta_T)^K\left\|\mathbf{H}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - \mathbf{P}_{\Delta_T}[f]\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}.$$

After replacing $\mathbf{H}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right]$ by its expression and using the triangle inequality we have

$$\left\|\mathbf{H}_{\Delta,f}\left[\mathbf{P}_{\Delta_T}[f]\right] - \mathbf{P}_{\Delta_T}[f]\right\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} \leq \mathfrak{C}\Delta_T,$$

which leads to

$$\big\|\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_{\Delta_T}[f]\big]-f\big\|_{L_p(\mathcal{D})}\leq\mathfrak{C}\Delta_T\mathfrak{K}(\Delta_T)^K, \tag{4.55}$$

$\mathfrak{C}$ depends on $\vartheta,\mathfrak{M},\mathfrak{O},K$. We conclude by injecting (4.19), (4.53), (4.54) and (4.55) into (4.52) and taking the supremum in ϑ over the compact set Θ.

We now control $\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K-\widetilde{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]$, the triangle inequality leads to

$$\begin{aligned}\big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K-\widetilde{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p}\\ \leq\sum_{m=1}^{K+1}\big(\mathbb{E}\big[\|\big(l_m(\Delta_T,\widehat{\vartheta_T})-l_m(\Delta_T,\vartheta)\big)\widehat{P_{\Delta_T,m}}\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p},\end{aligned}$$

where $\widehat{P_{\Delta_T,m}}$ does not depend on ϑ (see (4.22)). Cauchy-Schwarz inequality leads to

$$\begin{aligned}\mathbb{E}\big[\|\big(l_m(\Delta_T,\widehat{\vartheta_T})-l_m(\Delta_T,\vartheta)\big)\widehat{P_{\Delta_T,m}}\|_{L_p(\mathcal{D})}^p\big]^2\\ \leq\mathbb{E}\Big[\big|l_m(\Delta_T,\widehat{\vartheta_T})-l_m(\Delta_T,\vartheta)\big|^{2p}\Big]\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|_{L_p(\mathcal{D})}^{2p}\Big],\end{aligned}$$

where using part 1) of Theorem 2, the triangle inequality and that $T\geq1$ we have

$$\begin{aligned}\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|_{L_p(\mathcal{D})}^{2p}\Big]&\leq\mathbb{E}\Big[\|\widehat{P_{\Delta_T,m}}-\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathcal{D})}^{2p}\Big]+\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathcal{D})}^{2p}\\ &\leq\mathfrak{C}T^{-2\alpha(s,p,\pi)p}+\mathfrak{M}^{2p}\leq\mathfrak{C}\end{aligned} \tag{4.56}$$

where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi,\vartheta$. We conclude the proof with the following Lemma, proof of which is given in the Appendix.

Lemma 9. *Work under Assumptions 4 and 3. We have for all $r\geq2$*

$$\mathbb{E}\big[|l_m(\Delta_T,\widehat{\vartheta_T})-l_m(\Delta_T,\vartheta)|^r\big]\leq\mathfrak{C}\big(T^{1-r}+T^{-r/2}\big)$$

where $\mathfrak{C}$ depends on $r,\mathfrak{A},\mathfrak{a},\vartheta$.

It follows from (4.56) and Lemma 9 applied with $r=2p$ that

$$\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K-\widetilde{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}\big]^{1/p}\leq\mathfrak{C}\big(T^{1-1/(2p)}+T^{-1/2}\big),$$

where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi,\mathfrak{A},\mathfrak{a},\vartheta$. We deduce for $p\geq1$

$$\begin{aligned}\sup_{\vartheta\in\Theta}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K\big(\widehat{\vartheta}\big)-\widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p}\\ \leq\mathfrak{C}\big(T^{-(1-1/(2p))}+T^{-1/2}\big)\end{aligned}$$

where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi,\mathfrak{A},\mathfrak{a},K$. It is negligible compared to $T^{-\alpha(s,p,\pi)}$ since $\alpha(s,p,\pi)\leq1/2$. The proof of Theorem 2 is now complete.

Appendix

Proof of Proposition 1

Let $x \in \mathbb{R}$, we have by Assumption 1 that $\{R_\Delta \neq 0\} = \{X_\Delta \neq 0\}$, it follows from stationarity that

$$\begin{aligned}\mathbb{P}(\mathbf{D}^\Delta X_{S_1} \leq x) &= \mathbb{P}(X_\Delta \leq x | X_\Delta \neq 0)\\ &= \sum_{m=0}^{\infty} \mathbb{P}(X_\Delta \leq x | R_\Delta = m, R_\Delta \neq 0)\mathbb{P}(R_\Delta = m)\\ &= \sum_{m=1}^{\infty} p_m(\Delta)\mathbb{P}(X_\Delta \leq x | R_\Delta = m)\end{aligned}$$

where $\mathbb{P}(X_\Delta \leq x | R_\Delta = m) = \int_{-\infty}^{x} f^{\star m}(y)dy$ for $m \geq 1$. It follows

$$\mathbb{P}(\mathbf{D}^\Delta X_{S_1} \leq x) = \int_{-\infty}^{x} \mathbf{P}_\Delta[f](y)dy.$$

Proof of Lemma 1

We start with the second assertion. For $m \geq 1$ we have

$$p_m(\Delta) = \frac{\mathbb{P}(R_\Delta = m)}{1 - \mathbb{P}(R_\Delta = 0)}.$$

First we derive the lower bound

$$1 - \mathbb{P}(R_\Delta = 0) = 1 - \mathbb{P}(J_1 \geq \Delta) \geq \frac{1}{\mu}\int_0^\Delta 1 - F(\Delta)dx \geq \frac{\Delta}{2\mu} \tag{4.57}$$

since F is a cumulative distribution function; it is positive, increasing and continuous with $F(0) = 0$. Then there exists Δ_1 such that for all $\Delta \leq \Delta_1$ we have $F(\Delta) \leq \frac{1}{2}$. Second we have for all $m \geq 1$

$$\mathbb{P}(R_\Delta = m) \leq \mathbb{P}(J_1 + \ldots + J_m \leq \Delta) = \int_0^\Delta \tau_0 \star \tau^{\star m-1}(x)dx,$$

where for all $x \in [0, \Delta]$

$$\begin{aligned}\tau_0 \star \tau^{\star m-1}(x) = x^{m-1}\int_0^1 \tau_0(xt_1)\int_0^{1-t_1} \tau(xt_2)\ldots\\ \int_0^{1-t_1-\ldots-t_{m-2}} \tau(xt_{m-1})\tau(x(1-t_1-\ldots-t_{m-2}-t_{m-1}))dt_1\ldots dt_{m-1}.\end{aligned}$$

We derive

$$\begin{aligned}
&\tau_0 \star \tau^{\star m-1}(x) \\
&\leq x^{m-1} \sup_{t\in[0,x]} \tau_0(t) \big(\sup_{t\in[0,x]} \tau(t)\big)^{m-1} \int_0^1 \int_0^{1-t_1} \dots \int_0^{1-t_1-\dots-t_{m-2}} dt_1 \dots dt_{m-1} \\
&\leq \frac{1}{\mu} \big(\sup_{t\in[0,\Delta]} \tau(t)\big)^{m-1} \frac{x^{m-1}}{(m-1)!},
\end{aligned}$$

since

$$\int_0^1 \int_0^{1-t_1} \dots \int_0^{1-t_1-\dots-t_{m-2}} dt_1 \dots dt_{m-1} = \frac{1}{(m-1)!}.$$

It follows that

$$\mathbb{P}(R_\Delta = m) \leq \frac{1}{\mu} \big(\sup_{t\in[0,\Delta]} \tau(t)\big)^{m-1} \frac{\Delta^m}{m!}. \tag{4.58}$$

Since τ is continuous, there exists Δ_2 such that

$$\sup_{t\in[0,\Delta_2]} \tau(t) \leq 2\tau(0).$$

Taking $\Delta_0 = \Delta_1 \wedge \Delta_2$, (4.57) and (4.58) lead to the second assertion. The first one is straightforward from the previous computations.

Proof of Proposition 3

According to the definition of $\mathbf{L}_{\Delta,K}$ inequality (4.20) is immediate. The dependency in $\tau(0)$ and $\mathfrak{M}$ of the constant is a consequence of Lemma 1, part 2) of Proposition 2 and Lemma 2. A rearrangement of the terms enables to write $\mathbf{L}_{\Delta,K}$ as a sum of increasing powers of $\mathbf{P}_\Delta[f]^{\star m}$. Thus we have to prove that only the $K+1$ first convolution powers of $\mathbf{P}_\Delta[f]$ intervene and that the coefficient $l_m(\Delta)$ in front of $\mathbf{P}_\Delta[f]^{\star m}$ in the rearrangement satisfies

$$\big| l_m(\Delta) \big| \leq \mathfrak{C}_{\tau(0)} \Delta^{m-1}.$$

For that we show that for all $L \geq 1$ the Taylor expansion of order L in Δ of $\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big]$, that we denote $\widetilde{\mathbf{L}}_{\Delta,K,L}$, only depends on $\mathbf{P}_\Delta[f]^{\star m}$, $m = 1, \dots, L+1$ with coefficients such that $\widetilde{l}_{m,K}(\Delta) \leq \mathfrak{C}_{\tau(0)} \Delta^{m-1}$. We prove the result by induction on K. For $K = 1$ we immediately have the result by Lemma 1 since

$$\mathbf{H}_{\Delta,f}\big[\mathbf{P}_\Delta[f]\big] = 2\mathbf{P}_\Delta[f] - \sum_{m=1}^{\infty} p_m(\Delta) \mathbf{P}_\Delta[f]^{\star m},$$

it follows that

$$\widetilde{\mathbf{L}}_{\Delta,L,1} = (2 - p_1(\Delta)\mathbf{P}_\Delta[f] - \sum_{m=2}^{L+1} p_m(\Delta)\mathbf{P}_\Delta[f]^{\star m}$$

with $\widetilde{l}_{1,1}(\Delta) = (2 - p_1(\Delta)) \le 2$ and $\widetilde{l}_{m,1}(\Delta) = p_m(\Delta) \le \mathfrak{C}_{\tau(0)}\Delta^{m-1}$. Then using the definition of $\mathbf{H}_{\Delta,f}$ we have

$$\mathbf{H}_{\Delta,f}^{\circ(K+1)}\big[\mathbf{P}_\Delta[f]\big] = \mathbf{P}_\Delta[f] + \mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big] - \sum_{m=1}^{\infty} p_m(\Delta)\Big(\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big]\Big)^{\star m}.$$

The induction hypothesis and Lemma 1, with part 2) of Proposition 2 which ensures that $\mathbf{H}_{\Delta,f}^{\circ K}\big[\mathbf{P}_\Delta[f]\big] \in \mathcal{H}(s,\pi,\mathfrak{O},\mathfrak{N})$, lead to

$$\begin{aligned}
\widetilde{\mathbf{L}}_{\Delta,L,K+1} &= \mathbf{P}_\Delta[f] + \widetilde{\mathbf{L}}_{\Delta,L,K} - \sum_{m=1}^{L+1} p_m(\Delta)\Big(\widetilde{\mathbf{L}}_{\Delta,L,K}\Big)^{\star m} \\
&= \mathbf{P}_\Delta[f] + \sum_{m=1}^{L+1} \widetilde{l}_{m,L}(\Delta)\mathbf{P}_\Delta[f]^{\star m} - \sum_{m=1}^{L+1} p_m(\Delta)\Big(\sum_{m'=1}^{L+1} \widetilde{l}_{m',L}(\Delta)\mathbf{P}_\Delta[f]^{\star m'}\Big)^{\star m} \\
&= \sum_{m=1}^{L+1} \widetilde{l}_{m,L+1}(\Delta)\mathbf{P}_\Delta[f]^{\star m},
\end{aligned}$$

where $\widetilde{l}_{1,L+1}(\Delta) = 1$ and

$$\widetilde{l}_{m,L+1}(\Delta) = \widetilde{l}_{m,L}(\Delta) - \sum_{k=1}^{m} p_k(\Delta) \sum_{n_1+\ldots+n_k=m} \widetilde{l}_{n_1,L}(\Delta)\ldots\widetilde{l}_{n_k,L}(\Delta)$$

which we bound with Lemma 1 and the induction hypothesis by

$$\begin{aligned}
\big|\widetilde{l}_{m,L+1}(\Delta)\big| &\le \mathfrak{C}\Big(\Delta^{m-1} + \sum_{k=1}^{m} \Delta^{k-1} \sum_{n_1+\ldots+n_k=m} \widetilde{\Delta}^{n_1-1}\ldots\Delta^{n_k-1}\Big) \\
&= \mathfrak{C}\Big(\Delta^{m-1} + m\Delta^{m-1}\Big) \le \mathfrak{C}\Delta^{m-1},
\end{aligned}$$

where $\mathfrak{C}$ is a positive constant depending on $\tau(0)$ and K. We conclude the proof having $L = K$ and $l_m(\Delta) = \widetilde{l}_{m,K}$ for $m = 1,\ldots,K+1$.

Proof of Lemma 9

Preliminary

Lemma 10. *Work under assumptions 3 and 4, for all* $r \ge 2$

$$\mathbb{E}\big[|\widehat{\vartheta_T} - \vartheta|^r\big] \le \mathfrak{C}\big(T^{1-r} + T^{-r/2}\big),$$

where $\mathfrak{C}$ *depends on* $r, \mathfrak{A}, \mathfrak{a}, \vartheta$ *and* $\widehat{\vartheta_T}$ *is defined in Definition 2.*

Démonstration. Let $r > 2$, the proof is a consequence of Proposition 5.5 of Dedecker *et al.* [31] which is a Rosenthal type inequality for dependent data. Define

$$S_T = \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} Y_1$$

where $S_0 = X_0 = 0$ and the $Y_i = \mathbb{1}_{\mathbf{D}^{\Delta}X_i \neq 0} - q(\vartheta)$ are centered identically distributed random variables bounded by 1. To apply Proposition 5.5 of [31] we have to verify that (Y_i) is a sequence of $\theta_{1,\infty}-$dependent random variables. For that Proposition 2.3 of [31] ensures that it is sufficient to have a $\theta-$dependent sequence which is defined as follows with notation of [31] ; Let $\Gamma(u, v, k)$ be the set of (i, j) in $\mathbb{Z}^u \times \mathbb{Z}^v$ such that

$$i_1 < \ldots < i_u \leq i_u + k < j_1 < \ldots < j_v,$$

we have to show that for all $f \in \mathcal{F}_u$ the set of bounded function from $\mathbb{R}^u$ to $\mathbb{R}$ and for all $g \in \mathcal{G}_v$ the set of Lipschitz function from $\mathbb{R}^v$ to $\mathbb{R}$ with Lipschitz coefficient denoted Lip$g$ the sequence $\theta(k)$ defined as

$$\theta(k) = \sup_{u,v} \sup_{(i,j)\in\Gamma(u,v,k)} \sup_{f\in\mathcal{F}_u, g\in\mathcal{G}_v} \frac{\left|Cov\big(f(Y_{i_1}, \ldots, Y_{i_u}), g(Y_{j_1}, \ldots, Y_{j_v})\big)\right|}{v\|f\|_\infty \mathrm{Lip}g}$$

tends to 0. We denote as Y_i and Y_j respectively $(Y_{i_1}, \ldots, Y_{i_u})$ and $(Y_{j_1}, \ldots, Y_{j_v})$, and due to the fact that X is a renewal process Y_i and Y_j are independent if there exists r such that $i_u < r < j_1$ and $Y_r = 1 - p(\Delta)$ *i.e* there is a jump between Y_{i_u} and Y_{j_1}. We denote by A the event "there exists r such that $i_u < r < j_1$ and $Y_r = 1 - p(\Delta)$". It follows that

$$\begin{aligned}\left|Cov\big(f(X_i), g(X_j)\big)\right| &= \left|\mathbb{E}\big[\big(f(X_i) - \mathbb{E}[f(X_i)]\big)\big(g(X_j) - g(0_j)\big)\mathbb{1}_{\{A\}}\big]\right| \\ &\leq 2\|f\|_\infty \mathrm{Lip}g\mathbb{E}\big[\|X_j\|\mathbb{1}_{\{A\}}\big] \\ &\leq 2v\|f\|_\infty \mathrm{Lip}g\mathbb{P}(R_{k\Delta} \neq 0),\end{aligned}$$

since $\|X_j\| \leq v$ as the Y_i are bounded by 1, for every L_p norm $p \geq 0$, and $\mathbb{E}[\mathbb{1}_{\{A\}}]$ is bounded by $\mathbb{P}(R_{k\mathfrak{T}} \neq 0)$. We immediately derive that $\theta(k) \leq 2\mathbb{P}(R_{k\Delta} \neq 0)$ and by Assumption 3 we derive

$$\theta(k) \leq \mathfrak{C}\exp(-\mathfrak{a}(k\Delta)^{\mathfrak{g}'}), \tag{4.59}$$

where $\mathfrak{g} < \mathfrak{g}'$, it tends to 0. We verify the hypothesis of Proposition 5.5 of [31] and get for all $r > 2$

$$\mathbb{E}\big[|S_T|^r\big] \leq \mathfrak{C}\Big(\lfloor T\Delta^{-1}\rfloor \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} i^{r-2}\theta(i) + \big(\lfloor T\Delta^{-1}\rfloor \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} \theta(i)\big)^{r/2}\Big)$$

where $\mathfrak{C}$ depends on r. Since we have the upper bound (4.59), we derive applying (4.30) with $k = 0$ and $k = r - 2$

$$\sum_{i=1}^{\lfloor T\Delta^{-1} \rfloor} \theta(i) \leq \mathfrak{C}\Delta \quad \text{and} \quad \sum_{i=1}^{\lfloor T\Delta^{-1} \rfloor} i^{r-2}\theta(i)\theta(i) \leq \mathfrak{C}\Delta^{r-1}$$

where $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \vartheta$. It follows

$$\frac{1}{\lfloor T\Delta^{-1} \rfloor^r}\mathbb{E}\big[|S_T|^r\big] \leq \mathfrak{C}\big(T^{1-r} + T^{-r/2}\big),$$

where $\mathfrak{C}$ depends on $r, \mathfrak{A}, \mathfrak{a}, \vartheta$. The case $r = 2$ is a consequence of

$$\mathbb{E}\big[|S_T|^2\big] = \frac{1}{\lfloor T\Delta^{-1} \rfloor}\mathbb{V}(Y_1) + \frac{2}{\lfloor T\Delta^{-1} \rfloor^2} \sum_{1 \leq i < j \leq T} Cov(Y_i, Y_j)$$

and the upper bounds $\mathbb{V}(Y_1) \leq \mathfrak{C}\Delta$ where $\mathfrak{C}$ depends on ϑ and

$$|Cov(Y_i, Y_{i+k})| \leq \mathfrak{C}\exp(-\mathfrak{a}k\Delta).$$

We derive

$$\frac{1}{\lfloor T\Delta^{-1} \; rfloor^2}\mathbb{E}\big[|S_T|^2\big] = \mathfrak{C}T^{-1}$$

where $\mathfrak{C}$ depends on $\mathfrak{A}, \mathfrak{a}, \vartheta$. We conclude the proof using Assumption 4, for all $r \geq 2$

$$\begin{aligned}\mathbb{E}\big[|\widehat{\vartheta_T} - \vartheta|^r\big] &= \mathbb{E}\big[|q^{-1}(q(\widehat{\vartheta_T})) - q^{-1}(q(\vartheta))|^r\big] \\ &\leq \|q^{-1}\|_\infty \; \mathbb{E}\big[|q(\widehat{\vartheta_T}) - q(\vartheta)|^r\big] \leq \mathfrak{C}\big(T^{1-r} + T^{-r/2}\big),\end{aligned}$$

where $\mathfrak{C}$ depends on $r, \mathfrak{A}, \mathfrak{a}, \vartheta$. □

Completion of the proof of Lemma 9

The remaining of the proof is now based on the fact that under Assumption 4 the functions $\vartheta \to p_m(., \vartheta)$ are Lipschitz continuous. We show that their derivative with respect to ϑ is bounded, we have for $m \geq 1$ that

$$\begin{aligned}\partial_\vartheta[p_m(\Delta, \vartheta)] =& \frac{1}{\int_0^\Delta \tau_2(z, \vartheta)dz}\Big(\int_0^\Delta \partial_\vartheta[\tau_2(., \vartheta) \star \tau_1^{\star m-1}(., \vartheta)](z)dz \\ &- \int_0^\Delta \partial_\vartheta[\tau_2(., \vartheta) \star \tau_1^{\star m}(., \vartheta)](z)dz\Big) \\ &- \frac{\int_0^\Delta \partial_\vartheta[\tau_2(z, \vartheta)]dz}{\Big(\int_0^\Delta \tau_2(z, \vartheta)dz\Big)^2}\Big(\int_0^\Delta \tau_2(x, \vartheta) \star \tau_1^{\star m-1}(., \vartheta)(z)dz \\ &- \int_0^\Delta \tau_2(x, \vartheta) \star \tau_1^{\star m}(., \vartheta)(z)dz\Big)\end{aligned} \tag{4.60}$$

where $\tau_2(.,\vartheta)/\mu$ is the density of J_1. Immediate induction gives for $m \geq 1$

$$\begin{aligned}\partial_\vartheta[\tau_2(.,\vartheta) \star \tau_1^{\star m}(.,\vartheta)](z) =& \partial_\vartheta[\tau_2(.,\vartheta)] \star \tau_1^{\star m}(.,\vartheta)(z) \\ &+ m\tau_2(.,\vartheta) \star \partial_\vartheta[\tau_1(.,\vartheta)] \star \tau_1^{\star m-1}(.,\vartheta)(z)\end{aligned} \tag{4.61}$$

and

$$\int_0^\Delta g^{\star m}(x)dx \leq \mathfrak{C}\Delta^{m-1} \tag{4.62}$$

for some constant $\mathfrak{C}$ and any bounded function g supported by $(0,\infty)$. Moreover we have

$$\partial_\vartheta[\tau_2(z,\vartheta)] = -\int_0^z \partial_\vartheta\tau_1(x,\vartheta)dx,$$

and it follows from Assumption 4 that for Δ small enough we have $\forall z \leq \Delta$

$$0 < \frac{\tau_1(0,\vartheta)}{2} \leq \tau_1(z,\vartheta) \leq 2\tau_1(0,\vartheta), \tag{4.63}$$

and that its derivative is bounded over $[0,\Delta]$. Finally we bound (4.60), using (4.61) (4.62) and (4.63), we get

$$\big|\partial_\vartheta[p_m(\Delta,\vartheta)]\big| \leq \mathfrak{C}\Delta^{m-1},$$

where $\mathfrak{C}$ continuously depends on ϑ. Then taking the supremum in ϑ over the compact set Θ we derive

$$\big|\partial_\vartheta[p_m(\Delta,\vartheta)]\big| \leq \mathfrak{C}\Delta^{m-1},$$

where $\mathfrak{C}$ is a positive constant independent of ϑ. It follows that for $m \geq 1$, the functions $\vartheta \to p_m(.,\vartheta)$ are Lipschitz continuous and with Lemma 10 we derive

$$\begin{aligned}\mathbb{E}\big[\big|p_m(\Delta,\widehat{\vartheta_T}) - p_m(\Delta,\vartheta)\big|^r\big] &\leq \mathfrak{C}\Delta^{m-1}\mathbb{E}\big[\big|\widehat{\vartheta_T} - \vartheta\big|^r\big] \\ &\leq \mathfrak{C}\big(T^{1-r} + T^{-r/2}\big),\end{aligned}$$

where $\mathfrak{C}$ is a positive constant depending on $r, \mathfrak{A}, \mathfrak{a}, \vartheta$. We conclude the proof using that $l_m(\Delta,\vartheta) = l(p_1(\Delta,\vartheta),\ldots,p_m(\Delta,\vartheta))$ where l is Lipschitz in every argument and the argument are bounded by 1.

Abstract

We consider the macroscopic regime where a compound Poisson process is observed over $[0, T]$ at a sampling rate $\Delta = \Delta_T \to \infty$ as $T \to \infty$. General limit theorems suggest that the law of the increments converge in law to a parametric stable law, depending on a small number of parameters. Thus if the law of the process depends on more parameters, it might be impossible to identify them all in the macroscopic regime. In this last Chapter we try to quantify the loss of information in macroscopic regimes.

Keywords : Discretely observed random process, Renewal reward process, Pure jump Lévy process, Information loss.

Note

Chapter 5 is a work in progress.

Chapitre 5

Quantifying identifiability loss at macroscopic scales

5.1 Introduction

5.1.1 Motivation and statistical setting

Diffusion models are natural macroscopic models for phenomena observed at a large sampling rate. In finance asset prices or volumes change at discrete random times (see for instance Gerber and Shiu [42], Russell and Engle [89], Masoliver *et al.* [71] or Cont and de Larrard [26]), however it is common to use continuous diffusion processes in modelling (see for instance Masoliver *et al.* [71], Önalan [84], Cont and de Larrard [26] or Hong and Satchell [55]) when they are observed at a "large" sampling rates. In physics the opposition between large scale diffusion behaviour and random walks at small scale is popular ; for instance see Metzler and Klafter [77], Bouchaud and Geoges [14] or Uchaikin and Zolotarev [100] giving a wide range of physical examples, Fedotov and Méndez [34] for the modelling of a front wave or Fedotov and Iomin [35, 36] for the modelling of the proliferation tumor cells.

Yet the use of diffusion models may conceal the dynamics of the jumps that occur at smaller scales. In this Chapter we try to quantify the (possible) loss of information resulting from the use of diffusive models. In other words does the modelling by diffusions hide information on the jumps, information that could have been recovered otherwise. We come across problems of loss of information as introduced in Chapter 2 again, but this time we focus on macroscopic regimes.

In what follows, we derive results on a class of compound Poisson processes, denoted by $\mathcal{F}_0$ in the following. However, since a compound Poisson process is both a renewal reward process and a Lévy process some results derived in this Chapter hold for renewal reward processes and Lévy processes. Let R be a Poisson process of intensity ϑ, such that $R_0 = 0$.

Define the compound Poisson process X as

$$X_s = \sum_{i=1}^{R_s} \xi_i, \quad s \geq 0$$

where the (ξ_i) are independent and identically distributed random variables with density f with respect to the Lebesgue measure and independent of the (J_i). To identify the law of X, it is necessary to specify the pair $r = (\vartheta, f)$ giving the intensity and the jump density. Abusing notation slightly we will refer to r as the density of the compound Poisson process.

Definition 1. *The class $\mathcal{F}_0$ is the class of compound Poisson processes of density r such that the norm $\|r\|_{2,\mathcal{F}_0} := \|\vartheta f\|_2$ is finite*[1].

Suppose we have discrete observations of a process X in $\mathcal{F}_0$ over $[0, T]$ at a sampling rate $\Delta > 0$, namely we observe

$$\big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor \Delta}\big). \tag{5.1}$$

Definition 2. *The observations* (5.1) *are said to be on a macroscopic regime if $\Delta = \Delta_T \to \infty$ and $T/\Delta_T \to \infty$ as $T \to \infty$.*

The condition $T/\Delta_T \to \infty$ ensures there are asymptotically infinitely many observations. A typical example of macroscopic regime is a sampling rate Δ_T of the order of T^α as $T \to \infty$ for α in $(0, 1)$.

In a macroscopic regime, between two observations of X many jumps may occur. If X has finite variance, the central limit theorem applies and the law of the increments can be approximated by the law of a Gaussian random variable. The following Gaussian approximation in law

$$X_{i\Delta_T} - X_{(i-1)\Delta_T} \approx \sigma\big(W_{i\Delta_T} - W_{(i-1)\Delta_T}\big)$$

should hold, where W is a standard Wiener process and σ is positive. Hence in the macroscopic regime, only the variance parameter σ should be identifiable from (5.1). It follows that if r depends on more than 2 parameters, information should be lost. The asymptotic behaviour of the increments of a process X in $\mathcal{F}_0$ is not necessarily Gaussian, other stable laws can be encountered at the limit (see for instance Gnedenko and Kolmogorov [44], Kotulski [66] or Levy and Taqqu [69]). The limit law is parametric (see Gnedenko and Kolmogorov [44]) and may depend on fewer parameters than the number of parameter needed to describe the compound Poisson density r. In this Chapter we intend to answer the following questions which naturally arise :

i) When is the experiment generated by the increments of X asymptotically equivalent to a Gaussian experiment ?

1. Where $\|.\|_2$ denotes the usual L_2 norm, $\|f\|_2 = (\int_{\mathbb{R}} f(x)^2 dx)^{1/2}$.

- **ii)** Is it possible to recover the density r characterising a process in $\mathcal{F}_0$ from the observations (5.1) ?

Asymptotic equivalence of a Poisson experiment with variable intensity has been studied in Brown *et al.* [15]. Shevtsova [96] looks at the accuracy of Gaussian approximations for Poisson random sums.

5.1.2 Main results

We explore questions **i)** and **ii)**, but investigating them directly is sensitive, thus in Section 5.2 we first build and study a toy model. That toy model is a compound Poisson process with a drift that depends on a 2-dimensional parameter. The process studied in the toy model does not belong to $\mathcal{F}_0$. The identifiability of the parameter is investigated. In the sequel we say that the parameter is identifiable if it can be consistently estimated from the observations. We identify two distinct macroscopic regimes for that toy model.

- A regime where the parameter cannot be consistently estimated, giving a negative answer to **ii)**.
- A regime where the parameter can be estimated, giving a positive answer to **ii)**, but with slower optimal rates than the usual parametric ones (see Proposition 1 and Theorem 1 hereafter).

From the study of the toy model, we derive a lower bound in Theorem 2 which gives regimes where consistent estimation of the law generating a process in $\mathcal{F}_0$ is impossible, leading to a negative answer to **ii)**. Then we investigate question **i)** focusing on compound Poisson processes, in Theorem 3 we give asymptotic equivalence results. We distinguish regimes where the experiment generated by the observation of a process in $\mathcal{F}_0$ is asymptotically equivalent to a Gaussian experiment, giving a positive answer to **i)**. And regimes where compound Poisson processes depending on a too large number of parameter are not identifiable, giving a negative answer to **ii)**. In that latter case, the limit number of parameter beyond which consistent estimation is no longer possible can be related to α, if Δ_T is of the order of T^α when $T \to \infty$.

This Chapter is organised as follows, in Section 5.2 we construct and study a toy model and in Section 5.3 we establish the main Theorems 2 and 3. Section 5.5 is devoted to the proofs.

5.2 Loss of information in a toy model

5.2.1 Building up a toy model

Consider a Poisson process R with unknown intensity $\vartheta \in (0, \infty)$ and independent of the (ξ_i) which are independent and exponentially distributed random variables with unknown parameter $\lambda \in (0,\infty)$, we denote their density by

$$f_\lambda(x) = \lambda e^{-\lambda x}, \qquad x \in (0,\infty).$$

Define the Lévy process X as the sum of a compound Poisson process, with intensity ϑ and compound law f_λ, and a drift $-\vartheta/\lambda$

$$X_s = \sum_{i=1}^{R_s} \xi_i - \frac{\vartheta s}{\lambda}, \qquad s \geq 0. \tag{5.2}$$

The process X does not belong to $\mathcal{F}_0$. This toy model, without the drift, has been used by Alexandersson [2] to model rainfall. Suppose we observe increments of X over $[0,T]$ at a sampling rate $\Delta > 0$, namely we observe

$$\boldsymbol{X} = \big(X_{i\Delta} - X_{(i-1)\Delta}, \quad i = 1,\ldots,\lfloor T\Delta^{-1}\rfloor\big). \tag{5.3}$$

We obtain a family of statistical experiments indexed by Δ

$$\mathcal{X}^\Delta := \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{P}_\varpi^{T,\Delta}, \varpi \in \Theta\}\big),$$

where ϖ denotes the unknown parameter

$$\varpi = (\vartheta,\lambda) \in \Theta = (0,\infty)\times(0,\infty),$$

and $\mathbb{P}_\varpi^{T,\Delta}$ denotes the law of $\boldsymbol{X}$. The experiment $\mathcal{X}^\Delta$ is dominated by the Lebesgue measure μ_T on $\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}$. Since X is a Lévy process, the $X_{i\Delta} - X_{(i-1)\Delta}$ are independent and identically distributed, we denote by $p_{\varpi,\Delta}$ their density. The expression of the likelihood is the following

$$\frac{d\mathbb{P}_\varpi^{T,\Delta}}{d\mu_T}(\boldsymbol{X}, x) = \prod_{i=1}^{\lfloor T\Delta^{-1}\rfloor} p_{\varpi,\Delta}(X_{i\Delta} - X_{(i-1)\Delta}).$$

We intend to study the identifiability of the unknown parameter ϖ from (5.3) in the macroscopic regimes introduced in Definition 2. Since the (ξ_i) have finite variance, the central limit theorem gives the following approximation valid in law for Δ large enough

$$X_\Delta \approx \mathcal{N}(0, 2\vartheta\Delta/\lambda^2).$$

The parameter ϖ should not be identifiable when Δ is large, contrary to ϑ/λ^2.

5.2.2 Fisher information

We first study the behaviour of the Fisher information of $\mathcal{X}^\Delta$. Since the observations (5.3) are independent and identically distributed the Fisher information $I_{\lfloor T\Delta^{-1}\rfloor,\Delta}(\varpi)$ of $\mathcal{X}^\Delta$ verifies

$$I_{\lfloor T\Delta^{-1}\rfloor,\Delta}(\varpi) = \lfloor T\Delta^{-1}\rfloor I_{1,\Delta}(\varpi)$$

where

$$I_{1,\Delta}(\varpi) = \begin{pmatrix} -\mathbb{E}_{\mathbb{P}_\varpi}\left[\frac{\partial^2}{\partial\vartheta^2}\log p_{\varpi,\Delta}(X_\Delta,\vartheta,\lambda)\right] & -\mathbb{E}_{\mathbb{P}_\varpi}\left[\frac{\partial^2}{\partial\vartheta\partial\lambda}\log p_{\varpi,\Delta}(X_\Delta,\vartheta,\lambda)\right] \\ -\mathbb{E}_{\mathbb{P}_\varpi}\left[\frac{\partial^2}{\partial\lambda\partial\vartheta}\log p_{\varpi,\Delta}(X_\Delta,\vartheta,\lambda)\right] & -\mathbb{E}_{\mathbb{P}_\varpi}\left[\frac{\partial^2}{\partial\lambda^2}\log p_{\varpi,\Delta}(X_\Delta,\vartheta,\lambda)\right] \end{pmatrix}. \tag{5.4}$$

There is no closed form expression of $I_{\lfloor T\Delta^{-1}\rfloor,\Delta}(\varpi)$, the following proposition gives its asymptotic behaviour when Δ goes to infinity.

Proposition 1. *Let $\Delta = \Delta_T$ such that $\Delta_T \to \infty$ and $T/\Delta_T \to \infty$ as $T \to \infty$. We have*

$$\lim_{T\to\infty} I_{1,\Delta_T}(\varpi) = I(\varpi) := \begin{pmatrix} \frac{1}{2\vartheta^2} & -\frac{1}{\vartheta\lambda} \\ -\frac{1}{\vartheta\lambda} & \frac{2}{\lambda^2} \end{pmatrix}.$$

Moreover the eigenvalues of $I_{\lfloor T\Delta_T^{-1}\rfloor,\Delta_T}(\varpi)$, denoted by $e_{1,\Delta_T}(\varpi)$ and $e_{2,\Delta_T}(\varpi)$, verify

$$\begin{aligned} e_{1,\Delta_T}(\varpi) &= \frac{3}{4\lambda^2\vartheta + 16\vartheta^3}\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T} + O\Big(\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T^{3/2}}\Big) \\ e_{2,\Delta_T}(\varpi) &= \Big(\frac{2}{\lambda^2} + \frac{1}{2\vartheta^2}\Big)\lfloor T\Delta_T^{-1}\rfloor + \frac{3(7\lambda^4 + 40\lambda^2\vartheta^2 + 56\vartheta^4)}{8\lambda^2\vartheta^3(\lambda^2 + 4\vartheta^2)}\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T} + O\Big(\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T^{3/2}}\Big). \end{aligned}$$

Remark 1. *The matrix $I(\varpi)$ is the Fisher information with respect to (ϑ,λ) of an experiment consisting in one variable whose law is $\mathcal{N}(0, 2\vartheta/\lambda^2)$.*

We deduce from Proposition 1 that whenever $T/\Delta_T \to \infty$ and $T/\Delta_T^2 \to 0$ the Fisher information $I_{\lfloor T\Delta_T^{-1}\rfloor,\Delta_T}(\varpi)$ degenerates to a rank 1 matrix. Indeed we prove in Theorem 1 hereafter that it is not possible to build a consistent estimator of ϖ. On the contrary, when $T/\Delta_T \to \infty$ and $T/\Delta_T^2 \to \infty$ the parameter ϖ remains identifiable and consistent estimators of ϖ exists. However optimal estimators converge with a rate determined by the slowest eigenvalue $e_{1,\Delta_T}(\varpi)$ of the Fisher information which is $\lfloor T\Delta_T^{-2}\rfloor^{-1/2}$. It is a much slower rate than $\lfloor T\Delta_T^{-1}\rfloor^{-1/2}$, the square root of the sample size. In the case where there exists $\mathfrak{l} \in (0,\infty)$ such that $T/\Delta_T^2 \to \mathfrak{l}$, the slowest eigenvalue $e_{1,\Delta_T}(\varpi)$ is asymptotically finite and no consistent estimators of ϖ exist (see Theorem 1).

5.2.3 Loss of identifiability in macroscopic regimes : the toy model case

In the following Theorem $\|.\|$ denotes a norm on $\mathbb{R}^2$ and we set

$$diam(A) = \sup_{a_1,a_2\in A\times A} \|a_2 - a_1\|.$$

Theorem 1. *Let Δ_T be such that $\Delta_T \to \infty$ and $T\Delta_T^{-2} \to \mathfrak{l} \in [0, \infty)$ as $T \to \infty$. Then for all $\varpi_0 \in \Theta$ and $\delta > 0$ there exists $\mathcal{V}_\delta(\varpi_0) \subset \Theta$ a neighborhood of ϖ_0 such that diam$\big(\mathcal{V}_\delta(\varpi_0)\big) \leq \delta$, and*

$$\varliminf_{T\to\infty} \inf_{\widehat{\varpi}} \sup_{\varpi \in \mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi} - \varpi\|\big] > 0$$

where the infimum is taken over all estimators.

From Theorem 1, we see that there is no consistent estimator of ϖ when Δ_T goes rapidly to infinity. It was expected from the behaviour of the Fisher information in those regimes. Summarising the results obtained from that toy model, it appears that T/Δ_T^2 plays the role of a boundary.

- When $T/\Delta_T^2 \to 0$, for instance if Δ_T is of the order of T^α for $\alpha \in (1/2, 1)$, there is no consistent estimator of ϖ (see Theorem 1). The Fisher information of $\mathcal{X}^{\Delta_T}$ degenerates to a rank 1 matrix (see Proposition 1).
- When $T/\Delta_T^2 \to \mathfrak{l} > 0$, *i.e.* when $\alpha = 1/2$, there is no consistent estimator of ϖ (see Theorem 1). The first eigenvalue of the Fisher information of $\mathcal{X}^{\Delta_T}$ is finite (see Proposition 1).
- By Proposition 1, when $T/\Delta_T^2 \to \infty$, for instance for $\alpha \in (0, 1/2)$, optimal estimators of ϖ converge at the rate $\lfloor T\Delta_T^{-2}\rfloor^{-1/2}$. It is much slower than $\lfloor T\Delta_T^{-1}\rfloor^{-1/2}$, the square root of the sample size.

In that toy model, it is possible to distinguish the regimes for which ϖ is identifiable. In the next Section we derive non identifiability results for processes in $\mathcal{F}_0$, they are inspired by the study of that toy model.

5.3 Non identifiability at macroscopic scales in the general case

5.3.1 A lower bound

From Section 5.2 we derive a lower bound for the estimation from (5.1) of the law generating a process in $\mathcal{F}_0$.

Theorem 2. *Let $\Delta_T \to \infty$ be such that $T/\Delta_T \to \infty$ and $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ as $T \to \infty$. Then for all $r_0 \in \mathcal{F}_0$ and $\delta > 0$, there exist $\mathcal{V}_\delta(r_0)$, a neighborhood of r_0 such that diam$(\mathcal{V}_\delta(r_0)) \leq \delta$ and*

$$\varliminf_{T\to\infty} \inf_{\widehat{r}} \sup_{r \in \mathcal{V}_\delta(r_0)} \mathbb{E}_{\mathbb{P}_r}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r\|_{2,\mathcal{F}_0}\big] > 0$$

where the infimum is taken over all estimators.

The rate restriction $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ is technical. Indeed to prove Theorem 2 we show that, for Δ_T satisfying the rate restriction, the experiment $\mathcal{X}^{\Delta_T}$ introduced in Section 5.2 is close, for the total variation norm, to an experiment consisting of increments of a compound Poisson process. It permits to apply Theorem 1 to derive Theorem 2. Weakening that condition in $T/\Delta_T^2 = O(1)$, would ensure Theorem 2 to be true for regimes where Δ_T is of the order of $T^{1/2}$ as $T \to \infty$.

Theorem 2 gives no information on what happens when Δ_T is of the order of T^α as $T \to \infty$ for $\alpha \in (0, 1/2)$. Theorem 3 gives asymptotic equivalence results for processes in $\mathcal{F}_0$ in every macroscopic regime.

Remark 2. *It follows from Theorem 2 that whenever Δ_T is of the order of T^α for $\alpha \in (1/2, 1)$ it is not possible to estimate the density of a compound Poisson process from* (5.1)*. Compound Poisson processes are renewal reward processes and Lévy processes. From Theorem 2 we immediately derive that if Δ_T is such that $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ as $T \to \infty$, it is not possible to build consistent estimators of the law generating a renewal reward process or a Lévy process from* (5.1)*.*

5.3.2 An asymptotic equivalence result

Define for $K \geq 3$, the parameter $\rho = (\vartheta, m_2, \dots, m_K) \in \Sigma_K$ where Σ_K is a compact subset of $(0, \infty) \times \mathbb{R}^{K-1}$. Define for $\gamma > 0$ the function

$$h_\gamma : \rho \in \Sigma_K \to h_\gamma(\rho) = \left(\gamma\vartheta, \tfrac{m_2}{\gamma}, \dots, \tfrac{m_K}{\gamma}\right),$$

and the density function f_ρ with respect to the Lebesgue measure such that

$$\int x f_\rho(x)dx = 0, \qquad \text{and} \qquad \int x^k f_\rho(x)dx = m_k,\ k = 2, \dots, m_K. \tag{5.5}$$

Consider two compound Poisson processes Y and Z such that Y has intensity ϑ and compound density f_ρ satisfying (5.5) and Z has intensity $\gamma\vartheta$ and compound density $f_{h_\gamma(\rho)}$ satisfying (5.5).

Remark 3. *It is always possible to build a random variable with a density with respect to the Lebesgue measure imposing its $K-$first moments. Consider for instance the mixture of K Gaussian random variables*

$$\sum_{k=1}^K a_k \mathcal{N}(k, k),$$

and choose the (a_k) such that the K first moments match the desired ones. The construction is not unique since the same construction can be made with a mixture of any $M \geq K$ Gaussian random variables.

We also define the compound Poisson process U of intensity ϑ and of centred compound law whose second moment m_2 is finite and third moment m_3 is zero. The process U then depends on a parameter $\phi = (\vartheta, m_2) \in \Gamma$, where Γ is a compact subset of $(0,\infty) \times (0,\infty)$. Consider finally a centred Gaussian process W of quadratic variation $l(\phi) = \vartheta m_2$. Suppose we discretely observe Y, Z, U and W over $[0,T]$ at a sampling rate $\Delta > 0$, namely we observe

$$\big(Y_{i\Delta} - Y_{(i-1)\Delta}, i = 1, \dots, \lfloor T\Delta^{-1}\rfloor\big), \tag{5.6}$$
$$\big(Z_{i\Delta} - Z_{(i-1)\Delta}, i = 1, \dots, \lfloor T\Delta^{-1}\rfloor\big), \tag{5.7}$$
$$\big(U_{i\Delta} - U_{(i-1)\Delta}, i = 1, \dots, \lfloor T\Delta^{-1}\rfloor\big), \tag{5.8}$$
$$\big(W_{i\Delta} - W_{(i-1)\Delta}, i = 1, \dots, \lfloor T\Delta^{-1}\rfloor\big). \tag{5.9}$$

We consider the following families of statistical experiments indexed by Δ

$$\begin{aligned}
\mathcal{Y}_K^{\Delta} &:= \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{P}_\rho^{T,\Delta}, \rho \in \Sigma_K\}\big)\\
\mathcal{Z}_K^{\Delta} &:= \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{Q}_\rho^{T,\Delta}, \rho \in \Sigma_K\}\big)\\
\mathcal{U}^{\Delta} &:= \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{P}_\phi^{T,\Delta}, \phi \in \Gamma\}\big)\\
\mathcal{W}^{\Delta} &:= \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\mathbb{Q}_\phi^{T,\Delta}, \phi \in \Gamma\}\big),
\end{aligned}$$

where we denote by $\mathbb{P}_\rho^{T,\Delta}$ the law of (5.6), $\mathbb{Q}_\rho^{T,\Delta}$ the law of (5.7), $\mathbb{P}_\phi^{T,\Delta}$ the law of (5.8) and $\mathbb{Q}_\phi^{T,\Delta}$ the law of (5.9).

Theorem 3. *Let $\Delta_T \to \infty$ be such that $T/\Delta_T \to \infty$ as $T \to \infty$.*

1. *If $T\Delta_T^{-2} = o((\log(T/\Delta_T))^{-1/4})$, the experiments $\mathcal{U}^{\Delta_T}$ and $\mathcal{W}^{\Delta_T}$ are asymptotically equivalent.*
2. *Let $K \geq 3$, if $T\Delta_T^{-(K+1)/2} = o((\log(T/\Delta_T))^{-1/4})$, the experiments $\mathcal{Y}_K^{\Delta_T}$ and $\mathcal{Z}_K^{\Delta_T}$ are asymptotically equivalent.*

Remark 4. *We have $\mathcal{U}^{\Delta} \subset \mathcal{Y}_3^{\Delta}$ and equality only for experiments consisting in observing a compound Poisson process whose third moment is null. The two assertions of Theorem 3 do not always coincides.*

From Theorem 3.1. we deduce that when Δ_T goes rapidly to infinity, for instance if it is of the order of T^α as $T \to \infty$ for $\alpha \in (1/2, 1)$, the Gaussian approximation is valid. In other words it is not possible to distinguish the increments of a compound Poisson process from the increments of a Gaussian process. From Theorem 3.2. we derive that if Δ_T is of the order of T^α as $T \to \infty$ for $\alpha \in (0, 1/2)$ it is not possible to identify the $M \geq K_\alpha = \lceil \frac{2}{\alpha} - 1 \rceil$ first moments of the compound law and the intensity of a compound Poisson process from (5.1). Thus compound laws characterised by their $M \geq K_\alpha$ first moments cannot be identified.

Notice that $K = 1$ ($\rho = \vartheta$) is a similar case to the one treated in Chapter 2. The necessary assumption to estimate ρ is that the amount of data $T\Delta_T^{-1}$ is asymptotically large.

An estimator is provided in Chapter 2, in a special case, and is $\lfloor T\Delta_T \rfloor^{-1/2}$-optimal in all macroscopic regimes. The case $K = 2$ *i.e.* $\rho = (\vartheta, m_2)$ should behave similarly to the case treated in Section 5.2, where in some regime it is possible to estimate ρ but with slower rates of convergence than the usual parametric rates. In Section 5.3.1 a lower bound (Theorem 2) shows that it is not possible to estimate ρ when $T\Delta_T^{-2}$ is going to 0.

5.4 Discussion

Summary. From Proposition 1 and Theorems 1, 2 and 3 we distinguish the following macroscopic regimes that we express for Δ_T of the order of T^α, $\alpha \in (0,1)$ for readability. We set $K_\alpha = \lceil \frac{2}{\alpha} - 1 \rceil$. We add in the Table results immediately derived from Theorem 2 (see Remark 3) for renewal reward processes (a class we denote by $\mathcal{F}_1$) and for Lévy processes (class we denote by $\mathcal{F}_2$).

	Toy model (parameter $\varpi \in \mathbb{R}^2$)	Process in $\mathcal{F}_1$ or $\mathcal{F}_2$	Compound Poisson process
Slow regimes $\alpha \in (0, 1/2)$	$\sqrt{T\Delta_T^{-2}}$-optimal estimators		Non identifiability of the first $M \geq K_\alpha$ compound law moments and of the intensity
$\alpha = 1/2$	No consistent estimator		
Fast regimes $\alpha \in (1/2, 1)$	No consistent estimator	No consistent estimator	No consistent estimator Equivalent to Gaussian process

The last two columns of the table are incomplete. For the third column clues on how it can be completed are provided hereafter. The fourth column gives no information on compound Poisson processes depending on a small number of parameters, namely a number of parameter slower than K_α if Δ_T is of the order of T^α as $T \to \infty$ for $\alpha \in (0, 1/2)$. From the study of the toy model, we may conjecture that such processes can be estimated but with optimal rates of convergence that deteriorate as the number of parameters increases, as it was observed in the toy model for a 2 dimensional parameter.

Nonparametric estimation in macroscopic regimes is impossible. Theorem 2 ensures that for fast macroscopic regimes consistent estimation of processes in $\mathcal{F}_1$ or $\mathcal{F}_2$ is impossible. Theorem 3.2. completes the result and shows that at slow macroscopic scales, it is possible to build two different compound Poisson processes indistinguishable from (5.1). Thus it should be possible to derive a lower bound, similar of the one obtained in Theorem 2, ensuring that consistent (nonparametric) estimation from (5.1) of processes in $\mathcal{F}_0$ (and thus in $\mathcal{F}_1$ or $\mathcal{F}_2$ from Remark 3) is impossible ; making nonparametric inference for processes in $\mathcal{F}_1$ or $\mathcal{F}_2$ from (5.1) impossible in every macroscopic regimes.

The case where Δ_T goes to infinity slower than any power of T, *i.e.* $\alpha = 0$, has not been studied here. For Lévy processes, nonparametric estimation of the Lévy density is possible when Δ_T is fixed, possibly large (see for instance Neumann and Reiß [83] or Comte and Genon-Catalot [24]). In the case in between when Δ_T grows arbitrarily slowly to infinity, it may (presumably with deteriorated rates of convergence) or may not be possible to estimate (nonparametrically) processes in $\mathcal{F}_1$ or $\mathcal{F}_2$.

5.5 Proof

5.5.1 Proof of Proposition 1

Preparation

The increments of X are independent and identically distributed. First we compute the density of the Poisson part of X that we denote by $\mathbf{P}_\Delta[f_\lambda]$, keeping up with the notation of Chapters 3 and 4. We have for $x \geq 0$

$$\mathbf{P}_\Delta[f_\lambda](x) = \sum_{m=1}^{\infty} \mathbb{P}(R_\Delta = m | R_\Delta \neq 0) f_\lambda^{\star m}(x) = \frac{e^{-\vartheta\Delta}}{1-e^{-\vartheta\Delta}} \sum_{m=1}^{\infty} \frac{(\vartheta\Delta)^m}{m!} f_\lambda^{\star m}(x)$$

where $\star$ denotes the convolution product. Since f_λ is the density of an exponential distribution, $f_\lambda^{\star m}$ is the density of a gamma distribution, we derive for $x \geq 0$

$$\mathbf{P}_\Delta[f_\lambda](x) = \frac{e^{-\vartheta\Delta}}{1-e^{-\vartheta\Delta}} e^{-\lambda x} \vartheta\Delta\lambda \sum_{m=0}^{\infty} \frac{(\vartheta\Delta\lambda x)^m}{m!(m+1)!}.$$

Introduce the functions $k \in \mathbb{N}$ and $x \in [0, \infty)$ by

$$g_k(x) = \sum_{m=0}^{\infty} \frac{x^m}{m!(m+k)!}, \tag{5.10}$$

which is related to the modified Bessel function of the first kind $\mathcal{I}_k$ as follows

$$g_k(x) = \frac{1}{x^{k/2}} \mathcal{I}_k(2\sqrt{x}), \qquad x > 0, \tag{5.11}$$

where

$$\mathcal{I}_k(x) = \sum_{m=0}^{\infty} \left(\frac{x}{2}\right)^{2m+k} \frac{1}{m!\Gamma(m+k+1)}.$$

The density $\mathbf{P}_\Delta[f_\lambda]$ can be rewritten as

$$\mathbf{P}_\Delta[f_\lambda](x) = \frac{e^{-\vartheta\Delta}}{1-e^{-\vartheta\Delta}} e^{-\lambda x} \vartheta\Delta\lambda g_1(\vartheta\Delta\lambda x).$$

Adding the drift part we derive that the density $p_{\varpi,\Delta}$ of an increment is

$$p_{\varpi,\Delta}(x) = \frac{e^{-2\vartheta\Delta}}{1-e^{-\vartheta\Delta}} e^{-\lambda x} \vartheta\Delta\lambda g_1\big(\vartheta\Delta\lambda x + \vartheta^2\Delta^2\big), \qquad \forall x \geq -\vartheta\Delta/\lambda.$$

Some technical tools

To prove Proposition 1 we introduce some technical Lemmas.

Lemma 1. *The modified Bessel functions of the first kind $\mathcal{I}_k(x)$, $k \in \mathbb{N}$ verify for all $M \in \mathbb{N}$ as $x \to \infty$*

$$e^{-x}\mathcal{I}_k(x) = \frac{1}{(2\pi x)^{1/2}} \sum_{m=0}^{M} \frac{(-1)^m}{(2x)^m} \frac{\Gamma(k+m+\frac{1}{2})}{m!\Gamma(k-m+\frac{1}{2})} + O\Big(\frac{1}{x^{M+3/2}}\Big),$$

where the remainder depends on k and M.

Démonstration. Lemma 1 can be found in Watson [109]. □

We will need to compute the first moments of X_Δ. To that end we use its characteristic function given by Lévy-Kintchine formula

$$\phi_{X_\Delta}(w) = \mathbb{E}\big[e^{iwX_\Delta}\big] = \exp\big(\vartheta\Delta((1-iw/\lambda)^{-1} - 1 - iw/\lambda)\big)$$

and the fact that

$$\mathbb{E}\big[X_\Delta^m\big] = \frac{1}{i^m}\frac{\partial^m \phi_{X_\Delta}(w)}{\partial w^m}\Big|_{w=0}, \qquad m \in \mathbb{N}.$$

The asymptotic behaviour of those moments is given by the following Lemma, which apply in a more general setting than the one described in Section 5.2.

Lemma 2. *Let X_Δ be an increment of length Δ of a compound Poisson process whose compound law is centred and has moment up to order $K \in \mathbb{N}$. For Δ large enough we have for all $m \leq \lfloor \frac{K-1}{2} \rfloor$*

$$\mathbb{E}\big[X_\Delta^{2m}\big] \leq \mathfrak{C}_1\Delta^m \qquad \text{and} \qquad \big|\mathbb{E}\big[X_\Delta^{2m+1}\big]\big| \leq \mathfrak{C}_2\Delta^m,$$

where $\mathfrak{C}_1$ and $\mathfrak{C}_2$ continuously depend on ϑ.

Proof of Lemma 2. We prove the result we using Lévy-Kintchine formula and reasoning by induction on m. Lévy-Kintchine gives an explicit formula of the Fourier transform of X_Δ

$$\phi_{X_\Delta}(w) = \mathbb{E}\big[e^{iwX_\Delta}\big] = \exp\big(\vartheta\Delta(\widehat{f}(w) - 1)\big)$$

where $\widehat{f}(w) = \mathbb{E}[e^{iw\xi}]$ denotes the Fourier transform of the compound law ξ and ϑ the intensity of the associated Poisson process. Moreover we have the following connection between the moments of X_Δ and the derivatives of ϕ_{X_Δ}

$$\mathbb{E}\big[X_\Delta^m\big] = \frac{1}{i^m}\frac{\partial^m \phi_{X_\Delta}(w)}{\partial w^m}\Big|_{w=0}, \qquad m \in \mathbb{N}. \tag{5.12}$$

To show Lemma 2, we prove by induction the following property, for all $m \leq \left\lfloor \frac{K-1}{2} \right\rfloor$ we have

$$\frac{\partial^{2m}\phi_{X_\Delta}(w)}{\partial w^{2m}} = \big(P_{2m}(w,\Delta) + Q_{2m}(w,\Delta)\big)\exp\big(\vartheta\Delta(\widehat{f}(w)-1)\big)$$
$$\frac{\partial^{2m+1}\phi_{X_\Delta}(w)}{\partial w^{2m+1}} = \big(P_{2m+1}(w,\Delta) + Q_{2m+1}(w,\Delta)\big)\exp\big(\vartheta\Delta(\widehat{f}(w)-1)\big)$$

where the functions $\Delta \to P_{2m}(w,\Delta)$, $\Delta \to Q_{2m}(w,\Delta)$, $\Delta \to P_{2m+1}(w,\Delta)$ and $\Delta \to Q_{2m+1}(w,\Delta)$ are polynomials in Δ, the degree of Q_{2m} and Q_{2m+1} is lower than m and there exist

$$(c_{2m,j}(.), c_{2m+1,j}(.), j = 1,\dots,m)$$

C^1 functions continuously depending on ϑ, such that

$$P_{2m}(w,\Delta) = \sum_{j=1}^{m} c_{2m,j}(w)\widehat{f}'(w)^{2j}\Delta^{m+j}$$
$$P_{2m+1}(w,\Delta) = \sum_{j=1}^{m} c_{2m+1,j}(w)\widehat{f}'(w)^{2j-1}\Delta^{m+j}.$$

Straightforward computation permits to verify the result for $m = 1$

$$\frac{\partial^{2}\phi_{X_\Delta}(w)}{\partial w^{2}} = \big(\vartheta\Delta\widehat{f}^{(2)}(w) + (\vartheta\Delta\widehat{f}'(w))^2\big)\exp\big(\vartheta\Delta(\widehat{f}(w)-1)\big)$$
$$\frac{\partial^{3}\phi_{X_\Delta}(w)}{\partial w^{3}} = \big(\vartheta\Delta\widehat{f}^{(3)}(w) + 2\vartheta^2\Delta^2\widehat{f}'(w)\widehat{f}^{(2)}(w) + \vartheta\Delta\widehat{f}'(w)(\vartheta\Delta\widehat{f}^{(2)}(w) + (\vartheta\Delta\widehat{f}'(w))^2)\big)$$
$$\times \exp\big(\vartheta\Delta(\widehat{f}(w)-1)\big).$$

Assuming that the property holds for $m-1$, we have

$$\begin{aligned}\frac{\partial^{2m}\phi_{X_\Delta}(w)}{\partial w^{2m}} &= \frac{\partial}{\partial w}\frac{\partial^{2m-1}\phi_{X_\Delta}(w)}{\partial w^{2m-1}}\\ &= \big(\partial_w P_{2m-1}(w,\Delta) + \partial_w Q_{2m-1}(w,\Delta) + \vartheta\Delta\widehat{f}'(x)(P_{2m-1}(w,\Delta) + Q_{2m-1}(w,\Delta))\big)\\ &\quad\times \exp\big(\vartheta\Delta(\widehat{f}(w)-1)\big).\end{aligned}$$

Since

$$\begin{aligned}\partial_w P_{2m-1}(w,\Delta) &= c_{2m-1,1}(w)'\widehat{f}'(w)\Delta^m + c_{2m-1,1}(w)\widehat{f}''(w)\widehat{f}'(w)\Delta^m\\ &+ \sum_{j=1}^{m-2}\big(c_{2m-1,j+1}(w)'\widehat{f}'(w)^{2j+1}\Delta^{m+j} + c_{2m-1,j+1}(w)(2j+1)\widehat{f}''(w)\widehat{f}'(w)^{2j}\Delta^{m+j}\big)\end{aligned}$$

and

$$\vartheta\Delta\widehat{f}'(x)P_{2m-1}(w,\Delta) = \vartheta\sum_{j=1}^{m-1} c_{2m-1,j}(w)\widehat{f}'(w)^{2j}\Delta^{m+j}$$

we set

$$P_{2m}(w,\Delta) = \sum_{j=1}^{m-2} \big(c_{2m-1,j+1}(w)'\widehat{f}'(w) + c_{2m-1,j+1}(w)(2j+1)\widehat{f}''(w)\big)\widehat{f}'(w)^{2j}\Delta^{m+j}$$
$$Q_{2m}(w,\Delta) = \partial_w Q_{2m-1}(w,\Delta) + \vartheta\Delta\widehat{f}'(x)Q_{2m-1}(w,\Delta)$$

where P_{2m} has the desired property and we derive from the induction hypothesis that the degree of Q_{2m} is lower than m. Similar computations give the result for P_{2m+1} and Q_{2m+1}.

We conclude on the proof of Lemma 2 using the result we have just proven, (5.12) and the fact that $\widehat{f}(0) = 1$ and that f is centred : $\widehat{f}'(0) = 0$. It immediately follows that

$$\mathbb{E}\big[X_\Delta^{2m}\big] \leq \mathfrak{C}_1\Delta^m, \qquad \big|\mathbb{E}\big[X_\Delta^{2m+1}\big]\big| \leq \mathfrak{C}_2\Delta^m,$$

where $\mathfrak{C}_1$ and $\mathfrak{C}_2$ continuously depend on the intensity ϑ. □

Remark 5. *Lemma 2 provides a bound for the absolute value of the moments, but in what follows we need to bound $\mathbb{E}[|X_\Delta^{2m+1}|]$ for $m \geq 1$, provided it exists. Lemma 2 and the Cauchy-Schwarz inequality imply $\mathbb{E}[|X_\Delta^{2m+1}|] \leq \mathfrak{C}\Delta^{m+1/2}$.*

Completion of the proof of Proposition 1

Since the observations (5.3) are independent and identically distributed, the Fisher information of $\mathcal{E}^{\varpi,\Delta_T}$ verifies

$$I_{\lfloor T\Delta_T^{-1}\rfloor,\Delta_T}(\varpi) = \lfloor T\Delta_T^{-1}\rfloor I_{1,\Delta_T}(\varpi)$$

where

$$I_{1,\Delta_T}(\varpi) = \begin{pmatrix} I_{\Delta_T}(\vartheta,\vartheta) & I_{\Delta_T}(\vartheta,\lambda) \\ I_{\Delta_T}(\lambda,\vartheta) & I_{\Delta_T}(\lambda,\lambda) \end{pmatrix}$$

where $I_{\Delta_T}(\vartheta,\vartheta)$, $I_{\Delta_T}(\vartheta,\lambda)$, $I_{\Delta_T}(\lambda,\vartheta)$ and $I_{\Delta_T}(\lambda,\lambda)$ are defined in (5.4). From the symmetry of second derivatives and the regularity of $\varpi \to p_{\varpi,\Delta_T}$ we have $I_{\Delta_T}(\vartheta,\lambda) = I_{\Delta_T}(\lambda,\vartheta)$. From

(5.10) we immediately derive $g'_k(x) = g_{k+1}(x)$ and computations lead to

$$\begin{aligned}
I_{\Delta_T}(\vartheta,\vartheta) = \mathbb{E}_{\mathbb{P}_\varpi}\Big[&\frac{\Delta_T^2 e^{-2\vartheta\Delta_T}}{(1-e^{-\vartheta\Delta_T})^2} + \frac{\Delta_T^2 e^{-\vartheta\Delta_T}}{1-e^{-\vartheta\Delta_T}} + \frac{1}{\vartheta^2} - 2\Delta_T^2\frac{g_2(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}\\
&-(\lambda\Delta_T X_{\Delta_T}+2\vartheta\Delta_T^2)^2\\
&\times\Big(\frac{g_3(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)} - \Big(\frac{g_2(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}\Big)^2\Big)\Big]\\
I_{\Delta_T}(\vartheta,\lambda) = \mathbb{E}_{\mathbb{P}_\varpi}\Big[&-\Delta_T X_{\Delta_T}\frac{g_2(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}\\
&-\vartheta\Delta_T X_{\Delta_T}(\lambda\Delta_T X_{\Delta_T}+2\vartheta\Delta_T^2)\\
&\Big(\frac{g_3(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)} - \Big(\frac{g_2(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}\Big)^2\Big)\Big]\\
I_{\Delta_T}(\lambda,\lambda) = \mathbb{E}_{\mathbb{P}_\varpi}\Big[&\frac{1}{\lambda^2} - (\vartheta\Delta_T X_{\Delta_T})^2\Big(\frac{g_3(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)} - \Big(\frac{g_2(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}{g_1(\vartheta\Delta_T\lambda X_{\Delta_T}+\vartheta^2\Delta_T^2)}\Big)^2\Big)\Big].
\end{aligned}$$

Finally using (5.11), Lemma 1, Lemma 2, Remark 5 and usual Taylor expansions we derive the formula announced in Proposition 1 (computations are made with Mathematica.).

5.5.2 Proof of Theorem 1

Preliminary

Lemma 3. *Let Δ_T be such that $\Delta_T \to \infty$ and $T\Delta_T^{-2} \to \mathfrak{l} \in [0,\infty)$ as $T \to \infty$. We have for $\gamma > 0$*

$$\mathbb{E}_{\mathbb{P}_{\varpi_0}}\big[\log\big(\tfrac{g_1(\vartheta_0\lambda_0\Delta_T X_{\Delta_T}+\vartheta_0^2\Delta_T^2)}{g_1(\gamma^3\vartheta_0\lambda_0\Delta_T X_{\Delta_T}+\gamma^4\vartheta_0^2\Delta_T^2)}\big] = 2\vartheta_0\Delta_T(1-\gamma^2) + 3\log(\gamma) - \frac{9(\gamma^2-1)}{16\gamma^2\vartheta_0\Delta_T} + O\big(\tfrac{1}{\Delta_T^{3/2}}\big).$$

Démonstration. We derive the announced asymptotic expansion from (5.11), Lemma 1, Lemma 2 and Remark 5 (computations are made with Mathematica.). □

Completion of the proof of Theorem 1

We have the following inequality for all $\varpi_0 \in \Theta$ and $\delta > 0$

$$\sup_{\varpi\in\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big] \geq \int_{\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big]\mu(d\varpi) \tag{5.13}$$

where $\mathcal{V}_\delta(\varpi_0)$ is a neighborhood of ϖ_0 such that $diam(\mathcal{V}_\delta(\varpi_0)) < \delta$ and μ is the following measure on $\mathcal{V}_\delta(\varpi_0)$

$$\mu = \frac{1}{2}\big(\mathbb{1}_{\varpi_0} + \mathbb{1}_{h_\gamma(\varpi_0)}\big)$$

and $h_\gamma(\varpi_0) \in \mathcal{V}(\varpi_0)$ is a perturbation of ϖ_0 that we explicit later and $\mathbb{1}_\varpi$ denotes the Dirac distribution in ϖ. It follows that

$$\begin{aligned}\int_{\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big]\mu(d\varpi) &= \frac{1}{2}\big(\mathbb{E}_{\mathbb{P}_{\varpi_0}}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi_0\|\big] + \mathbb{E}_{\mathbb{P}_{h_\gamma(\varpi_0)}}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-h_\gamma(\varpi_0)\|\big]\big) \\ &= \frac{1}{2}\Big(\mathbb{E}_{\mathbb{P}_{\varpi_0}}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi_0\|\big] + \mathbb{E}_{\mathbb{P}_{\varpi_0}}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-h_\gamma(\varpi_0)\|\frac{d\mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor}}\big]\Big) \\ &\geq \mathbb{E}_{\mathbb{P}_{\varpi_0}}^{\lfloor T\Delta_T^{-1}\rfloor}\Big[\frac{e^{-s}}{2}(\|\widehat{\varpi}-\varpi_0\| + \|\widehat{\varpi}-h_\gamma(\varpi_0)\|)\mathbb{1}_{\Big\{\frac{d\mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor}} > e^{-s}\Big\}}\Big] \qquad (5.14)\end{aligned}$$

for any $s > 0$. We use the triangle inequality to obtain a lower bound of (5.14) that does not depend on $\widehat{\varpi}$. We get

$$\int_{\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big]\mu(d\varpi) \geq \frac{e^{-s}}{2}\|\varpi_0 - h_\gamma(\varpi_0)\|\mathbb{P}_{\varpi_0}\Big(\frac{d\mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor}} > e^{-s}\Big).$$

From Markov inequality and

$$\|\mathbb{P}-\mathbb{Q}\|_{TV} = \int |d\mathbb{P} - d\mathbb{Q}|$$

we derive that for any $s > 0$ and $\mathbb{P}$ and $\mathbb{Q}$ some probability we have

$$\mathbb{P}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}} > e^{-s}\Big) \leq 1 - \frac{1}{1-e^{-s}}\|\mathbb{P}-\mathbb{Q}\|_{TV}$$

where $\|.\|_{TV}$ denotes the total variation norm. It follows that for all $s > 0$

$$\int_{\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big]\mu(d\varpi) \geq \|\varpi_0 - h_\gamma(\varpi_0)\|\frac{e^{-s}}{2}\Big(1 - \frac{1}{1-e^{-s}}\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV}\Big).$$

Hence,

$$\int_{\mathcal{V}_\delta(\varpi_0)} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\varpi}-\varpi\|\big]\mu(d\varpi) \geq \|\varpi_0 - h_\gamma(\varpi_0)\|\Phi\big(\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV}\big) \qquad (5.15)$$

where

$$\Phi(x) = \sup_{s\in(0,\infty)} \frac{e^{-s}}{2}\Big(1 - \frac{1}{1-e^{-s}}x\Big) = \Phi(x) = \frac{(1-\sqrt{x})^2}{2}, \qquad x \in [0,1]. \qquad (5.16)$$

Thus for Φ to be strictly positive x must remain bounded away from 1.

Then if it is possible to choose h_γ such that

– $\|\mathbb{P}_{\varpi_0,\Delta_T}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0),\Delta_T}^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV} \leq \mathfrak{C}_1 < 1$ for some constant $\mathfrak{C}_1$,

– and $\|\varpi_0 - h_\gamma(\varpi_0)\| \geq \mathfrak{C}_2 > 0$ for some constant $\mathfrak{C}_2$ that may depend on ϖ_0 we obtain the lower bound announced in Theorem 1. We define the function

$$h_\gamma : \varpi \to h_\gamma(\varpi) = (\gamma^2\vartheta, \gamma\lambda)$$

where γ is a positive constant. By Pinsker's inequality we get

$$\big\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \sqrt{K\big(\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor}, \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\big)} = \sqrt{\tfrac{\lfloor T\Delta_T^{-1}\rfloor}{2} K\big(\mathbb{P}_{\varpi_0}, \mathbb{P}_{h_\gamma(\varpi_0)}\big)}, \tag{5.17}$$

where K is the Kullback divergence and

$$\begin{aligned} K\big(\mathbb{P}_{\varpi_0,\Delta_T}, \mathbb{P}_{h_\gamma(\varpi_0),\Delta_T}\big) &= \int_{-\infty}^{\infty} \big(\log(p_{\varpi_0}(x)) - \log(p_{h_\gamma(\varpi_0)}(x))\big) p_{\varpi_0}(x)dx \\ &= \int_{-\infty}^{\infty} \big(-2\vartheta_0\Delta(1-\gamma^2) - \lambda_0 x(1-\gamma) - 3\log(\gamma) \\ &\qquad + \log\big(\tfrac{g_1(\vartheta_0\lambda_0\Delta_T x + \vartheta_0^2\Delta_T^2)}{g_1(\gamma^3\vartheta_0\lambda_0\Delta_T x + \gamma^4\vartheta_0^2\Delta_T^2)}\big)\big) p_{\varpi_0}(x)dx \\ &= -2\vartheta_0\Delta(1-\gamma^2) - 3\log(\gamma) + \mathbb{E}_{\mathbb{P}_{\varpi_0}}\big[\log\big(\tfrac{g_1(\vartheta_0\lambda_0\Delta_T X_{\Delta_T} + \vartheta_0^2\Delta_T^2)}{g_1(\gamma^3\vartheta_0\lambda_0\Delta_T X_{\Delta_T} + \gamma^4\vartheta_0^2\Delta_T^2)}\big)\big]. \end{aligned}$$

It follows from Lemma 3 that

$$K\big(\mathbb{P}_{\varpi_0}, \mathbb{P}_{h_\gamma(\varpi_0)}\big) = \frac{9(1-\gamma^2)}{16\gamma^2\vartheta_0\Delta_T} + O\big(\tfrac{1}{\Delta_T^{3/2}}\big),$$

and

$$\big\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \sqrt{\frac{9(1-\gamma^2)}{32\gamma^2\vartheta_0}\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T} + O\big(\tfrac{T}{\Delta_T^{5/2}}\big)}. \tag{5.18}$$

Then if $T/\Delta_T^2 \to 0$, we have from (5.18) that for T large enough there exits $\mathfrak{C}_1 < 1$ such that

$$\big\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \mathfrak{C}_1 < 1.$$

The inequality holds for any choice of γ! Otherwise when $T/\Delta_T^2 \to \mathfrak{l} > 0$, we choose γ such that

$$0 < \frac{1}{\frac{16\vartheta_0}{9\mathfrak{l}} + 1} < \gamma^2, \tag{5.19}$$

and from (5.18) there also exists $\mathfrak{C}_1 < 1$ such that

$$\|\mathbb{P}_{\varpi_0}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_{h_\gamma(\varpi_0)}^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV} \leq \mathfrak{C}_1 < 1.$$

Notice that γ can be chosen either greater or lower than 1.

Second we bound from below $\|\varpi_0 - h_\gamma(\varpi_0)\|$, we work with the L_2 norm since all norms are equivalent. We have

$$\|\varpi_0 - h_\gamma(\varpi_0)\| = \sqrt{(1-\gamma^2)^2\vartheta_0^2 + (1-\gamma)^2\lambda_0^2} = |1-\gamma|\sqrt{(1+\gamma)^2\vartheta_0^2 + \lambda_0^2}.$$

We choose $\gamma \neq 1$ such that (5.19) is satisfied and such that $\|h_\gamma(\varpi_0) - \varpi_0\| < \delta$ to ensure $h_\gamma(\varpi_0) \in \mathcal{V}_\delta(\varpi_0)$, which is always possible since we can chose either $\gamma > 1$ or $\gamma < 1$ avoiding boundary issues. It follows that there exists $\mathfrak{C}_2 > 0$ depending on γ and ϖ_0 such that

$$\|\varpi_0 - h_\gamma(\varpi_0)\| \geq \mathfrak{C}_2 > 0. \tag{5.20}$$

We conclude the proof of Theorem 1 by plugging (5.18) and (5.20) into (5.15) and taking limits.

Remark 6. *To bound the total variation norm in* (5.17) *we prefer the Kullback divergence over the Hellinger distance since the logarithm renders easier the manipulation of the density* $p_{\varpi,\Delta}$ *(Lemma 3).*

5.5.3 Proof of Theorem 2

Notation

We keep up with the notation of Section 5.2, we denote as previously by ϖ the parameter $(\vartheta, \lambda) \in (0,\infty) \times (0,\infty)$ and define $t_\varpi = (\vartheta, f_\lambda)$. Consider the compound Poisson process V of intensity $\frac{8}{9}\vartheta$ and of jump density is f_λ, namely

$$V_s = \sum_{i=1}^{N_s} \epsilon_i, \quad s \geq 0 \tag{5.21}$$

where N is a Poisson process of intensity $\frac{8}{9}\vartheta$ and independent of the (ϵ_i) which are centred independent and identically distributed with density f_ϖ defined as

$$f_\varpi(x) = \tfrac{2}{3}\lambda e^{-\frac{2}{3}\lambda\left(x+1/(\frac{2}{3}\lambda)\right)}, \quad x \geq -1/\lambda. \tag{5.22}$$

We observe

$$\left(V_{i\Delta} - V_{(i-1)\Delta}, \quad i = 1, \ldots, \lfloor T\Delta^{-1}\rfloor\right). \tag{5.23}$$

and we denote by $\mathbb{Q}_\varpi^{\lfloor T\Delta_T^{-1}\rfloor}$ the law of (5.23).

Remark 7. *The choice of the multiplicative constants* $\frac{8}{9}$ *and* $\frac{2}{3}$ *in front of* ϑ *and* λ *ensures that* X_Δ *and* V_Δ *have same moments of order 2 and 3.*

Preliminary

The model generated by (5.23) is similar to the toy model introduced in Section 5.2 ; in (5.3) we have a Lévy process with a drift component whereas in (5.23) we observe a compound Poisson process belonging to $\mathcal{F}_0$. We cannot directly use the result of Theorem 1, instead we introduce the following Lemma which ensures that those two models are close for the total variation norm.

Lemma 4. *Let $\Delta_T \to \infty$ as $T \to \infty$ such that*

$$T/\Delta_T \to \infty \quad \text{and} \quad T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4}). \tag{5.24}$$

Then we have for any compact subset Θ of $(0,\infty) \times (0,\infty)$

$$\sup_{\varpi \in \Theta} \big\| \mathbb{P}_\varpi^{\lfloor T\Delta_T^{-1} \rfloor} - \mathbb{Q}_\varpi^{\lfloor T\Delta_T^{-1} \rfloor} \big\|_{TV} \to 0.$$

Proof of Lemma 4. The announced result is is implied by (see for instance Tsybakov [99])

$$\big\| \mathbb{P}_\varpi - \mathbb{Q}_\varpi \big\|_{TV} = o\big((T/\Delta_T)^{-1}\big), \tag{5.25}$$

uniformly over Θ, since each experiment is the $\lfloor T\Delta_T^{-1} \rfloor$-fold product of independent and identically distributed random variables[2]. Let us further denote by p_{ϖ,Δ_T} and q_{ϖ,Δ_T} the densities of X_{Δ_T} and of V_{Δ_T} respectively. We have

$$\big\| \mathbb{P}_\varpi - \mathbb{Q}_\varpi \big\|_{TV} = \int_{\mathbb{R}} |p_{\varpi,\Delta_T}(x) - q_{\varpi,\Delta_T}(x)| dx \leq I + II + III,$$

where, applying successively the triangle inequality and Cauchy-Schwarz inequality, we have for any $\eta > 0$,

$$\begin{aligned} I &= \sqrt{2\eta}\Big(\int_{\mathbb{R}} \big(p_{\varpi,\Delta_T}(x) - q_{\varpi,\Delta_T}(x)\big)^2 dx\Big)^{1/2}, \\ II &= \mathbb{P}_\varpi\big(|V_{\Delta_T}| \geq \eta\big), \\ III &= \mathbb{P}_\varpi\big(|X_{\Delta_T}| \geq \eta\big). \end{aligned}$$

Set $\eta = \eta_T = \kappa\sqrt{\Delta_T \log(T/\Delta_T)}$. We claim that for $\kappa^2 > 3\vartheta$, the terms I, II and III are $o\big((T/\Delta_T^{-1})\big)$ hence the result. For the term II we apply Bernstein inequality : assume $U_1, \ldots, U_n$ are centred independent random variables satisfying for a constant $\mathfrak{K}$

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}[U_i^m] \leq \frac{m!}{2}\mathfrak{K}^{m-2}, \quad \forall m \geq 2$$

2. For instance, by using the bound (see Tsybakov [99] pp. 83 – 90)

$$\big\| \mathbb{P}^{\otimes n} - \mathbb{Q}^{\otimes n} \big\|_{TV} \leq \sqrt{2}\big(1 - \big(1 - \tfrac{1}{2}\|\mathbb{P} - \mathbb{Q}\|_{TV}\big)^n\big)^{1/2}.$$

then for all $s > 0$ we have

$$\mathbb{P}\Big(\sum_{i=1}^{n} U_i \geq t\Big) \leq \exp\Big(-\frac{s^2}{s\mathfrak{K} + n + \sqrt{2s\mathfrak{K}n + n^2}}\Big). \tag{5.26}$$

By construction we have that $V_{\Delta_T} = \sum_{i=1}^{N_{\Delta_T}} \epsilon_i$ where ϵ_i has density f_ϖ given by (5.22), it follows that $\mathbb{E}[\epsilon_i] = 0$. We readily obtain from the fact that ϵ_i is a centred exponentially distributed random variable and the convex inequality that for all $m \geq 2$

$$\mathbb{E}\big[|\epsilon_i|^m\big] \leq 2^m \tfrac{2}{3}\lambda \int_0^\infty x^m e^{-\frac{2}{3}\lambda x} dx = 2\Big(\frac{3}{\lambda}\Big)^{m-1} m!$$

we set $\mathfrak{K} = 2.9/\lambda^2$. We now apply (5.26) conditional on N_{Δ_T} and derive that

$$\begin{aligned}
&\mathbb{P}_\varpi\big(|V_{\Delta_T}| \geq \eta_T\big)\\
&\quad\leq 2\mathbb{E}_{\mathbb{P}_\varpi}\Big[\exp\Big(-\frac{\eta_T^2}{\eta_T\frac{9}{\lambda^2}+N_{\Delta_T}+\sqrt{2\eta_T\frac{9}{\lambda^2}N_{\Delta_T}+N_{\Delta_T}^2}}\Big)\Big]\\
&\quad\leq 2\Big(\exp\Big(-\frac{\eta_T^2}{\eta_T\frac{9}{\lambda^2}+\kappa'\Delta_T+\sqrt{2\eta_T\frac{9}{\lambda^2}\kappa'\Delta_T+\kappa'^2\Delta_T^2}}\Big)\Big] + \mathbb{P}_\varpi\big(N_{\Delta_T} \geq \kappa'\Delta_T\big)\Big),
\end{aligned}$$

for every $\kappa' > 0$. We use Chernov inequality to bound the last term of the inequality. After replacing η_T by its value we get for T large enough

$$\mathbb{P}_\varpi\big(|V_{\Delta_T}| \geq \eta_T\big) \leq (T/\Delta_T)^{-\frac{\kappa^2}{3\kappa'}} + \exp\big(-\Delta_T(\kappa'\log(\kappa'/(\tfrac{8}{9}\vartheta)) - (\kappa' - \tfrac{8}{9}\vartheta))\big).$$

If we have $\kappa' < \kappa^2/3$ and $\kappa' > \vartheta$ then II is of the right order. For the term III we proceed as for II since

$$\mathbb{P}_\varpi\big(|X_{\Delta_T}| \geq \eta_T\big) \leq \mathbb{P}_\varpi\big(|V_{\Delta_T}| \geq \eta_T - \big|\tfrac{N_{\Delta_T}}{\lambda} - \tfrac{\vartheta\Delta_T}{\lambda}\big|\big).$$

We do not develop computations further. Thus II and III have the right order. It remains to bound the main term I. For that we use that $p_{\varpi,\Delta}$ is the density of an increment of a compound Poisson process with drift and $q_{\varpi,\Delta}$ the density of an increment of a compound Poisson process. By Plancherel equality we obtain the following

$$\begin{aligned}
&\int_{\mathbb{R}} \big(p_{\varpi,\Delta_T}(x) - q_{\varpi,\Delta_T}(x)\big)^2 dx\\
&\qquad= (2\pi)^{-1}\int_{\mathbb{R}} |\widehat{p}_{\varpi,\Delta_T}(\xi) - \widehat{q}_{\varpi,\Delta_T}(\xi)|^2 d\xi\\
&\qquad= (2\pi)^{-1}\int_{\mathbb{R}} \Big| e^{\vartheta\Delta_T\big(\frac{1}{1-i\xi/\lambda}-1\big)} e^{-\frac{i\xi\vartheta\Delta_T}{\lambda}} - e^{\frac{8}{9}\vartheta\Delta_T\big(\frac{1}{1-i3\xi/(2\lambda)}e^{-i3\xi/(2\lambda)}-1\big)}\Big|^2 d\xi\\
&\qquad\leq IV + V + VI,
\end{aligned}$$

with

$$IV = (2\pi)^{-1} \int_{|\xi| \le \rho\sqrt{\Delta_T}} \Big| \exp\big(\vartheta\Delta_T\big(\frac{1}{1 - i\frac{\xi}{\sqrt{\Delta_T}\lambda}} - 1\big)\big) \exp(-i\tfrac{\xi\vartheta\sqrt{\Delta_T}}{\lambda}) - \exp\big(\tfrac{8}{9}\vartheta\Delta_T\big(\frac{1}{1 - i\frac{3\xi}{2\sqrt{\Delta_T}\lambda}} e^{-i\frac{3\xi}{2\sqrt{\Delta_T}\lambda}} - 1\big)\big)\Big|^2 \frac{d\xi}{\sqrt{\Delta_T}},$$

$$V = (2\pi)^{-1} \int_{|\xi| \ge \rho\sqrt{\Delta_T}} \Big| \exp\big(\vartheta\Delta_T\big(\frac{1}{1 - \frac{i\xi}{\sqrt{\Delta_T}\lambda}} e^{-\frac{i\xi}{\sqrt{\Delta_T}\lambda}} - 1\big)\big)\Big|^2 \frac{d\xi}{\sqrt{\Delta_T}},$$

$$VI = (2\pi)^{-1} \int_{|\xi| \ge \rho\sqrt{\Delta_T}} \Big| \exp\big(\tfrac{8}{9}\vartheta\Delta_T\big(\frac{1}{1 - i\frac{3\xi}{2\sqrt{\Delta_T}\lambda}} - 1\big)\big)\Big|^2 \frac{d\xi}{\sqrt{\Delta_T}},$$

for any $\rho \ge 0$ and where we replaced ξ by $\xi/\sqrt{\Delta_T}$. By a first order expansion, we have that IV is less than

$$\int_{|\xi| \le \rho\sqrt{\Delta_T}} e^{-2\frac{\vartheta\xi^2}{\lambda^2}} \frac{\xi^8\alpha^2(\frac{\xi}{\sqrt{\Delta_T}})}{\Delta_T^2} e^{2\frac{\xi^4}{\Delta_T}\alpha\left(\frac{\xi}{\sqrt{\Delta_T}}\right)} \frac{d\xi}{\sqrt{\Delta_T}}$$

for some bounded function $\xi \rightsquigarrow \alpha(\xi)$. Set $\overline{\alpha} = \sup_x |\alpha(x)|$. We thus obtain that IV is less than a constant times

$$\int_{\mathbb{R}} e^{-2(\vartheta - \rho^2\overline{\alpha})\frac{\xi^2}{\lambda^2}} \frac{\xi^8\overline{\alpha}^2}{\Delta_T^2} \frac{d\xi}{\sqrt{\Delta_T}}.$$

We choose ρ such that $\vartheta - \rho^2\overline{\alpha} > 0$, the term IV is then of order $\Delta_T^{-5/2}$. For the term V and VI we readily obtain that

$$\begin{aligned} V &= (2\pi)^{-1} \int_{|\xi| \ge \rho} \Big| \exp\big(\vartheta\Delta_T\big(\tfrac{1}{1 - \frac{i\xi}{\lambda}} e^{-\frac{i\xi}{\lambda}} - 1\big)\big)\Big|^2 d\xi \\ &= (2\pi)^{-1} \int_{|\xi| \ge \rho} \exp\big(2\vartheta\Delta_T\big(\tfrac{\cos(\xi/\lambda)}{1 + \frac{\xi^2}{\lambda^2}} - \tfrac{\sin(\xi/\lambda)}{1 + \frac{\xi^2}{\lambda^2}} - 1\big)\big) d\xi \\ &\le (2\pi)^{-1} \int_{|\xi| \ge \rho} \exp\Big(2\vartheta\Delta_T \tfrac{-\frac{\xi^2}{\lambda^2}}{1 + \frac{\rho^2}{\lambda^2}}\Big) d\xi, \end{aligned}$$

$$VI = (2\pi)^{-1} \int_{|\xi| \ge \rho} \big| \exp\big(\tfrac{8}{9}\vartheta\Delta_T\big(\tfrac{1}{1 - \frac{i3\xi}{2\lambda}} - 1\big)\big)\big|^2 d\xi \le (2\pi)^{-1} \int_{|\xi| \ge \rho} \exp\Big(\tfrac{16}{9}\vartheta\Delta_T \frac{1 - \frac{9\xi^2}{4\lambda^2}}{1 + \frac{9\rho^2}{4\lambda^2}}\Big) d\xi,$$

it follows that if we choose $\rho > \frac{2}{3}\lambda$, then V and VI are of order $e^{-\mathfrak{C}\Delta_T}$ for some positive constant $\mathfrak{C}$ that depends on ρ and ϑ.

In conclusion, we have that $\int_{\mathbb{R}} \big(p_{\vartheta,\Delta_T}(x) - q_{\vartheta,\Delta_T}(x)\big)^2 dx$ is dominated by the term IV and is thus of order $\Delta_T^{-5/2}$. It follows that I is of order $\eta_T^{1/2}\Delta_T^{-5/4}$ and the choice $\eta_T = \kappa\sqrt{\Delta_T \log(T/\Delta_T)}$ implies $I = o\big((T/\Delta_T)^{-1}\big)$ thanks to the restriction condition $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$. We conclude the proof taking the supremum over Θ. □

Completion of the proof of Theorem 2

We have for all $r_0 \in \mathcal{F}_0$ and $\delta > 0$

$$\sup_{r\in\mathcal{V}_\delta(r_0)} \mathbb{E}_{\mathbb{P}_r}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r\|_{2,\mathcal{F}_0}\big] \geq \sup_{r_\varpi\in\mathcal{V}_\delta(r_{\varpi 0})} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0}\big]$$

where $\mathcal{V}_\delta(r_0)$ is a neighborhood of r_0 such that $diam(\mathcal{V}_\delta(r_0)) < \delta$, r_ϖ is defined from (5.21) and (5.22) and $\mathcal{V}_\delta(r_{\varpi 0})$ is a neighborhood of $r_{\varpi 0}$ such that $diam(\mathcal{V}_\delta(r_{\varpi 0})) < \delta$ and $\mathcal{V}_\delta(r_{\varpi 0}) \subset \mathcal{V}_\delta(r_0)$. Notice that

$$\inf_{\widehat{r}} \sup_{r_\varpi\in\mathcal{V}_\delta(r_{\varpi 0})} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0}\big] = \inf_{\widehat{r}\in\mathcal{V}_\delta(r_{\varpi 0})} \sup_{r_\varpi\in\mathcal{V}_\delta(r_{\varpi 0})} \mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0}\big].$$

Otherwise if $\widetilde{r} \notin \mathcal{V}_\delta(r_{\varpi 0})$, define $\Pi_{\mathcal{V}_\delta(r_{\varpi 0})}$ the projector over $\mathcal{V}_\delta(r_{\varpi 0})$, we immediately get for all $r_\varpi \in \mathcal{V}_\delta(r_{\varpi 0})$

$$\|\widetilde{r}-r_\varpi\|_{2,\mathcal{F}_0} \geq \|\Pi_{\mathcal{V}_\delta(r_{\varpi 0})}[\widetilde{r}]-r_\varpi\|_{2,\mathcal{F}_0}.$$

It follows that for all $\widehat{r},\ r_\varpi$ in $\mathcal{V}_\delta(r_{\varpi 0})$ we have

$$\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0} \leq 2(\delta + \|r_{\varpi 0}\|_{2,\mathcal{F}_0}). \tag{5.27}$$

The remainder of the proof is then a consequence of Scheffé's theorem, let F be a bounded function we have

$$\big|\mathbb{E}_{\mathbb{P}_r}[F(X)] - \mathbb{E}_{\mathbb{Q}_r}[F(X)]\big| \leq \|F\|_\infty \int |d\mathbb{P}_r - d\mathbb{Q}_r| = 2\|F\|_\infty \|\mathbb{P}_r - \mathbb{Q}_r\|_{TV}.$$

It follows from (5.27)

$$\mathbb{E}_{\mathbb{P}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0}\big] \geq \mathbb{E}_{\mathbb{Q}_\varpi}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\varpi\|_{2,\mathcal{F}_0}\big] - 2\big(2(\delta + \|r_{\varpi 0}\|)\|\mathbb{P}_\varpi - \mathbb{Q}_\varpi\|_{TV}\big).$$

We conclude the proof with Lemma 4, Theorem 1 and taking limits.

5.5.4 Proof of Theorem 3

First we prove part 2) of the Theorem ; we plan to prove that for Δ_T satisfying the rate restriction

$$T/\Delta_T^{(K+1)/2} = o((\log(T/\Delta_T))^{-1/4}), \tag{5.28}$$

the experiments $\mathcal{Y}_K^{\Delta_T}$ and $\mathcal{Z}_K^{\Delta_T}$ are asymptotically equivalent as $T \to \infty$. Experiments $\mathcal{Y}_K^{\Delta_T}$ and $\mathcal{Z}_K^{\Delta_T}$ live on the same state space $\mathbb{R}^{\lfloor T\Delta_T^{-1}\rfloor}$ and have smooth densities with respect to the Lebesgue measure, therefore to show the asymptotic equivalence of both experiments it is sufficient to show (see Le Cam and Yang [68])

$$\sup_{\rho\in\Sigma_K} \big\|\mathbb{P}_\rho^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{Q}_{h_\gamma(\rho)}^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \to 0 \ \text{ as } \ T \to \infty.$$

The bound

$$\big\|\mathbb{P}_{\rho}^{\lfloor T\Delta_T^{-1}\rfloor}-\mathbb{Q}_{h_\gamma(\rho)}^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV}\to 0 \text{ as } T\to\infty.$$

is implied in turn by the bound

$$\big\|\mathcal{L}(Y_{\Delta_T})-\mathcal{L}(Z_{\Delta_T})\big\|_{TV}=o\big((T/\Delta_T)^{-1}\big), \tag{5.29}$$

since each experiment is the $\lfloor T\Delta_T^{-1}\rfloor$-fold product of independent and identically distributed random variables[3]. To show that latter result we apply the same methodology as for the proof of Lemma 4. Let us further denote by $p_{\Delta_T,\rho}$ and $q_{\Delta_T,h_\gamma(\rho)}$ the densities of Y_{Δ_T} and Z_{Δ_T} respectively. We have

$$\|\mathcal{L}(Y_{\Delta_T})-\mathcal{L}(Z_{\Delta_T})\|_{TV}=\int_{\mathbb{R}}|p_{\Delta_T,\rho}(x)-q_{\Delta_T,h_\gamma(\rho)}(x)|dx\leq I+II+III,$$

where, applying successively the triangle inequality and Cauchy-Schwarz, for any $\eta>0$,

$$\begin{aligned}I&=\sqrt{2\eta}\Big(\int_{\mathbb{R}}\big(p_{\Delta_T,\rho}(x)-q_{\Delta_T,h_\gamma(\rho)}(x)\big)^2dx\Big)^{1/2},\\ II&=\mathbb{P}_\rho\big(|Y_{\Delta_T}|\geq\eta\big),\\ III&=\mathbb{P}_\rho\big(|Z_{\Delta_T}|\geq\eta\big).\end{aligned}$$

Set $\eta=\eta_T=\kappa\sqrt{\Delta_T\log(T/\Delta_T)}$. We claim that for $\kappa^2>2\vartheta m_2$, the terms I, II and III are $o\big((T/\Delta_T^{-1})\big)$ hence (5.29) and the result. For the term II we use that since $K\geq 3$, Y_{Δ_T} is a compound Poisson process whose compound law has finite variance m_2 we have

$$\frac{Y_{\Delta_T}}{\sqrt{\Delta_T}}\to\mathcal{N}(0,\vartheta m_2),\qquad \text{as } \Delta_T\to\infty.$$

Let $D\sim\mathcal{N}(0,\vartheta m_2)$, we have

$$\begin{aligned}\mathbb{P}_\rho\big(|Y_{\Delta_T}|\geq\eta_T\big)\leq{}&\mathbb{P}\big(|D|\geq\kappa\sqrt{\log(T/\Delta_T)}\big)\\ &+\big|\mathbb{P}\big(|D|\geq\kappa\sqrt{\log(T/\Delta_T)}\big)-\mathbb{P}_\rho\big(|Y_{\Delta_T}/\sqrt{\Delta_T}|\geq\kappa\sqrt{\log(T/\Delta_T)}\big)\big|\end{aligned}$$

where we readily obtain that

$$\mathbb{P}\big(|D|\geq\kappa\sqrt{\log(T/\Delta_T)}\big)\leq 2(T/\Delta_T)^{-\kappa^2/(2\vartheta m_2)}=o\big((T/\Delta_T)^{-1}\big)$$

3. For instance, by using the bound (see Tsybakov [99] pp. 83 – 90)

$$\big\|\mathbb{P}^{\otimes n}-\mathbb{Q}^{\otimes n}\big\|_{TV}\leq\sqrt{2}\big(1-\big(1-\tfrac{1}{2}\|\mathbb{P}-\mathbb{Q}\|_{TV}\big)^n\big)^{1/2}.$$

and the second term is bounded using Edgeworth series. Indeed since $K \geq 3$, the compound law has finite moment of order 3 and we obtain from Edgeworth series that

$$\begin{aligned}
\big|\mathbb{P}\big(|D| \geq & \kappa\sqrt{\log(T/\Delta_T)}\big) - \mathbb{P}_\rho\big(|Y_{\Delta_T}/\sqrt{\Delta_T}| \geq \kappa\sqrt{\log(T/\Delta_T)}\big)\big| \\
&\leq \Big|\frac{\mathfrak{C}}{\sqrt{\Delta_T}}\frac{\partial^3}{\partial x^3}\int_x^\infty e^{-\frac{s^2}{2\vartheta m_2}}ds\Big|_{x=\kappa\sqrt{\log(T/\Delta_T)}}\Big| \\
&= \frac{\mathfrak{C}}{\sqrt{\Delta_T}}\Big|1-\frac{\kappa\sqrt{\log(T/\Delta_T)}}{\vartheta m_2}\Big|e^{-\frac{\kappa^2\log(T/\Delta_T)}{2\vartheta m_2}} = \frac{\mathfrak{C}\sqrt{\log(T/\Delta_T)}}{\sqrt{\Delta_T}}(T/\Delta_T)^{-\kappa^2/(2\vartheta m_2)} \\
&= o\big((T/\Delta_T)^{-1}\big)
\end{aligned}$$

where $\mathfrak{C}$ continuously depends on ϑ, m_2 and m_3 and which is $o\big((T/\Delta_T)^{-1}\big)$ for $\kappa^2 \geq 2\vartheta m_2$. The term III is treated similarly as II, we do not reproduce computations. Thus II and III have the right order and the choice of κ and the bounds on II and III can be made independent of ρ taking the supremum of the upper bound over the compact set Σ_K in ϑ, m_2 and m_3.

It remains to bound the main term I. By Plancherel equality we obtain the following explicit expression :

$$\begin{aligned}
\int_{\mathbb{R}}\big(p_{\Delta_T,\rho}(x) - q_{\Delta_T,h_\gamma(\rho)}(x)\big)^2dx &= (2\pi)^{-1}\int_{\mathbb{R}}(\widehat{p}_{\Delta_T,\rho}(\xi) - \widehat{q}_{\Delta_T,h_\gamma(\rho)}(\xi))^2d\xi \\
&= (2\pi)^{-1}\int_{\mathbb{R}}\Big(e^{\vartheta\Delta_T(\widehat{f}_\rho(\xi)-1)} - e^{\gamma\vartheta\Delta_T(\widehat{f}_{h_\gamma(\rho)}(\xi)-1)}\Big)^2d\xi \\
&= (2\pi)^{-1}\int_{\mathbb{R}}\Big(e^{\vartheta\Delta_T\big(\widehat{f}_\rho(\frac{\xi}{\sqrt{\Delta_T}})-1\big)} - e^{\gamma\vartheta\Delta_T(\widehat{f}_{h_\gamma(\rho)}(\frac{\xi}{\sqrt{\Delta_T}})-1)}\Big)^2\frac{d\xi}{\sqrt{\Delta_T}} \\
&\leq IV + V + VI,
\end{aligned}$$

with

$$\begin{aligned}
IV &= (2\pi)^{-1}\int_{|\xi|\leq\eta\sqrt{\Delta_T}}\Big(e^{\vartheta\Delta_T(\widehat{f}_\rho(\frac{\xi}{\sqrt{\Delta_T}})-1)} - e^{\gamma\vartheta\Delta_T(\widehat{f}_{h_\gamma(\rho)}(\frac{\xi}{\sqrt{\Delta_T}})-1)}\Big)^2\frac{d\xi}{\sqrt{\Delta_T}}, \\
V &= (2\pi)^{-1}\int_{|\xi|\geq\eta\sqrt{\Delta_T}}e^{2\vartheta\Delta_T\big(\widehat{f}_\rho(\frac{\xi}{\sqrt{\Delta_T}})-1\big)}\frac{d\xi}{\sqrt{\Delta_T}}, \\
VI &= (2\pi)^{-1}\int_{|\xi|\geq\eta\sqrt{\Delta_T}}\widehat{q}_{\Delta_T,h_\gamma(\rho)}(\tfrac{\xi}{\sqrt{\Delta_T}})\frac{d\xi}{\sqrt{\Delta_T}},
\end{aligned}$$

for any $\eta \geq 0$. Since f_ρ and $f_{h_\gamma(\rho)}$ have finite $K-$first moments, we have the following expansion for any bounded w

$$\widehat{f}_\rho(w) - \Big(1 - \frac{m_2w^2}{2} + \ldots + \frac{i^Km_Kw^K}{K!}\Big) = w^{K+1}\alpha_1(w)$$

and

$$\widehat{f}_{h_\gamma(\rho)}(w) - \Big(1 - \frac{m_2w^2}{2\gamma} + \ldots + \frac{i^Km_Kw^K}{K!\gamma}\Big) = w^{K+1}\alpha_2(w)$$

for some bounded functions $w \rightsquigarrow \alpha_1(w)$ and $w \rightsquigarrow \alpha_2(w)$. It follows that that IV is less than

$$\int_{|\xi|\leq \eta\sqrt{\Delta_T}} \Big| e^{-\vartheta m_2 \frac{\xi^2}{2}+\ldots+i^K\vartheta m_K \frac{\xi^K}{\sqrt{\Delta_T}^{K-2}K!}} \Big|^2 \frac{\xi^{2K+2}}{\Delta_T^{K-1}} \alpha^2\big(\tfrac{\xi}{\sqrt{\Delta_T}}\big) e^{2\frac{\xi^{K+1}}{\sqrt{\Delta_T}^{K-1}}\alpha\left(\frac{\xi}{\sqrt{\Delta_T}}\right)} \frac{d\xi}{\sqrt{\Delta_T}}$$

for some bounded function $\xi \rightsquigarrow \alpha(\xi)$. Set $\overline{\alpha} = \sup_x |\alpha(x)|$. We thus obtain that IV is less than a constant times

$$\int_{|\xi|\leq \eta\sqrt{\Delta_T}} \exp\Big(- \big(\vartheta(m_2 - \sum_{k=2}^{\lfloor K/2 \rfloor} \tfrac{(-1)^k m_{2k}\eta^{2k-2}}{2(2k)!}) + 2\eta^{K-1}\overline{\alpha}\big)\xi^2\Big) \frac{\xi^{2K+2}}{\Delta_T^{K-1}} \overline{\alpha}^2 \frac{d\xi}{\sqrt{\Delta_T}}.$$

If we pick η such that

$$\vartheta(m_2 - \sum_{k=2}^{\lfloor K/2 \rfloor} \tfrac{(-1)^k m_{2k}\eta^{2k-2}}{2(2k)!}) > 2\eta^{K-1}\overline{\alpha},$$

the term IV is of order $\Delta_T^{-(2K-1)/2}$. For the terms V and VI we apply Riemann-Lebesgue Lemma that ensures that

$$\lim_{|\xi|\to\infty} \widehat{f}_\rho(\xi) = 0 \quad \text{and} \quad \lim_{|\xi|\to\infty} \widehat{f}_{h_\gamma(\rho)}(\xi) = 0.$$

Then V and VI are of order $e^{-\eta\vartheta\Delta_T}$.

In conclusion, we have that $\int_{\mathbb{R}} \big(p_{\Delta_T,\rho}(x) - q_{\Delta_T,\rho}(x)\big)^2 dx$ is dominated by the term IV and is thus of order $\Delta_T^{-(2K-1)/2}$. It follows that I is of order $\eta_T^{1/2}\Delta_T^{-(2K-1)/4}$ and the choice $\eta_T = \kappa\sqrt{\Delta_T \log(T/\Delta_T)}$ implies $I = o\big((T/\Delta_T)^{-1}\big)$ thanks to the rate restriction (5.28). Finally we take the supremum in ρ over the compact set Σ_K. The proof of part 1) is the same as for part 2) having $K = 3$ and replacing ρ by ϕ, m_3 by 0, $\mathcal{Y}_3^{\Delta_T}$ by $\mathcal{U}^{\Delta_T}$ and $\mathcal{Z}_3^{\Delta_T}$ by $\mathcal{W}^{\Delta_T}$. Proof of Theorem 3 is complete.

Remark 8. *To prove Theorem 3 of Chapter 2 we also showed an asymptotic equivalence but we had then the harsher rate restriction*

$$T/\Delta_T^{1+1/4} = o((\log(T/\Delta_T))^{-1/4}).$$

Indeed we studied then a compound Poisson process which was not absolutely continuous with respect to the Lebesgue measure, we had to add a regularising kernel from which the condition was imposed. It is unnecessary here because the compound law is absolutely continuous with respect to the Lebesgue measure.

Bibliographie

[1] Ait-Sahalia, Y. and Mykland, P.A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, **71**, 483–549.

[2] Alexandersson, H. (1985). A Simple Stochastic mode of a Precipitation Process. *Journal of climate and applied meteorology*, **24**, 1285–1295.

[3] Alvarez, E.E. (2005). Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey. *Methodology and Computing in Applied Probability*, **7**, 119–130.

[4] Anderson, P. L. and Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resources Research,* **34**, 2271–2280.

[5] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2011). Modeling microstructure noise with mutually exciting point processes. *Arxiv preprint,* 11013422v1.

[6] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2012). Scaling limits for Hawkes processes and application to financial statistics. *Arxiv preprint,* 12020842v1.

[7] Baricz, Á. (2008). Functional inequalities involving Bessel and modified Bessel functions of the first kind. *Expositiones Mathematicae*, **26**, 279–293.

[8] Bauwens, L. and Hautsch, N. (2006). Modelling high frequency financial data using point processes. *Discussion paper.*

[9] Bec, M. and Lacour, C. (2012). Adaptive kernel estimation of the Lévy density, *Hal preprint*, 00583221, version 2.

[10] Berkowitz, B., Cortis, A., Dentz, M. and Scher, H. (2006). Modeling non-Fickian transport in geological formations as a continuous time random walk. *Reviews of Geophysics,* **44**.

[11] Billingsley, P. (1999). *Convergence of probability measures.* Second edition. Wiley Series in Probability and Statistics.

[12] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, **81**, 637–654.

[13] Bøgsted, M. and Pitts, S. (2010). Decompounding random sums : a nonparametric approach. *Annals of the Institute of Statistical Mathematics,* **62**, 855–872.

[14] Bouchaud, J-P. and Georges, A. (1990). Anomalous diffusion in disordered media : statistical mechanism, models and physical applications. *Physics Reports*, **195**, 127–293.

[15] Brown, L.D., Carter, A.V., Low, M.G. and Zhang, C-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *The Annals of Statistics*, **32**, 2074–2097.

[16] Buchmann, B. and Grübel, R. (2003). Decompounding : an estimation problem for Poisson random sums. *The Annals of Statistics,* **31**, 1054–1074.

[17] Buchmann, B. and Grübel, R. (2004). Decompounding Poisson random sums : recursively truncated estimates in the discrete case. *Annals of the Institute of Statistical Mathematics*, **56**, 743–756.

[18] Buroni, R., Caniparoli, L. and Vezzani, A. (2011). Lévy walks and scaling in quenched disordered media. *Arxiv preprint*, 10032161v3.

[19] Cohen, A. (2003). *Numerical Analysis of wavelet methods.* Studies in Mathematics and its Applications, **32**, North-Holland Publishing.

[20] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **3**, 514–543.

[21] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2010). Nonparametric estimation for a stochastic volatility model. *Finance and Stochastics*, **14**, 49–80.

[22] Comte, F. and Genon-Catalot, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, **119**, 4088–4123.

[23] Comte, F. and Genon-Catalot, V. (2009). Nonparametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Statistica Neerlandica*, **64**, 290–313.

[24] Comte, F. and Genon-Catalot, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'I.H.P., Probability and Statistics*, **46**, 595–617.

[25] Comte, F. and Genon-Catalot, V. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *The Annals of Statistics*, **39**, 803–837.

[26] Cont, R. and de Larrard, A. (2011). Price dynamics in a Markovian limit order market. *Arxiv preprint,* 1104.4596v1.

[27] Cuppen, H.M., Morata, O. and Herbst, E. (2006). Monte Carlo simulations of H_2 formation on stochastically heated grains. *Arxiv preprint,* 0601554v1.

[28] Daley, D.J. et Vere-Jones, D. (1988). *An introduction to the theory of point processes.* Springer, New York.

[29] Al Dayri, K.A. (2012). Market Microstructure and Modeling of the Trading Flow. *PhD thesis, École Polytechnique.*

[30] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet Thresholding. *The Annals of Statistics*, **24**, 508–539.

[31] Dedecker, J., Doukhan, P. Lang, G., León, R.J. Louhichi, S. and Prieur, C. (2007). *Weak Dependence. With Examples and Applications.* Springer. Lecture Notes in Statistics.

[32] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

[33] Errais E., Giesecke, K. and Goldberg, L.R. (2006). Pricing credit from the top down with affine point processes. *Preprint.*

[34] Fedotov, S. and Méndez, V. (2002). Continuous-time random walks and traveling fronts. *Physical review E*, **66**, 030102.

[35] Fedotov, S. and Iomin, A. (2007). Migration and Proliferation Dichotomy in Tumor-Cell Invasion. *Physical Review Letters,* **98.** 118101.

[36] Fedotov, S. and Iomin, A. (2008). Probabilistic approach to a proliferation and migration dichotomy in the tumor cell invasion. *Arxiv preprint*, 0711.1304v2.

[37] Feller, W. (1971). *An Introduction to Probability Theory and Its Application.* Volume II, Second Edition. Wiley series in Probability and Mathematical Statistics.

[38] Figueroa-López, J.E. and Houdré, C. (2006). Risk bounds for the nonparametric estimation of Lévy processes. *IMS Lecture Notes-Monograph Series, High dimensional probability*, **51**, 96–116.

[39] Garavaglia, E. and Pavani, R. (2011). About earthquake forecasting by Markov renewal processes. *Methodology and Computing in Applied Probability,* **13**, 155–169.

[40] Genon-Catalot, V., Laredo, C. and Picard, D. (1992). Nonparametric Estimation of the Diffusion Coefficient by Wavelets Methods. *Scandinavian Journal of Statistics*, **19**, 317–335.

[41] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusive coefficient for multi-dimensional diffusion processes. *Annales de l'I.H.P., section B*, **29**, 119–151.

[42] Gerber, H.U. and Shiu, E. (1998). Pricing perpetual options for jump processes. *The North American Actuarial Journal*, **2**, 101–112.

[43] Gill, R.D. and Keidin, N. (2010). Product-limit estimators of the gap time distribution of renewal process under different sampling patterns. *Arxiv preprint,* 10030182v1.

[44] Gnedenko, V. and Kolmogorov, A.N. (1962). *Limit distributions for sums of independent random variables.* Addison-Wesley Publishing company, inc.

[45] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, **38**, 711–737.

[46] Gobet, E., Hoffmann, M. and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, **32**, 2223–2253.

[47] Guédon, Y. and Cocozza-Thivent, C. (2003). Nonparametric estimation of renewal processes from count data. *The Canadian Journal of Statistics,* **31**, 191–223.

[48] Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *Journal of Nonparametric Statistics*, **21**, 321–343.

[49] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications.* Springer-Verlag, New York.

[50] Hawkes, A. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**, 83–90.

[51] Hawkes, A.G. (1971b). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society : Series B*, **33**, 438–443.

[52] Helmstetter, A. and Sornette, D. (2002). Diffusion of epicenters of earthquake aftershocks, Omori's law and generalized continuous-time random walk models. *The American Physical Society*, **66**, 061104.

[53] Hewlett, P. (2006). Clustering of order arrivals, price impact and trade path optimisation. *Working paper*, University of Oxford.

[54] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, 135–163.

[55] Hong, K.J. and Satchell, S. (2012). Defining Single Asset Price Momentum in Terms of a Stochastic Process. *Theoretical Economics Letters*, **2**, 274–277.

[56] Huelsenbeck, J.P., Larget, B., Swofford, D. (2000). A Compound Poisson Process for Relaxing the Molecular Clock. *Genetics Society of America*, **154**, 1879–1892.

[57] Ibragimov, I.A. and Hasminskii, R.Z (1981). *Statistical Estimation. Asymptotic Theory.* Springer-Verlag.

[58] Jacod, J. (1993). Random sampling in estimation problems for continuous Gaussian processes with independent increments. *Stochastic Processes and their Applications,* **44**, 181–204.

[59] Jacod, J. (2006). Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli*, **12**, 383–401.

[60] Jeon, J., Tejedor, V., Burov, S., Barkai, E., Selhuber-Unkel, C., Berg-Sørensen, K., Oddershede, L. and Matzler, R. (2010). In vivo anomalous diffusion and weak ergodicity breaking of lipid granules. *Arxiv preprint,* 10100347v2.

[61] Jongbloed, G., van der Meulen, F. H. and van der Vaart, A. W. (2005). Nonparametric inference for Lévy driven Ornstein-Uhlenbeck processes. *Bernoulli,* **11**, 759–791.

[62] Jurlewicz, A., Wyłomańska, A. and Żebrowski, P. (2009). Coupled continuous-time random walk approach to the Rachev-Rüschendorf model for financial data. *Physica A*, **388**, 407–418.

[63] Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, **9**, 283–344.

[64] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, **24**, 211–229.

[65] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.

[66] Kotulski, M. (1995). Asymptotic Distributions of the Continuous-Time Random Walks : A Probabilistic Approach. *Journal of Statistical Physics*, **81**, 777–792.

[67] Lawrence, J.K., Cadavid, A.C., Ruzmaikin, A. and Berger, T.E. (2001). Spatio-Temporal Scaling of Solar Surface Flows. *Arxiv preprint*, 0101224v4.

[68] Le Cam, L. and Yang, L.G. (2000) *Asymptotics in Statistics : Some Basic Concepts*. 2nd edition. Springer-Verlag, New York.

[69] Levy, J.B. and Taqqu, M.S. (2000). Renewal reward processes with heavy-tailed inter-renewal times and heavy-tailed rewards. *Bernoulli*, **6**, 23–44.

[70] Lindvall, T. (1992). *Lectures on the coupling method*. Dover Publications.

[71] Masoliver, J., Montero, M., Perelló, J. and Weiss, G.H. (2008). Direct and inverse problems with some generalizations and extensions. *Arxiv preprint,* 0308017v2.

[72] Masuda, H. (2006). Likelihood Estimation of Stable Levy Processes from Discrete Data. *MHF Preprint Series*.

[73] Meerschaert, M.M. and Scheffler, H-P. (2004). Limit theorems for continuous-time random walks with infinite mean waiting times. *Journal of Applied Probability,* **41**, 623–638.

[74] Meerschaert, M.M. and Scheffler, H-P. (2005). Limit theorems for continuous time random walks with slowly varying waiting times. *Statistics & Probability Letters,* **71**, 15–22.

[75] Meerschaert, M.M. and E. Scalas, E. (2006). Coupled continuous time random walk in finance. *Physica A*, **370**, 114–118.

[76] Metzler, R. and Klafter, J. (2000). The random walk's guide to anomalous diffusion : a fractional dynamics approach. *Physics Reports,* **339**, 1–77.

[77] Metzler, R. and Klafter, J. (2004). The restaurant at the end of the random walk : recent developments in the description of anomalous transport by fractional dynamics. *Journal of Physics A,* **37**, 161–208.

[78] Moharir, P.S. (1992). Estimation of the compounding distribution in the compound Poisson process model for earthquakes. *Proceedings of the Indian Academy of Science*, **101**, 347–359.

[79] Montroll, E.W. and Weiss, G.H. (1965). Random Walks on Lattices. II. *Journal of Mathematical Physics,* **6,** 167–181.

[80] Montroll, E.W. and Scher, H. (1973). Random Walks on Lattices. IV. Continuous-Time Walks and Influence of Absorbing Boundaries. *Journal of Statistical Physics,* **9**, 101–135.
distribution from indirect measurement. *Bernoulli,* **13 (2)**, 365–388.

[81] Nasell, I. (1974). Inequalities for Modified Bessel Functions. *Mathematics of Computing*, **28**, 253–256.

[82] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *The Annals of Statistics,* **24**, 2399–2430.

[83] Neumann, M. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, **15**, 223–248.

[84] Önalan, Ö. (2010). Fractional Ornstein-Uhlenbeck Processes Driven by Stable Lévy Motion in Finance. *International Research Journal of Finance and Economics*, **42**, 129-139.

[85] Pham, D.T. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Mathematische Operations Forschung Und Statistik - Statistics*, **12**, 61–73.

[86] Reiß, M. (2006). Nonparametric volatility estimation on the real line from low-frequency data. *The art of semiparametrics*, 32–48, Physica, Heidelberg.

[87] Reynaud-Bouret, P. and Schbath, S. (2009). Adaptive estimation for Hawkes processes ; application to genome analysis. *The Annals of Statistics*, **39**, 2781–2822.

[88] Rodriguez-Iturbe, I., Cox, D.R. and Isham, V. (1988). A Point Process Model for Rainfall : Further Developments. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **417**, 283–298.

[89] Russell, J.R. and Engle, R.F. (2005). A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times. *Journal of Business and Economic Statistics*, **23**, 166–180.

[90] Sabhapandit, S. (2011). Record Statistics of Continuous Time Random Walk. *Arxiv preprint*, 10081762v2.

[91] Sato, K-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

[92] Scalas, E. (2004). Five years of continuous-time random walks in Econophysics. *Proceedings of WEHIA 2004*.

[93] Scalas, E., Gorenflo, R., Luckock, H., Mainardi, F., Mantelli, M. and Raberto, M. (2005). Anomalous waiting times in high-frequency financial data. *Arxiv preprint,* 0505210v1.

[94] Scalas, E. (2006). The application of continuous-time random walks in finance and economics. *Physica A,* **362**, 225–239.

[95] Schumer, R., Baeumer, B. and Meerschaert, M.M. (2011). Extremal behavior of a coupled continuous random walk. *Physica A,* **390**, 505–511.

[96] Shevtsova, I. (2007). On the accuracy of the normal approximation to the distributions of Poisson random sums. *Proceedings in Applied Mathematics and Mechanics*, **7**, 2080025–2080026.

[97] Shimizu, Y. (2006). Density estimation of Lévy measures for discretely observed diffusion processes with jumps. *Journal of The Japan Statistical Society,* **36**,37–62.

[98] Sørensen, M. (2008). Parametric inference for discretely sampled stochastic differential equations. *CREATES Reasearch Paper*.

[99] Tsybakov, A.B (2008). *Introduction to Nonparametric Estimation*. Springer.

[100] Uchaikin, V.V. and Zolotarev, V.M. (1999). *Chance and stability. Stable Distributions and their Applications.* VSP, Utrecht.

[101] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[102] Vardi, Y. (1982). Nonparametric estimation in renewal processes. *The Annals of Statistics*, **10**, 772–785.

[103] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli*, **13**, 672–694.

[104] van Es, B., Spreij, P. and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli*, **9**, 451–465.

[105] van Es, B., Spreij, P. and van Zanten, H. (2011). Nonparametric methods for volatility density estimation. *Advanced mathematical methods for finance*, 293–312.

[106] Vlahos, L., Isliker, H., Kominis, Y. and Hizanidis, K. (2008). Normal and Anomalous Diffusion : A Tutorial. *Arxiv preprint*, 08050419v1.

[107] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, **41**, 220–242. (parametric)

[108] Watkins, N.W. and Credgington, D. (2008). A kinetic equation for linear fractional stable motion with applications to space plasma physics. *Arxiv preprint* 08032833v1.

[109] Watson, G.N (1922). *A Treatise on the Theory of Bessel Functions*. Cambridge University Press.